Title: The word frequency effect in short-term serial recall

Author: Leonie Miller

Year: 2010

Repository DOI:

The word frequency effect in short-term serial recall

Leonie M. Miller
University of Wollongong

The word frequency effect in short-term serial recall
A thesis submitted in fulfilment of the
requirements for the award of the degree

Doctor of Philosophy

from

UNIVERSITY OF WOLLONGONG

By

**Leonie M. Miller, B. Ed. (Sec. Math., with Distinction), B. Math. (Hons),
B. Psyc. (Hons)**

School of Psychology

June 2010

Thesis Certification

CERTIFICATION

I, Leonie M. Miller, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Psychology, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Leonie M. Miller

June 9 2010

Abstract

Recent research into the nature of the frequency effect in immediate serial recall has revealed that some aspects of the mnemonic influence of word frequency over the short-term are not well accommodated by current explanations of the effect (i.e. item-based redintegration). In particular, the finding that how well a word is recalled is dependent on the relationship between that word's frequency and the frequencies of other list items suggests that processes far greater in complexity than previously assumed underpin the encoding, retention and retrieval of to-be-remembered material. This thesis assesses the word frequency effect according to two lines of investigation. It firstly examines the relationship of word frequency with a second lexical-semantic variable, concreteness, and determines that (i) the size of frequency effect obtained is influenced by the concreteness of the stimuli, and that (ii) these variables appear to behave similarly across serial positions. The architecture of language-based models of STM is argued to be consistent with these findings. A second series of studies considers the influence of item arrangement in lists of mixed frequency and uncovers the presence of directionally-sensitive and non-directional associative effects operating in the early and late serial positions, respectively. These results are also considered to be most compatible with language-based explanations of memory given their capacity to reflect associativities that have developed through natural language use as well as their accommodation of early-stage lexical-semantic influences. However the transformation from directional to non-directional associativity as list recall proceeds requires further research to better articulate the responsible mechanism(s). Possible future avenues of investigation are presented, and the research is discussed with reference to broader theoretical issues (e.g. the separation of item and order mechanisms, unitary versus two-store accounts of short-term memory).

Acknowledgements

Firstly, I would like to acknowledge and to thank my supervisor Steven Roodenrys for his continued support in completing the many facets of this project, and his enduring patience with my perfectionism. I owe thanks to my co-supervisor Stephen Palmisano for his words of encouragement and advice. I would also like to recognise the assistance of the staff of the School of Psychology during my time as a postgraduate student; their interest in my work and support in the day-to-day aspects of study has meant a great deal to me. To my fellow postgraduate students, thank you for sharing, normalising, and cheering me over the line. Lastly, to my friends and family I must express my deep gratitude for the patience and support given to me throughout this process – thank you for enduring my absences and preoccupations with grace and good humour. I promise to be more available to you all very soon!

Table of Contents

Page

Page

Page

Page

Page

Page

Page

Page

Page

Page

List of Tables

Table Page

List of Figures

Figure Page

Chapter 1

The Frequency Effect in Serial Recall

1.1 Background

Some time ago, Hulme et al. (1997) stated that the effect of word frequency on short-term memory (STM) was complex and not well understood. More than a decade later, despite continued investigation (e.g. Allen & Hulme, 2006; Hulme, Stuart, Brown, & Morin, 2003; Jefferies, Frankish, & Lambon Ralph, 2006a; Morin, Poirier, Fortin, & Hulme, 2006; Saint-Aubin & LeBlanc, 2005; Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000, 2009; Tse & Altarriba, 2007; Woodward, Macken, & Jones, 2008), specifically in the area of short-term serial recall, this assessment still applies; mechanisms that explain memory performance associated with word frequency are confined to conceptual accounts (see Hulme et al., 2003; Stuart & Hulme, 2009; Saint-Aubin & LeBlanc, 2005) and the lack of activity in the development of computational models explaining the range of frequency-related recall phenomena attests to both the complexity of the role (or roles) of word frequency within serially-ordered recall, and the limited understanding of how this variable influences performance. Across experimental manipulations, the observed effects of word frequency have motivated a shift in the bases of explanations of serial recall to ones that consider list context and pre-experimental association (Jefferies et al., 2006a; Hulme et al., 2003; Stuart & Hulme, 2000, 2009), distinctiveness-based processing (Saint-Aubin & LeBlanc, 2005), or the impacts of word boundaries on co-articulation (Woodward et al., 2008), and contrast with earlier accounts focusing on either item-based differences in knowledge of the phonological forms of items (Hulme et al., 1997), presumed rehearsal rate differences between high frequency (HF) and low frequency (LF) words (Baddeley, Thomson, & Buchanen, 1975; Wright, 1979), or the interaction of a long-term memory (LTM) effect with the selective retrieval from short-term or long-term stores as a function of serial position (M.J. Watkins, 1977).

1.2 Motivation of the thesis

Miller (2004) conducted a study examining the relationship between serial recall performance and the semantic neighbourhood of items, and included word frequency as an additional predictor. This experiment was an analogue to Experiment 4 reported by Roodenrys, Hulme, Lethbridge, Hinton, and Nimmo (2002) where phonological neighbourhood measures and word frequency were regressed onto the item recall of individual participants (Lorch & Myers, 1990). Miller (2004) however was surprised to observe a marked attenuation in the frequency effect in this case, an order of magnitude smaller than that found by Roodenrys et al. (2002). A comparison between these experiments using traditional regression methods, that is, collapsing performance across participants, showed that the frequency effect achieved by Miller (2004) accounted for approximately 5% of the total variance, while the equivalent analysis of the Roodenrys et al. (2002) data implied word frequency accounted for approximately 30% of the total variance. Subsidiary analyses of the Miller (2004) stimuli failed to find any confound with other variables known to influence the size of the frequency effect.

However, Miller (2004) had used items from a restricted concreteness domain, as all words were highly concrete. Furthermore, the procedure in this task involved the use of mixed frequency lists; HF and LF items were randomly allocated to serial positions, and despite the Roodenrys et al. (2002) data producing a substantial effect, other recent literature (e.g. Hulme et al., 2003) had shown that the behaviour of mixed lists was not well accounted for.

1.3 Objectives of the thesis

The primary objective of this thesis, motivated by the experience of Miller (2004), is to contribute to the understanding of the frequency effect in serial recall. To this end it pursues two lines of investigation. Firstly, it examines the effect word concreteness has on the size of the frequency effect in the serial recall of words as a follow-up to Miller (2004). This is explored in two experiments that use factorial manipulation on pure lists. The second focus of study involves a closer inspection of the frequency effect when list composition and item arrangement within the list are varied. Following Hulme et al. (2003), the serial recall performance on three forms of mixed list is investigated in four experiments.

1.4 Structure of the thesis

Before presenting and discussing the experimental data, it is appropriate to define the key terms and provide a review of the research involving word frequency and serial recall. The definitions of terms are covered in Chapter 2, along with a summary of early two-store accounts of memory, and a précis of the nature of frequency effect in other tasks. Much of the early research on the frequency effect in STM was conducted using memory span experiments, and these are reviewed in Chapter 3. Chapters 4 - 6 deal with the observations of the frequency effect in experiments involving the serial recall of pure or mixed supraspan lists. Specifically, Chapter 4 covers the research performed using pure lists, while Chapter 5 examines those studies that have supported an item co-occurrence interpretation of the frequency effect (Hulme et al., 2003; Stuart & Hulme, 2000). Chapter 6 presents alternative conceptions of the frequency effect motivated primarily by research on mixed lists (e.g. Jefferies et al., 2006a; Saint-Aubin & LeBlanc, 2005; Saint-Aubin & Poirier, 2005).

Chapter 7 examines the relationship between word frequency and memory for order and considers evidence for the locus of effect of word frequency and lexical-semantic variables more generally in STM tasks, while Chapter 8 summarises the capacities or otherwise of computational models to explain the frequency effect in its variant forms. Models that have been endorsed in the literature by other researchers for example, the feature model (Nairne, 1990; Neath & Nairne, 1995) and the Temporal Context Model (TCM) (Howard & Kahana, 2002a, 2002b) and SIMPLE (G.D.A. Brown, Neath, & Chater, 2007) are given particular examination.

The frequency by concreteness experiments (Chapters 10 and 11) are prefaced by a review of the word concreteness effect in serial recall in Chapter 9. These experiments confirm the presence of an interaction between frequency and concreteness.

Chapters 12-15 report a series of experiments that were designed to determine whether the frequency effect in mixed lists is better described as a generalised list-level effect, as initially conceived by Hulme et al. (2003; Stuart & Hulme, 2000, 2009), or as an effect that is sensitive to pre-experimental inter-item associations of adjacent list items (Howard & Kahana, 2002b; Hulme et al., 2003). These experiments find support for a directional, that is, item-to-item influence associated with word frequency, although this effect is apparent for early but not late list items.

Chapter 16 contains a General Discussion that considers how well the results of the pure and mixed list-based investigations conform to existing views of the frequency effect in short-term serial recall (e.g. Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000; 2009). This chapter also considers other recent conceptual accounts of serial recall (e.g., Allen & Hulme, 2006; Majerus, 2009; Ward, Tan, & Bhatarah, 2009) and how these relate to the outcomes of the studies conducted. More general memory mechanisms (e.g. rehearsal - Ward et al., 2009, and recall processes - Nairne, Ceo, & Reysen, 2007) are proposed as processes that might be responsible for the restriction of item-to-item associativity to early serial positions, highlighting a potential interaction between systems that manage item and order information in memory (Majerus, 2009). Lastly, Chapter 16 considers limitations of the current research and identifies areas for future investigation thought to extend the understanding of the complexities of the frequency effect in short-term serial recall.

Chapter 2

Definitions and Historical Perspectives

2.1 Introduction

This chapter provides definitions to the key terms relevant to this thesis. It includes early conceptions of STM and the frequency effect, offering a starting point for the review of the literature regarding the word frequency effect in serial recall, presented in the following chapters.

2.2 Word frequency

Word frequency, an index of experience with verbal items (Wright, 1979), is the most studied variable of the properties of language items in STM (Morin et al., 2006). Measures of word frequency estimate how often a word is encountered in written or spoken language (Stuart & Hulme, 2000), and are typically obtained from established databases that are based on large text corpora (e.g. CELEX - Baayen, Pipenbrook & Rijan, 1993; G.D.A. Brown, 1984; Kucera & Francis, 1967; Thorndike-Lorge, 1944). These counts are usually expressed in terms of the number of instances per million words of text. Word frequency therefore reflects differences in knowledge of, or familiarity with, the phonological forms of language items and, as such, is an LTM variable (Hulme et al., 1997). Due to the evolving nature of language, measures that best reflect the language habits of the population under investigation should be used (Gregg, 1976).

Word frequency is highly confounded with other attributes of language. High-frequency (HF) words have greater complexity in meaning and more associates than low-frequency (LF) words (Deese, 1959, 1960; Gregg, 1976). Word frequency is also positively associated with semantic class size and negatively associated with age of acquisition (AoA) (Dent, Johnston, & Humphreys, 2008; Gregg, 1976; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994). High frequency items are more likely to be rated as more pleasant than LF words (Gregg, 1976), and are articulated faster than LF words (e.g. Roodenrys et al., 1994).

2.3 Word frequency effects in language processing tasks and other memory tasks

Word frequency is a variable that influences performance on a wide range of tasks. High-frequency words are processed faster or produced as responses more quickly than LF words in word discrimination (e.g. lexical decision tasks, Glanzer & Adams, 1990; Gregg, 1976; Luce & Pisoni, 1998; Monsell, Doyle, & Haggard, 1989), identification in noise (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990) and print (Monsell et al., 1989), picture-naming (Dent et al., 2008), visual naming, semantic categorisation and noun/adjective classification (Monsell et al., 1989). High frequency words also demonstrate better identification with increased levels of noise than do LF words (Howes, 1957). The ease of solution with anagrams is greater for HF than LF words (Mayzner & Tresselt, 1958). In the absence of context, polysemous nouns are encoded according to their most frequent meaning (Winnograd & Conn, 1971). High-frequency words are produced as category exemplars faster than LF words (Freedman & Loftus, 1971) and words are emitted more quickly as exemplars for higher frequency categories (Loftus & Freedman, 1972).

Therefore, HF words enjoy a performance advantage across a wide range of tasks. However, in word recognition memory tasks, LF words exhibit a higher hit rate and lower false alarm rate than HF words (Diana & Reder, 2006; Glanzer & Adams, 1990), but this relationship does not extend to very low frequency items (Mulligan, 2001).

In list learning, lists of HF words are learned to criterion faster than lists of LF words (Sumby, 1963), however in mixed lists LF items are learned faster than in pure lists (May & Tryk, 1970). Postman, Turnage and Silverstein (1964) found HF words produced better memory for items than LF words in a running memory span task. The general pattern found in free recall is that the HF advantage in pure lists is not observed in mixed lists (e.g. DeLosh & McDaniels, 1996; M.J. Watkins, LeCompte, & Kim, 2000); the typical finding is that, in comparison to pure list performance, the recall of HF words is impaired, while recall for LF items is enhanced (Gregg, 1976; Hulme et al., 2003). In some circumstances this has resulted in an LF advantage in mixed lists (e.g. DeLosh & McDaniels, 1996). Therefore, in this case the frequency effect is malleable and dependent on the composition of the presented list.

2.4 Verbal STM

Verbal STM has been described as a labile memory system (Baddeley, 1972; Baddeley & Ecob, 1970), otherwise referred to as primary memory (James, 1950; Murdock, 1967; Waugh & Norman, 1965) that contains the changing verbal contents of an individual's current mental activity, and underpins their ability to recite verbatim the last words that they have heard or spoken (Waugh & Norman, 1965). While such retention is not thought to critically support comprehension (Baddeley, 1986), more recently verbal STM has been argued to play a key role in the vocabulary development of children (Baddeley & Hitch, 1994; Baddeley, Gathercole, & Papagno, 1998). In contrast, long-term or secondary memory is the memory component that stores accumulated experience and knowledge in a more durable form (Baddeley & Ecob, 1970).

A key component of verbal STM is memory for order, the retention of material according to its presented sequence. The preservation of order of verbal information is a necessary feature of a mechanism associated with the processing of language. The number of items recalled in order, immediately after sequence presentation, has been considered an indicator of STM capacity (J. Brown, 1958; Waugh & Norman, 1965).

The weight of early findings using STM tasks suggested that performance was highly dependent on the use of a phonological code (e.g. Baddeley, 1966a; Baddeley & Hitch, 1974; Conrad, 1964; Conrad & Hull, 1964). The prevalence of phonological errors in serial recall (Conrad, 1964), and the impairment to recall when lists contained phonologically similar items (Baddeley, 1966a; Conrad, 1964; Conrad & Hull, 1964; Murray, 1968) reflected a system that was predominantly speech-based. By comparison, long-term memory (LTM) performance was argued to be mostly reliant on semantic encoding, to which STM was considered, on the whole, insensitive (Baddeley, 1966a, 1966b, 1972; Baddeley & Levy, 1971).

2.5 Measures of serial recall

Short-term serial recall tasks involve the maintenance of a series of representations of phonological forms presented in a list (Hulme et al., 1997). These forms are usually words although in some cases items can be nonwords (e.g. Hulme, Maughan, & Brown, 1991). Tasks involve the auditory or visual presentation of a list of items followed a recall phase where the participant attempts to recall the items in their presentation order.

The number of items in the lists employed in STM tasks are less than for lists used in studies more focused on LTM processing (e.g. free recall) (Klein, Addis, & Kahana, 2005), typically 5-7 items. Experiment-specific list length varies according to the stimulus properties under investigation, the STM capacities of the participants, and the nature of the task.

2.5.1 Word span

Memory (or word) span tasks test short-term recall capacity by determining the greatest list length that can be recalled without error in serial order after a single presentation (Hulme et al., 1991; Waugh, 1960). While this task has been utilised traditionally in experimental psychology as a means to explore memory capacities (e.g. Brener, 1940; Ebbinghaus, 1913), it is also used in clinical psychology and cognitive neuropsychology to provide estimates of memory functioning (e.g. Milner, 1968), and is a task included in intelligence tests (e.g. digit span - Weschler scales, Baddeley et al., 1998; Hulme et al., 1991). The average adult span is 5-6 items (Brener, 1940), but this is dependent on the to-be-remembered material (Broadbent, 1987). For example, average digit span has been estimated to lie within 5-8 items (Atkinson & Shiffrin, 1968).

The method used to establish memory span can vary; some procedures adjust list length in response to participant performance (e.g. Hulme et al., 1991; Hulme et al., 1997; M.J. Watkins, 1977) while others present sets of lists with predetermined lengths (the method of constant stimuli, e.g. Drewnowski & Murdock, 1980). Procedures also vary in terms of whether a termination criterion is applied based on performance (Hulme et al., 1991; Hulme et al., 1997) or whether testing ceases after a fixed number of trials (Engle, Nations, & Cantor, 1990; M.J. Watkins, 1977). The precision of span scores is also influenced by the number of trials presented to a participant with list length close to their span limit (Hulme et al., 1991).

While memory span is considered to be a relatively uncontaminated measure of STM capacity (although see Hulme et al., 1991 for a different perspective), it is diagnostically limited because it does not capture types of recall errors, and therefore does not provide information that points to likely mechanisms constraining memory performance (Roodenrys et al., 2002). Additionally, span measures do not yield data revealing how differences in test variables can change as a function of the serial position of items.

2.5.2 Serial recall of supraspan lists

Memory span is contrasted with the immediate serial recall (ISR) of supraspan lists, more commonly referred to as the ISR task (Poirier & Saint-Aubin, 1996). This task is the most frequently utilised test of order retention among STM researchers (Roodenrys & Quinlan, 2000). In this paradigm, lists just greater than average span are presented to participants and their recall performance across list positions is measured.

The recall of items in supraspan lists is argued to be a complex process (M.J. Watkins, 1977). The presentation of lists of item length above normal span can encourage participants to adopt strategies that maximise their recall. Unlike memory span tasks that produce a single estimate of STM capacity, recall in ISR tasks produce performance patterns across a fixed number of serial positions, revealing information about how STM recall operates and breaks down when overloaded (M.J. Watkins, 1977). The breakdown in recall performance for supraspan list length occurs in a particular way (Tan & Ward, 2007). Characteristic serial position curves exhibit enhanced memory for primacy items (those items presented in the foremost portion of the list), a smaller recency effect involving the recall of final list items, while the recall for the medial items is poorest. Error patterns provide information about the nature of operations and the types of constraints that exist in STM functioning (Drewnowski & Murdock, 1980; Roodenrys et al., 2002).

Measures used in serial recall analysis reflect different aspects of memory functioning (Poirier & Saint-Aubin, 1996). The scoring of recall performance using a strict criterion, that is when correct recall is conditional on the item being recalled in its presented position, is a measure that confounds item and order information (Murdock, 1976; Poirier & Saint-Aubin, 1995), and therefore is sensitive to differences in either. In contrast, the scoring of recall data using a lenient criterion, where the number of correctly recalled items regardless of order is taken, is a measure of item memory.

Differences in order memory can also be examined using conditionalised rates of order errors (Poirier & Saint-Aubin, 1996). These are established by expressing, for each condition, the number of items recalled in an incorrect serial position as a proportion of the number of items recalled regardless of position. Conditionalising the order error data in this way is argued to reveal effects of order memory unconfounded by effects of item memory, and expresses the rate of order errors given a particular level of item memory (Murdock, 1976).

It is noteworthy that this type of error analysis assumes independence between item and order errors. That is, the failure to recall an item, referred to as an omission, is assumed to reflect a breakdown in item memory and is classified an item error. However, failed recall makes it impossible to determine whether the unrecovered item was from another position in the list (Jefferies et al., 2006a). Therefore such analysis assumes that these events, should they occur, remain in proportion to the number of realised order errors across conditions.

2.6 Early accounts of STM

2.6.1 Unitary versus two-store conceptions

In simplest terms, the nature of STM and its relationship to LTM were defined by two opposing theoretical approaches. Continuum theorists argued that STM did not operate differently to LTM; the mechanisms that determine memory performance would be consistent across short- and long- term memory tasks (e.g. Melton, 1963; Postman, 1961; Underwood & Postman, 1960). While STM was that part of the memory continuum closest to the present moment, and STM tasks displayed qualitatively different responses to tasks involving memory over the longer term, STM encoding was permanent and subject to interference in the same way that long-term memories were encoded and forgotten (Melton, 1963). In either case, the major source of forgetting was due to the processing of similar material interfering with the selection of the correct memory representation at recall.

Other theorists however argued that memory performance across time scales reflected, at the very least, differential processing for short-term and long-term tasks (J. Brown, 1958, 1964), or the operation of two distinct memory stores (e.g. Atkinson & Shiffrin, 1968, 1971; Broadbent, 1987; Waugh & Norman, 1965). This latter position viewed STM as a capacity limited store that was subject to different processes from those operating in LTM. Short-term trace degradation was variously argued to reflect decay, the loss of information due to the trace ‘fading away’ with the passage of time (e.g. Atkinson & Shiffrin, 1968; J. Brown, 1958, 1964; Conrad, 1964), or displacement from a capacity limited store when capacity had been exceeded (Waugh & Norman, 1965). Furthermore, the process of rehearsal was considered central to maintaining memory performance, either through re-instantiation of a degraded short-term trace (J. Brown, 1958), or in the prevention of item loss and facilitation of encoding into LTM (Waugh & Norman, 1965).

The debate between interference and decay theorists regarding the properties of STM has been a long-lasting and ongoing one (see for example, Altmann, 2009; Baddeley, 2002, 2007; Nairne, 2002; Lewandowsky & Oberauer, 2008; Lewandowsky, Oberauer, & Brown, 2008, 2009; Surprenant & Neath, 2009). The idea of distinct long-term and short-term storage systems can be traced back to the nineteenth century (Atkinson & Shiffrin, 1971), and the concept of short-term decay had been linked to Thorndike’s (1923) theory of disuse (Melton, 1963); that ‘modifiable’ connections decrease in strength over time unless they are re-established (p.172). An accumulation of evidence and argument in the late nineteen-fifties and early nineteen-sixties (e.g. J. Brown, 1958; Conrad, 1964; Hebb, 1961; Peterson & Peterson, 1959) was seen to support the presence of decay as an active influence on STM.

However, McGeoch (1932) had maintained that, in general, forgetting in memory was due to factors other than disuse, namely the similarity of material processed during the period of retention, and the effects of encoding context on recall. Furthermore, Melton (1963) argued that closer inspection of a keystone defense of STM decay, the reduction in short-term recall associated with filled retention intervals (see Peterson & Peterson, 1959) revealed that interference from other list items, and proactive interference from items in previous trials made substantial contributions to short-term forgetting (see also, Keppel & Underwood, 1962; Murdock, 1961). With respect to memory for serial order, Keppel and Underwood (1962; Postman, 1961, 1962; Turnage, 1967; Underwood & Postman, 1960) proposed that associative interference might influence memory performance; unlearning over time would occur because pre-experimentally learned associations between items interfered with the integrity of the to-be-remembered sequence. A second interference-based explanation of STM suggested that the influence of interference on the contents of STM would be similar to exposure to an ‘acid-bath’, with trace degradation dependent on the time items were left unrehearsed (Posner & Konick, 1966); in this case interference would influence the retention of item traces rather than contribute to the difficulty of item selection at retrieval.

The continuum approach to memory was regarded as the dominant view until the mid 1960’s, when favour for the introspective coherence of decay-based explanations of STM, the precedence of information processing designs containing short-term and long-term stores (Broadbent, 1987), the assertion of the phenomenological differences between short-term and long-term memory (Waugh & Norman, 1965), and dissociations of STM and LTM with neuropsychological patients (e.g. Milner, 1968), supported the development of two-store models. Murdock (1967) referred to the two-store position as the ‘modal model’, although in his description of the generalised two-store account, displacement of items was responsible for the loss of information across brief time periods (p. 428).

2.6.2 Two-store models

2.6.2.1 Atkinson and Shiffrin (1968, 1971)

Atkinson and Shiffrin (1968, 1971) argued that human memory processes indicated a division between long-term and short-term memory stores (LTS, STS). Evidence for this dichotomy was found in anterograde amnesic patients who exhibited normal STM functioning coupled with an inability to transfer information to, or retrieve information from the LTS (Milner, 1968). According to this account, the STS contained auditory-verbal traces that experienced rapid decay in the absence of rehearsal. Decay was thought to complete within a 30 second period and had been demonstrated experimentally; for example, Peterson and Peterson (1959) had shown that recall of a consonant trigram, when rehearsal had been prohibited by a counting task, reduced markedly as the delay increased from 3 to 18 seconds. However, Atkinson and Shiffrin (1971) conceded that other results (e.g. Atkinson & Shiffrin, 1971; Reitman, 1971) suggested that interference from the intervening arithmetic task rather than the passage of time was the more influential factor in determining the level of recall.

Atkinson and Shiffrin (1968) considered that a speech-based code dominated processing in the STS, evident by sensitivity of correct recall to the phonological similarity of items. Conrad (1964) demonstrated that visually distinct but phonologically similar items induced confusions in an STM task, and that these corresponded to the confusions made by participants when attempting to identify items in noise. Items that sounded alike were most likely to be interchanged. Traces in the LTS were by contrast, durable, of any sensory modality, and potentially time-stamped (Atkinson & Shiffrin, 1968). Furthermore, while long-term traces were primarily subject to interference, decay operated in cases where weak or insufficient long-term coding had occurred.

The short-term processing of information was believed to lead to its automatic transfer to the LTS (Atkinson & Shiffrin, 1968). That such transfer existed was argued to be reflected in the enhanced recall of sequences repeated at intermittent intervals (Hebb, 1961). Occupancy within the rehearsal buffer dictated the degree of accumulation of information in the LTS; items maintained by rehearsal for longer periods of time would be more likely to form established long-term traces. Additionally, the efficiency of transfer to LTM depended on the number of items in the rehearsal buffer. Early list items would be encoded more efficiently as greater relative attention could be directed to this process until the rehearsal buffer reached capacity. The presentation of new items after this capacity was exhausted would result in the displacement of old items from the store (Atkinson & Shiffrin, 1971).

Recall was presumed to involve an inspection of the contents of the STS and direct retrieval from this buffer when appropriate, and a search of the long-term store when items had otherwise not been located. In the latter case, it was proposed that the most recently presented items in the LTS were differentiated from other LTM contents by their associated time-stamps, placing them along a temporal dimension (Atkinson & Shiffrin, 1968). The recency effect, the better recall for the final list items, was presumed to reflect retrieval from the current contents in the STS for both serial and free recall. Similarly, better recall associated with the primacy portion of the list resulted from items that were encoded into and retrieved from the LTS. These effects, present in both serial and free recall tasks, were therefore thought to arise from the operation of common mechanisms.

2.6.2.2 The working memory model (Baddeley & Hitch, 1974)

The Atkinson and Shiffrin (1968, 1971) model had used as its focus the control of cognition as the central feature of the STS, and this was equated with terms like working memory (Atkinson & Shiffrin, 1971; Shiffrin, 1993). Baddeley and Hitch (1974) fractionated this ‘short-term store’ into a working memory model that itself comprised two slave systems for the retention of verbal and visuospatial information respectively, and a central executive component that acted primarily as an attentional control system (Baddeley & Hitch, 1994). The central executive directed processing and task switching between slave systems during complex problem solving, and as initially conceived, possessed some additional storage capacity.

The short-term verbal storage and retention system, the articulatory (and later phonological - Baddeley, 1997) loop, was characterised as a serially ordered, speech-based store that was responsible for memory span performance (Baddeley, 1986). It contained a phonological short-term store and a subvocal rehearsal mechanism, and was envisaged as a fixed capacity system that was subject to time-based decay, where phonologically encoded items held in the store required rehearsal in the loop to prevent trace decay beyond the point of retrieval (Baddeley, 1986; Baddeley et al., 1998; Baddeley et al., 1975). At recall, items were identified from existing traces resident in the phonological short-term store.

The support for this system was argued to come from a number of sources. The basis for phonological encoding was found in the phonological similarity effect (Conrad & Hull, 1964), the impaired recall for similarly sounding items, and the finding that articulatory suppression (AS), the prevention of subvocal rehearsal by the continuous overt rehearsal of non-target material, eliminated this effect for visual presentation (Murray, 1968). This second result was explained in terms of the prevention of the phonological recoding of visual stimuli into the short-term store by AS; if recall, in this case, was not dependent on phonological encoding, then it would not be sensitive to the phonological similarity of items and no effect would arise.

Baddeley and Hitch (1974) performed a series of experiments using concurrent tasks that tested whether the capacity responsible for memory span was the same resource that determined the recency effect in free recall. While Atkinson and Shiffrin (1968, 1971) had maintained that the superior recall for the final list items in free recall would be due to their accessability from the short-term store, Baddeley and

Hitch (1974) showed that free recall recency was not impaired when a memory span task was inserted between the presentations of each list item. Under these conditions the abolition of the recency effect would be expected if the short-term store had been occupied by the intervening memory span task. Therefore, Baddeley and Hitch (1974) reasoned that free recall recency and memory span were products of two different systems.

The phonological loop component of the Baddeley and Hitch (1974) working memory model became an influential account of verbal STM for the following two decades. Importantly, it focused on the processing and retention of speech-based material in isolation of LTM effects, and was argued to reflect a memory capacity not responsible for the formation of the recency effect in free recall, and thus encouraged a divergence in the research of order-free and serially ordered memory (Stuart & Hulme, 2009).

2.7 The unit-sequence hypothesis and the word frequency effect in serial recall

Early theories regarding the effects of frequency on memory tasks focused on the strengths of associations between items that were derived from natural language use (Deese, 1959, 1960; Postman, 1961, 1962; Underwood & Postman, 1960). With respect to serial recall, the unit-sequence interference hypothesis (Postman, 1961, 1962; Turnage, 1967; Underwood & Postman, 1960) proposed that ordered memory for HF items should be worse than for LF items because the stronger pre-existing associations between HF items would interfere with the order of the test sequence, inducing faster forgetting than for LF words.

Baddeley and Scott (1971) tested this proposal and demonstrated that there was no difference in the rate of forgetting of short sequences of HF and LF words over a filled delay period that extended from 0 – 30 seconds. A recall advantage to HF words was reported in all experiments that matched sequences on the number of items. One experiment that tested the recall of 3-word LF sequences against 4-word HF sequences found a recall advantage to LF words, but in all cases the differences in effect observed with immediate recall did not change across retention intervals, indicating that HF words suffered no greater rate of forgetting than LF words.

Baddeley and Scott (1971) concluded that unit-sequence interference was unlikely to be an active influence on the nature of short-term forgetting.

2.8 Conclusion

Across many cognitive tasks, word frequency is a much studied variable. Frequency is associated with a number of other linguistic factors that leave open the possibility of confounding effects with additional variables (e.g. Deese, 1960). Most tasks reveal an advantage for HF words however some activities (e.g. free recall) demonstrate that effects are variable depending on the list context (Gregg, 1976). Prior to the systematic investigation of the effect of frequency on serial recall, influential models of STM assumed that memory functioning operated according to distinct stores for short-term and long-term memory. These models emphasised either the transfer of information from STM to LTM (Atkinson & Shiffrin, 1968, 1971) or the phonological attributes of studied material in the absence of LTM influences (Baddeley, 1986; Baddeley & Hitch, 1974). An early interference-based hypothesis making specific frequency-based predictions (the unit-sequence interference hypothesis) was found to be unsupported (Baddeley & Scott, 1971) suggesting that this type of influence was not present within the typical timescale of STM operation.

Chapter 3

Word Frequency and Memory Span

3.1 Introduction

The majority of early studies into the frequency effect in serial recall involved word span procedures. These studies therefore addressed whether STM capacity varied when the familiarity of the list stimuli was manipulated.

3.2 Memory span of pure and mixed lists

M.J. Watkins (1977) examined the word span of pure- and mixed-frequency lists. Lists were constructed so that list halves contained items drawn from either HF or LF stimulus sets. In pure lists, both list halves were drawn from the same pool, while mixed lists were formed when the list halves came from opposing sets.

M.J. Watkins (1977) sought to demonstrate that span was a result of complicated activity involving multiple processes rather than a single process (STM). On the basis of evidence in the free recall literature (e.g. Raymond, 1969; Sumby, 1963) and in serial recall studies (O.C. Watkins & Watkins, 1977), he argued that the frequency effect occurred for the early but not late serial positions in the list. He reasoned there must be an episodic LTM component to performance; early list items would be recalled from LTM, while items in the recency positions of the list would be recalled from STM. Consequently, as variables associated with LTM should produce differences in recall for early but not final list items, an impact on span measures should be observable when list composition was suitably manipulated.

More specifically, M.J. Watkins (1977) claimed that if span was a result of a single process, STM, then lists whose halves vary according to different levels of an LTM variable should produce equivalent memory span. Conversely, differences in span with mixed lists of this form would be evidence that STM tasks are complex, utilising multiple processes in recall, and reliant on long-term knowledge.

The results confirmed the frequency effect for pure lists, with span for HF lists greater than LF lists. Differences in mixed lists were also significant as span was greater when HF items filled the first half of the list rather than LF items, though spans for mixed lists were intermediate with respect to pure list span scores. Span values of lists where HF words occupied the second half of list items were greater than lists where LF items occupied the second half of list items. Therefore frequency continued to pose on influence on recall into the second half of the list. Despite this, M.J. Watkins (1977) argued that an LTM influence across all list positions was not necessarily implied. He concluded that the pattern of observations was consistent with an account of short-term recall that involved at least two processes.

3.3 A relationship between word frequency and rehearsal

Wright (1979) raised the possibility that M.J. Watkins' (1977) interpretation had overlooked effects of frequency on processes other than internal memory activities, specifically differences in word pronunciation. He suggested the results of M.J. Watkins (1977) could be explained within a single-process account that was sensitive to differences in articulation duration, and therefore one that did not rely on LTM knowledge.

Wright (1979) examined spoken duration according to three different methods; comparing the reading rate of lists of words with the rate of repetition for short, well-learned lists (following Baddeley, et al., 1975) and with the spoken rate of HF and LF words when items were spoken individually or in simple carrier phrases.

He found that LF words took more time to pronounce than HF words matched on the number of letters, to the extent that the amount of time required to produce a six-word utterance of HF words was equivalent to that for a sequence of five LF words. Furthermore he determined that LF words have more phonemes than HF words with the same number of letters, and these phonemes tended to take longer to pronounce.

Wright (1979) reasoned articulation speed might well reflect rehearsal speed producing an advantage to items that were rehearsed more often within a trial. Effects arising from differences in rehearsal speed could be well accommodated by the account of the phonological loop (Baddeley & Hitch, 1974); if HF items were articulated faster than LF words they would benefit from the increased rate of rehearsal used to maintain the integrity of item traces before output.

The difference found between the High-Low and Low-High conditions in M.J. Watkins' (1977) experiment was also explainable in these terms, if additional assumptions were made. The first of these involved the covert rehearsal of list items prior to output. If HF items were produced more quickly than LF items, then in a relative sense, words in the last serial positions of a list whose first half comprised LF items would be recalled later than the corresponding words in a list whose first items were HF words. If time to output was related to the success of recall then performance on High-Low lists would be superior to the performance on Low-High lists. Secondly, if the rehearsal of part-lists during presentation influenced the level of recall, then lists possessing HF items in the front end would give greater rehearsal opportunities for items in the second half of the list. If there was a linear relationship between the number of rehearsals for an item and the likelihood of recall, then High-Low lists would be better recalled than Low-High lists.

Geffen and Luszcz (1983) made a qualification to Wright's (1979) observation of shorter list reading times for HF than LF words. They replicated his experiment measuring both articulation durations and pauses between words, and found that the differences in overall reading times were due to differences in pauses; the articulation of HF and LF items *per se* were similar. Geffen and Luszcz (1983) were therefore less convinced of an articulatory difference based on word frequency.

3.4 Memory span of pure and mixed lists across age groups

Kausler and Puckett (1979) replicated M.J. Watkins' (1977) word span task, on samples of young and elderly participants. As M.J. Watkins (1977) had claimed an LTM contribution was of greater benefit to span when HF items were presented in the front half of the list, and it was thought this contribution might attenuate for older participants, it was predicted that the difference in performance between age groups would be greater for High-Low than Low-High lists. The study also sought to establish age-related trends for serially ordered memory, and the memory span task specifically.

Memory spans were smaller for elderly than young participants for all list types. List type effects across age groups showed that while span for HF lists was the greatest, and span for LF lists was the least, the spans for High-Low and Low-High lists were equivalent and intermediate. Furthermore, span performances on mixed lists did not differ when age groups were examined separately. The interaction was driven by a greater difference in the recall of pure HF to pure LF lists for young than older participants.

Kausler and Puckett (1979) suggested subtle variations in method might be responsible for the disparity between their results and those of M.J. Watkins (1977), but that it was also possible the span difference found by M.J. Watkins (1977) might be the consequence of a Type I error. The authors argued that this experiment provided evidence for an age-related decrement in memory span due to less efficient processing in the rehearsal and organisation of list items.

3.5 Memory span and articulatory suppression

Tehan and Humphreys (1988) examined word span in a factorial design manipulating word frequency and word class, as a test of the phonological loop's (Baddeley & Hitch, 1974) capacity to explain differences in recall. Specifically, they wished to determine whether rehearsal rates covaried with differences in span as word length had been shown to (Baddeley et al., 1975), and whether AS, assumed to block rehearsal and remove pronunciation duration differences, would eliminate effects observed in control conditions when visual presentation was used. Rehearsal speed was approximated by the length of time it took participants to read lists of items from each stimulus condition (Baddeley et al., 1975; Wright, 1979).

Analysis of the rehearsal speed data showed that despite matching stimuli on spoken duration (in terms of the time taken to pronounce each word individually), effects of frequency and word class were present, as was an interaction between these factors. The time taken to read lists of HF words was less than for LF words. Comparisons for the main effect of word class were not reported, however the interaction showed that differences between word class sets emerged for LF, but not HF items.

The span data revealed that HF words were better recalled than LF words, function words were recalled less well than either adjectives or nouns, and word span was worse under AS conditions than control conditions where presumably participants were able to rehearse freely. However, neither the interaction between frequency and suppression conditions, nor the word class and suppression conditions were significant; differences in span were unaltered when participants engaged in AS throughout the presentation of items. The three-way interaction was also found to be non-significant, indicating that the differences between HF and LF words across suppression conditions did not vary by word class.

These results demonstrated that differences in measures, assumed to reflect rehearsal rate, could occur between HF and LF words (Geffen & Luszcz, 1983; Wright, 1979), even when items are matched on pronunciation length and these differences were found to covary with observed differences in memory span. However, the patterns between reading rates and word class did not correspond in a consistent fashion; span differences were found for HF and LF sets of each word class but 'rehearsal' differences between word classes were found for LF words only. Specifically, the word class that had the slowest reading rate (nouns) had the highest span. These results challenged the generality of the phonological loop model as an explanatory mechanism.

A second concern regarding the accuracy of the phonological loop came from the observation that the frequency and word class effects in memory span were unaffected by AS using visual presentation. According to the phonological loop hypothesis, AS should inhibit the rehearsal of items during presentation and therefore limit any rehearsal-based differences on performance (Baddeley et al., 1975). Tehan and Humphreys (1988) concluded that frequency and word class effects could not be explained using the same mechanisms that described the effects of word length. They suggested that additional contributions to word span may be sourced from the central executive in the working memory model (Baddeley & Hitch, 1974).

Lastly, Tehan and Humphreys (1988) cautioned that finding a relationship between rehearsal rate measures and word span for different materials was insufficient evidence for the operation of the phonological loop. Word frequency had been shown to demonstrate such a relationship, although clearly, the failure to eliminate the effect under AS diminished the model's explanatory power.

Gregg, Freedman and Smith (1989) also sought to determine whether stimuli similar to those used by M.J. Watkins (1977) might show articulation rate differences. The time taken to complete ten repetitions of visually presented three-item sequences was measured and a reliable effect of articulation rate was found; HF words were spoken substantially faster than LF words.

Using similar reasoning to Tehan and Humphreys (1988), Gregg et al. (1989) argued that a suitable test of the single process theory of STM was to examine span measures for mixed lists of the type tested by M.J. Watkins (1977) under suppression conditions. Differences in span when the rehearsal mechanism was disabled would be evidence that short-term recall involved another process sensitive to frequency and not directly related to speech properties (Gregg et al., 1989).

Gregg et al. (1989) included AS as a between-subjects variable that was performed throughout the presentation and recall phases of the task. The control condition failed to replicate the key outcome from M.J. Watkins' (1977) experiment. That is, while the mixed list conditions demonstrated an intermediate level of performance, spans for High-Low and Low-High lists were the same. The suppression condition lowered the overall level of recall, and reduced the recall of HF lists more than LF lists, but the difference between mixed lists was again non-significant. Therefore, these results replicated the observations made by Kausler and Puckett (1979).

While the frequency effect was shown to reduce in size under AS, as would be expected if the blocking of rehearsal preferentially harmed the recall of items with greater articulation rates (Baddeley et al., 1975), the presence of a significant frequency effect under suppression did not fully conform to the prediction of the phonological loop. Gregg et al. (1989) reasoned that the presence of an effect must be due to either incomplete suppression of rehearsal, or because STM interacted with LTM in some way. They noted, despite Wright's (1979) argument that M.J. Watkins' (1977) data did not discount a single process explanation of STM, Wright (1979) had suggested STM performance might be governed by rehearsal processes in combination with an additional memory store that was sensitive to differences in articulation rate. If this were the case, greater efficiency in the articulation of HF than LF words might lessen the effort required to maintain items, allowing more resources to be devoted to the encoding of items in LTM.

Gregg et al. (1989) compared the delayed serial recall of pure frequency lists under control and suppression conditions to determine whether the frequency effect was sensitive to the blocking of articulatory rehearsal in a context when LTM encoding is important. Despite suppression lessening the levels of overall recall, a frequency effect was present to the same extent in the recall of lists in control and suppression conditions. Furthermore, a free recall measure produced similar results. Accordingly, Gregg et al. (1989) discounted a frequency by suppression interaction in the span task as a reflection of the greater impact of suppression on the encoding of

HF items into LTM. Instead, the authors interpreted these results to indicate that HF words are more easily rehearsed and maintained up to the point of retrieval than LF words, and this advantage is reduced by suppression.

Lastly, Gregg et al. (1989) examined whether the failure to replicate M.J. Watkins' (1977) results might be a product of differing rehearsal strategies by groups of participants performing the task. In this experiment participants were required to rehearse only the first or second halves of the mixed lists. The lists with HF words occupying the first half were better recalled than lists with LF words in the first half regardless of which list half was rehearsed; this effect, although smaller than that observed by M.J. Watkins' (1977) was consistent with it. Span improved when items in the second half of the lists were rehearsed however there was no interaction. Therefore, differences between the span experiments using mixed half lists as stimuli could not be explained by the use of rehearsal strategies that emphasised different portions of the list. Gregg et al. (1989) suggested that variations in performance could be explained by selective rehearsal of either LF (M.J. Watkins, 1977) or HF items (Experiment 2, Gregg et al., 1989; Kausler & Puckett, 1979) but this explanation implies that task execution is subject to ad hoc influences.

Unlike M.J. Watkins (1977) who saw span as a combination of a short-term process insensitive to word frequency coupled with an LTM component dependent upon it, Gregg et al. (1989) viewed span as a function of two frequency-sensitive memory stores coupled with the strategic control of rehearsal. However, as described by Baddeley and Hitch (1974) the phonological loop contained no interface to long-term knowledge. Accordingly, the central executive in the working memory model was nominated by the authors as a suitable conduit through which LTM effects could operate on memory span.

3.6 STM and phonological LTM

The LTM contribution to the frequency effect in memory span proposed by M.J. Watkins (1977) was thought to originate from differences in episodic encoding. A second line of research examined whether differences in long-term phonological representations of items might be, in part, responsible for performances in memory span tasks, and was initiated by Hulme et al. (1991). While their study examined the effect of lexicality on serial recall, that is the greater recall of lists of words than nonwords, it marked a turning point in the interpretation of LTM effects in STM, and motivated much of the research into the frequency effect for the following decade.

These authors acknowledged that much of the data regarding short-term span and manipulations of speech-based variables (e.g. Baddeley et al., 1975) was well accommodated by the articulatory loop (Baddeley & Hitch, 1974), and in particular, the observation that memory span for different materials correlated with the number of items of each material type that could be recited in just under 2 seconds (Schweickert & Boruff, 1986). However, this view of STM was premised on the assumption that memory span was not influenced by additional factors, and specifically a contribution from LTM. Despite the articulatory loop's appeal of parsimony, numerous sources of evidence suggested that STM, and memory span more precisely, were not a product of a speech-based system acting in isolation.

The failure of AS to fully eliminate a frequency effect in span (Gregg et al., 1989), not only threatened the generality of the phonological loop but also reflected the operation of and contribution from a memory store sensitive to word frequency. Hulme et al. (1991) reasoned that LTM would be an obvious candidate.

Hulme et al. (1991) argued that the nature of the contribution from LTM was based specifically on the long-term phonological representations of items, thus differentiating their position from other researchers who had proposed LTM influences (e.g. Gregg et al., 1989; M.J. Watkins, 1977). Additionally, the role of long-term representations was in the reconstruction of partially decayed items at the point of retrieval. This assistance, and therefore effects borne from it, would apply to all items in the list, not for items residing in specific list portions (e.g. primary list positions - c.f. M.J. Watkins, 1977).

This position was supported by two experiments that examined the effect of lexicality on memory span. Lexicality reflects the status of a verbal item as either a word or a nonword. While words are argued to possess long-term phonological representations that are acquired by their use, due to lack of familiarity, lexical representations for nonwords should be nonexistent. Therefore, comparison of memory span between words and nonwords should provide the limiting case of the benefit that long-term phonological knowledge provides to STM. Furthermore, if nonwords lack LTM representations, memory span for nonwords should reflect an estimate of the operation of the articulatory loop in the absence of LTM support.

Word span and speech rate measures were taken from the same participants tested on sets of words and nonwords, varying on spoken duration. Analysis of the memory span data revealed effects of the spoken length of item and lexicality, but no interaction. The speech rate data was found to contain an effect of item duration, but not an effect of lexicality and there was no interaction.

The relationships between memory span and speech rate for words and nonwords were determined by linear regression; while no difference in the slope of these relationships was found, a reliable difference in the intercepts by lexicality was established; the intercept of the speech rate-memory span function was greater for words than nonwords.

The presence of a consistent, non-zero gradient relating speech rate to memory span for both stimulus types was seen as evidence of the common operation of the articulatory loop. This function indicated that STM capacity for both words and nonwords corresponded, approximately, to the amount of material rehearsable within the critical 2-second period. However, the difference in span between words and nonwords, identified in the span data and indexed by the greater intercept for words in the speech rate-memory span functions, clearly pointed to a source for the lexicality effect that was independent of speech rate. Hulme et al. (1991) nominated phonological LTM as the source of this contribution. Following Baddeley et al. (1975), the authors proposed that memory span could be described by the equation,

$$S = c + kr \ ,$$

where memory span, *S*, is the sum of two terms, namely *c*, the intercept of the speech rate memory span function and an estimate of the LTM contribution, and a term that is the product of *k*, the change in span with the change in speech rate, with the speech rate, *r*, of the stimuli. The parameter *k* could alternatively be viewed as a measure of the decay period within the articulatory loop.

Hulme et al. (1991) applied this model to a second experiment testing the lexicality effect in memory span when nonwords were defined as unfamiliar foreign words. They measured participants' memory span of either English words or Italian words (and therefore English nonwords) before and after a 3-day training period where translations of the items were learned. It was hypothesised that span would improve across test phases for the group with Italian words as training would create long-term representations for these items. Furthermore, this difference should be observable as an increase in the intercept of the speech rate-memory span function for these stimuli, while the slope was expected to remain constant.

Conversely, the learning of Italian translations to the English items presented to the second group was not anticipated to impact span performance, as LTM representations for English words would be, by definition, established. The inclusion of the learning phase for this second group controlled for the possibility that the processing of information regarding a small set of items, rather than the creation of LTM representations *per se*, improved memory span.

Memory span for Italian words was found to be larger for short words than long words and larger after than before training. Furthermore, an interaction was present such that the improvement for short was not as great as for long words. The relationships between memory span and speech rate for items before and after training were calculated for each participant, and analysis determined that significant differences in both slopes and intercepts existed.

The corresponding analyses of the memory span data for English words showed an effect of item length, but no effect of time of testing, and no interaction. Thus, as predicted, the learning activity in between testing phases did not alter memory span. The relationship between speech rate and memory span for English words was found to be constant across tests.

A final analysis directly compared memory span and speech rate measures of English and Italian words before and after training. Span was greater for English than Italian words, and greater for short than long words. The improvement to span was greater for Italian than English words and greater for long than short Italian words. This experiment demonstrated a lexicality effect, operationalised as greater memory span for English than Italians words. The recall of English words was characterised by a relationship between speech rate and memory span with a non-zero intercept, argued to reflect a contribution from phonological LTM. Span was also shown to improve for Italian words after training, while a corresponding improvement for English words was nonexistent. The increase in span could not be attributed to an increase in the speech rate of items at the second time of testing. Accordingly, the speech rate memory span function for Italian words after training was found to contain a non-zero intercept, reflecting the newly learned LTM information that assisted the recall of items in STM. The unexpected finding of a significant change in slope of speech rate-memory span function for Italian words across testing periods was explained as an artefact of the development of phonological knowledge regarding language items. If the likelihood of short-term trace degradation was related to word length, then longer items should be more prone to loss of information in the short-term trace than shorter items. Without knowledge of their phonological structure there was no means by which items could be reconstructed for recall, leaving longer items with a marked disadvantage in this regard. Accordingly, the availability of long-term representations would be arguably more useful to long than short items, and result in a greater improvement in span performance for these stimuli.

Cumulatively, these results were viewed as supporting a two-component model of memory span, where the articulatory loop governed the maintenance of short-term phonological traces prior to recall. The loop would be capable of maintaining a limited amount of speech-based information in the form of an articulatory code. These traces would undergo passive decay and require rehearsal to retain sufficient intactness for recognition purposes, a process sensitive to the speech rate of items but not to their lexical status. This process would be complemented by the knowledge of the language structure of items at retrieval; lost information of partially degraded traces could be regained through a process of pattern completion, based on long-term phonological representations. Furthermore, Hulme et al. (1991) emphasised that the processes of rehearsal and pattern completion would apply to all items in the list.

While Hulme et al. (1991) had argued that the acquisition of long-term phonological knowledge was responsible for improvements to memory span, their study did not provide conclusive evidence for this; as the task involved the learning of translations of Italian words, it could be argued that the effect on word span was due to enhanced semantic knowledge drawn from learning the translations of Italian words. This point was addressed by Hulme, Roodenrys, Brown, and Mercer (1995), who used stimuli from the Hulme et al. (1991) experiment to test participants on memory span for words and nonwords. Word span was measured before and after a familiarisation phase that focussed on the phonological forms of items. They found that differences existed between the recall of words and nonwords after speech rate had been controlled, and that improvements for memory span after training were greater for nonwords than words. These results were seen to support Hulme et al.'s (1991) contention that the long-term phonological representations of words contribute to item recall in memory span.

The connection between STM and long-term phonological knowledge was an important theoretical development, as it linked the operation of STM to speech mechanisms thought to manage the perception and production of verbal items. This association would create an ongoing influence in the interpretation of effects from LTM variables on short-term recall and word frequency in particular (e.g. Hulme et al., 1997; Roodenrys et al., 2002).

3.7 Word frequency, word length and speech rate

Hulme et al. (1997) examined more closely the relationship between memory span and speech rate with respect to word frequency, in an analogue to the investigation by Hulme et al. (1991) into the lexicality effect (Experiment 1). High frequency words are recognised in speech at lower signal-to-noise ratios than LF words (Howes, 1957), and given the evidence for a link between STM and speech mechanisms, it was proposed that similar processes operated in the recall of items in STM. Thus, following Roodenrys et al. (1994), the basis of an LTM contribution to the frequency effect involved differences in the lexical entries of words.

This experiment sought to determine whether HF and LF words show similar relationships between speech rate and memory span across items of different spoken duration. It was predicted that memory span for HF words would be superior to span for LF words (Roodenrys et al., 1994), and that this difference would contain some component independent of any differences in speech rate between stimuli.

Word sets were manipulated on the basis of word frequency and word length. Analysis of the word span data identified significant effects of word frequency and word length, but no interaction between these factors. Span measures were found to be different for each level of word length. An equivalent analysis on the speech rate data revealed the same pattern of results. Therefore, both memory span and speech rate were influenced by the frequency of items as well as their length.

After the effects of speech rate were controlled for, HF words were again better recalled than LF words. Once more, it was argued that some non-articulatory process must be responsible for the frequency effect in span (Gregg et al., 1989; Roodenrys et al., 1994; Tehan & Humphreys, 1988). Hulme et al. (1997) identified differences in the accessibility of long-term phonological representations of the items as a reason for the superior performance of HF words. Better access to these representations would enhance a pattern completion process thought to apply to partially degraded traces at the point of recall.

The concept of a pattern completion process was referred to as redintegration, and had formerly appeared in a number of guises in the work of several researchers (Brown & Hulme, 1995; Cowan, 1992; Hulme et al., 1991; Nairne, 1990; Schweickert, 1993). Hulme et al. (1997) viewed redintegration as an automatic process, and one likely to engage mechanisms normally used for speech perception and speech production.

3.8 Word frequency and LTM variables

3.8.1 Age of acquisition (AoA)

Roodenrys et al. (1994) examined whether the effects of age of acquisition (AoA) and word frequency could be separated in memory span tasks, as these variables are highly correlated. The authors reported the results of two span tasks utilising pure lists that manipulated frequency or AoA. These demonstrated that while a frequency effect was present in memory span, AoA effects were absent. Furthermore, two naming procedures determined that the strength of manipulation of AoA was capable of producing differences in speech production. Lastly, when speech rate differences associated with the frequency of the stimuli were taken into account using speech rate measures as covariates, a frequency effect was still present.

Roodenrys et al. (1994) interpreted these results as evidence that a frequency effect, not attributable to either AoA or speech rate, existed in the short-term recall of words. This contribution was argued to result from the relative difference in accessibility of HF and LF long-term phonological representations thought to support the completion of partially decayed short-term traces in much the same way that long-term representations for words were thought to produce a short-term recall advantage over nonwords (Hulme, et al., 1991).

3.8.2 Phonological neighbourhood effects

Redintegration was argued to use mechanisms involved in speech processing (Hulme et al., 1991; Hulme et al., 1997; Roodenrys et al., 1994; Schweickert, 1993).

Schweickert (1993; Hulme et al., 1997) likened redintegrative processes to those used in speech production when speech errors are corrected. Alternatively, it was possible that redintegration might draw on speech perception mechanisms, involving processes used to recognise a partially degraded signal in noise. The correspondence of word frequency effects between memory span and identification in noise (Luce & Pisoni, 1998) was consistent with this position.

A phonological neighbour is defined as any word that differs from a target word by a single phoneme (Roodenrys et al., 2002; Vitevitch, 2002). Items with few phonological neighbours are better recognised in noise than items from large neighbourhoods, and items with LF neighbours are better recognised in noise than items with HF neighbours (Luce & Pisoni, 1998). Roodenrys et al. (2002) explored the differences in memory span for word frequency and phonological neighbourhood size, and word frequency and phonological neighbourhood frequency, in two experiments. They predicted that the direction of the relationships found with auditory perception would be replicated for memory span across all variables.

Both experiments found standard frequency effects after controlling for speech rate differences, while the effects of phonological neighbourhood size and neighbourhood frequency were eliminated or substantially reduced when speech rate was considered. Therefore, the frequency effect was seen to be essentially independent of the neighbourhood properties of words. These results however did not directly challenge the possibility that the source of the frequency effect, in terms of redintegration, was due to speech perception processes.

The effects of neighbourhood size and neighbourhood frequency were contrary to prediction. Accordingly, Roodenrys et al. (2002) acknowledged that the idea redintegration makes use of speech perception mechanisms, at least for phonological neighbourhood variables, required revision. An alternative view positioned effects from these lexical properties of words in speech production processes. This possibility was consistent with the direction of speech rate differences observed during experimentation and observed patterns in visual naming (Andrews, 1997). Roodenrys et al. (2002) proposed the retrieval of speech motor programs for the articulation of words as a potential production-related source of phonological neighbourhood effects.

3.8.3 The language-based model of R.C. Martin, Lesch and Bartha (1999)

Roodenrys et al. (2002) referred to the language-based model of R.C. Martin, Lesch and Bartha (1999) as a plausible framework to explain how phonological lexical variables may be engaged in late-stage redintegration. Language-based models have evolved primarily in response to the observed relationships between language processing and STM performance of neuropsychological patients with various forms of memory impairment (N. Martin & Saffran, 1997; R.C. Martin et al., 1999). As such, this class of model focuses on the mechanisms and capacities responsible for language processing, more specifically those relating to speech perception and speech production (Majerus, 2009).

The perspective of language-based models argues that, in effect, verbal STM is part of the psycholinguistic architecture that provides short-term storage for language-based activities, through the activation and maintenance of representations across multiple levels (phonological, lexical, semantic and syntactic) within the LTM network. This occurs through interactive activation (N. Martin & Saffran, 1997; N. Martin, 2009), with the degree of activation dependent on the connection strength of bi-directional links between representations at adjacent levels, coupled with the rate of decay of activation. The maintenance of activation results from ongoing feedforward and feedback activation to linked nodes at adjacent levels, generating a stable, mutually supporting pattern of activity. The selection of candidates for retrieval is based on the level of activation and the availability of language representations (N. Martin, 2009).

Within this approach, two prevailing views exist regarding the degree to which verbal STM is separate to the LTM knowledge store. A unitary view holds that STM comprises that subset of LTM (the language processor) that is currently activated (N. Martin & Saffran, 1997), while a second view proposes separate STM storage in the form of buffers in addition to activated LTM representations (R.C. Martin et al., 1999; Romani, McAlpine, & Martin, 2008). In the latter case, buffers record representations of items in a sequence at encoding, and integrate their contents with representations in the LTM knowledge store via the action of feedback, continually updating representations that feedback from LTM in return. Traditionally, these models have considered decay as a mechanism that can alter patterns of activation over time (N. Martin & Saffran, 1997; R.C. Martin et al., 1999; Romani et al., 2008) either in the activations of LTM representations themselves (N. Martin & Saffran, 1997) or within the STM buffers (R.C. Martin et al., 1999). In some cases STM deficits have been attributed to accelerated decay rates associated with either phonological or semantic processing (e.g. N. Martin & Saffran, 1997).

A depiction of the structure of the model proposed by R.C. Martin et al. (1999) is presented in Figure 3.1. Separate input and output phonological buffers are posited on the basis of neuropsychological evidence suggesting dissociation between speech production and speech perception mechanisms (e.g. patient MS, R.C. Martin et al., 1999). The separation of buffers for phonological and lexical-semantic information has similarly been motivated by dissociated performances of patients who selectively exhibit phonological or lexical-semantic deficits in language processing and short-term memory tasks (Martin & Lesch, 1996). Maintenance of items is achieved through the continued activation of information in the short-term buffers by the long-term knowledge structure and in turn, feedback from the buffers to the long-term store. All levels in the long-term knowledge store continue to provide activation feedback and feedforward after item presentation, and as a consequence, the contents of the buffers will reflect interactions between the levels within LTM.

3.8.3.1 A role for interference in the language processor?

Language-based models have assumed that short-term recall essentially reflects the activation patterns present in the language processor as a function of interactive activation and decay, with impairment primarily a reflection of accelerated decay rates (N. Martin & Saffran, 1997; R.C. Martin et al., 1999). However, more recently researchers have argued that some forms of STM deficit can be explained in terms of interference acting on the language network (e.g. Hamilton & Martin, 2005; 2007; R.C. Martin & Hamilton, 2007). According to this position, the inability to inhibit previously presented items leads to high levels of proactive interference and diminished performance in STM tasks. Bearing in mind the difficulties associated with the generalisation based on neuropsychological cases, this research poses that inhibition, in terms of a capacity to delete ‘just-used’ information, might be an important process in memory function over the short-term.

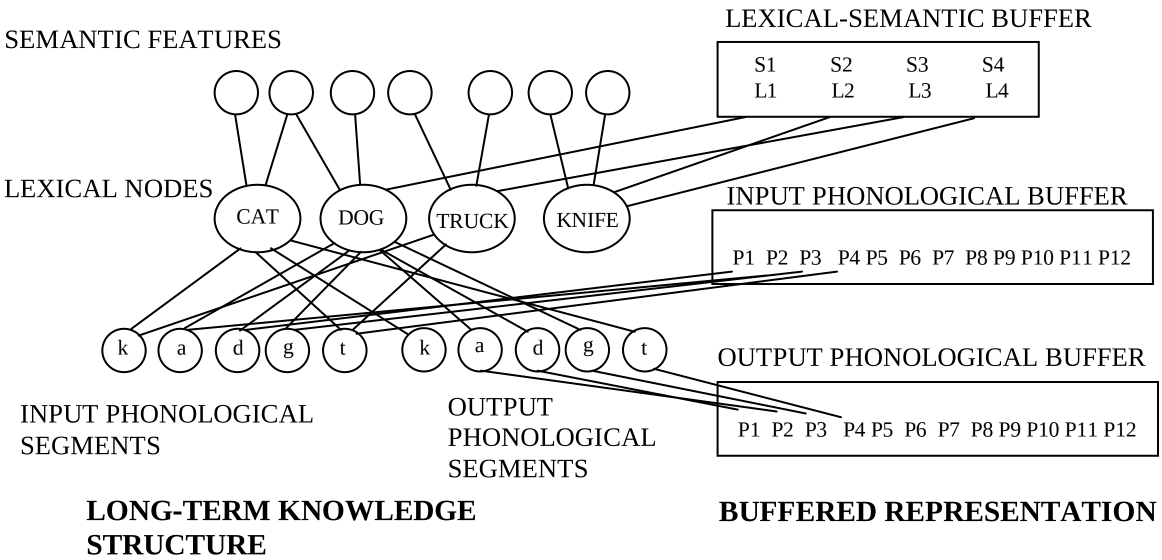


Figure 3.1 A language-based model of STM with lexical–semantic, phonological input and phonological output buffers. Phonological buffers exist for both speech perception and speech production pathways. Reprinted from *Journal of Memory and Language*, 41, R.C. Martin, M. F. Lesch, and M.C. Bartha, “Independence of input and output phonology in word processing and short-term memory”, p. 6, Copyright (1999), with permission from Elsevier.

3.8.4 Redintegration and LTM language representations

Although language-based models of memory can be conceived as stand-alone explanations of short-term memory phenomena, Roodenrys et al. (2002) sought to incorporate the model of R.C. Martin (1999) within the process of late-stage redintegration. Specifically, redintegration was proposed to be activated by the entry of degraded phonological information to the long-term knowledge system via the input phonological buffer. However, while the results of this study would suggest that the locus of phonological neighbourhood effect occurred in the output or production side of the model, possibly in terms of feedback from the output buffer to the long-term output phonological units, the authors did not specify the mechanisms or locus of the frequency effect as the data did not provide clear indications of its origin.

3.9 Summary

The evidence from memory span studies indicated that the frequency effect for pure lists was a robust and replicable one (Gregg et al., 1989; Kausler & Puckett, 1979; Tehan & Humphreys, 1988; M.J. Watkins, 1977). However, the results for mixed frequency lists, as defined by M.J. Watkins' (1977) are problematic; two instances reported the presence of an effect, advantaging lists beginning with HF items (Experiment 4, Gregg et al., 1989; M.J. Watkins, 1977), but two further experiments produced null effects (Experiment 2, Gregg et al., 1989; Kausler & Puckett, 1979). Furthermore, no systematic variation in method or procedure was identified to resolve the variability in these outcomes (Gregg et al., 1989).

M.J. Watkins (1977) had originally interpreted the effect found with mixed lists as evidence for multiple memory processes operating on memory span; differences in the frequency with which words are encountered in language influenced the level of episodic encoding of items in LTM, and this contributed to the success of recall. Furthermore, this effect operated for early rather than late lists items as indicated by free recall and serial recall studies (e.g. Sumby, 1963; O.C. Watkins & Watkins, 1977).

However, the interpretation of the frequency effect in STM was complicated by a possible confound between word frequency and articulation duration (Geffen & Luscz, 1983; Gregg et al., 1989; Tehan & Humphreys, 1988; Wright, 1979). High frequency words are articulated faster than LF words of the same length. A STM account emphasising differences in rehearsal rates, the phonological loop (Baddeley & Hitch, 1974), was promoted as a single, short-term process capable of linking differences in the articulation of stimuli to differences in span; relative to LF words, greater rehearsal of HF items would lead to greater retention of short-term traces, and better memory span (Wright, 1979). Despite this difference, the presence of reliable frequency effects under conditions of AS (Gregg et al., 1989; Tehan & Humphreys, 1988) pointed to a second process or mechanism sensitive to the familiarity of words that contributed to short-term recall. Tehan and Humphreys (1988) nominated the central executive in the working memory model as the additional contributor to memory span, while Gregg et al. (1989) saw the central executive as a pathway to episodic LTM. In addition, Gregg et al. (1989) thought that the strategic control of rehearsal was needed in order to explain performance patterns and inconsistencies found on mixed lists.

Memory span studies beyond this point adopted the redintegration framework and tested pure frequency lists (Hulme et al, 1997; Roodenrys et al., 1994; Roodenrys et al., 2002). The frequency effect was found to exist after differences in speech rate were accounted for, and was generally shown to be independent of other variables reflecting lexical properties of language items, namely AoA (Roodenrys et al., 1994) and phonological neighbourhood size (Roodenrys et al., 2002).The frequency effect in memory span was also found to be independent of word length (Hulme et al., 1997), however frequency and word class interacted such that differences between word classes existed for LF but not HF words (Tehan & Humphreys, 1988).

Initially, redintegration was seen as a reconstructive process that recruited speech-based mechanisms in the retrieval of items (Hulme et al., 1991). Schweickert (1993; Hulme et al., 1997) nominated a potential locus of redintegration, proposing its operation to be analogous with the correction of a speech error, and therefore an effect of speech production. While Roodenrys et al. (2002) found the frequency effect in memory span to be consistent with the frequency effect in identification in noise (Luce & Pisoni, 1998), they demonstrated that the span effects for phonological neighbourhood variables did not correspond to performances on perception tasks, and suggested instead that speech production mechanisms might be for the influence of neighbourhood factors in short-term recall.

Chapter 4

Word frequency and the serial recall of pure supraspan lists

4.1 Introduction

With the exception of O.C. Watkins and Watkins (1977), most research on the frequency effect involving the serial recall of supraspan lists has been conducted within the last 15 years. The results for pure lists are summarized in this chapter. The majority of these studies have presented evidence suggesting the frequency effect to be well behaved and understandable in terms of the redintegration/reconstruction hypotheses where successful retrieval is dependent on individual item properties (Hulme et al., 1997; Poirier & Saint-Aubin, 1996).

4.2 Word frequency and the modality effect

The modality effect is the superior recall of the last few items in serial recall when presentation is auditory rather than visual (Penney, 1989; O.C. Watkins & Watkins, 1977). O.C. Watkins and Watkins (1977) examined the behaviour of the word frequency effect in serial recall across presentation modalities.

Serial position was grouped into primacy and recency segments and analysed according to this segmentation. The analysis identified better recall in HF than LF lists, better recall for auditory than visual presentation, and better recall for primacy than recency positions of the serial recall curve. An interaction between position and modality revealed that recall with auditory presentation was greater than with visual presentation for the recency portion of the curve; this is the typical modality effect. Frequency also interacted with serial position such that a frequency effect was present for the primacy but not recency positions in serial recall. Furthermore, there was no modality by frequency interaction, as the frequency effect for both auditory and visual presentations was similar across serial positions. The three-way interaction between modality, frequency and position was also non-significant.

The presence of an interaction between word frequency and serial position was interpreted as a reflection of either a complex process or the superposition of two simple processes. O.C. Watkins and Watkins (1977) considered the STM-LTM distinction found in explanations of free recall performance as a potential account. That is, the items in the primacy portion of the curve would be encoded into ‘durable’, that is episodic, LTM that was sensitive to the word frequency of items, while the recency items would be recalled from a capacity-limited store that was not affected by frequency or other lexical variables. More generally, O.C. Watkins and Watkins (1977) did not consider the proposition that serial recall performance was due entirely to STM or a single process a tenable one.

4.3 Word frequency, and item and order information

Poirier and Saint-Aubin (1996) explored the nature of the frequency effect in relation to the two components of information important in serial recall, namely item identity and item sequence. Their objective was to determine whether word frequency influenced item memory, order memory or both.

Reconstruction theories (e.g. Hulme et al., 1991; Roodenrys et al., 1994; Schweickert, 1993) implied that word frequency should be an item memory effect, as the effect was derived from differences in the access to LTM representations after the selection of a short-term trace. Therefore such accounts assumed that the order of the partially degraded traces, on which reconstruction was based, was determined by some other mechanism.

Furthermore, alternative accounts of memory outlined a role for word frequency in the encoding of order information. These proposed that HF words created stronger inter-item associations than LF words (Deese, 1960; Sumby, 1963; Whiteman, Nairne, & Serra, 1994). Furthermore, the Theory of Distributed Associative Memory (TODAM, Lewandowsky & Murdock, 1989) modelled the frequency effect in serial recall by assuming better inter-item associations between items in HF lists.

Poirier and Saint-Aubin (1996) used open stimulus sets to emphasise item memory in testing. This ensured that item familiarity could not inadvertently change the relative emphasis of the experimental task to one focusing on the memory of the order of items. However, given order memory was also of interest, the authors included phonological similarity, an effect acknowledged to disrupt the order of recalled items (Conrad, 1964; Conrad & Hull, 1964), as a manipulation in the task to test its sensitivity to order memory.

Additionally, because phonological similarity was known to impact order memory, Poirier and Saint-Aubin (1996) argued that a difference in recall for similarity conditions should be evident for the strict but not lenient scoring of the serial recall data. An effect of word frequency however, was expected to be present in both measures, as item recall is implicated in both strict and lenient scoring.

Strict scoring revealed effects of similarity, frequency and serial position. Across frequency conditions, recall performance was ranked by frequency level; HF words were better recalled than medium frequency words that were in turn better recalled than LF words. A frequency by serial position interaction revealed a frequency effect for all positions but the last. The significance of the frequency by phonological similarity interaction was not reported and is therefore assumed to have been non-significant.

4.3.1 Item analysis

Lenient scoring produced an effect of frequency, but not of phonological similarity. Item memory performances on all levels of frequency were different from each other and in the rank order found for strict scoring.

To test for differences in order memory Poirier and Saint-Aubin (1996) used conditionalised order error rates, to avoid confounding differences in item memory with differences in order memory (Murdock, 1976). This data identified that the number of order errors was influenced by the phonological similarity of items, but not word frequency, and no interaction between these effects was present.

The results supported accounts proposing that the frequency effect was a result of a late-stage deblurring process, arising after short-term trace selection (Hulme et al., 1991; Poirier & Saint-Aubin, 1995; Roodenrys et al., 1994; Schweickert, 1993), and refuted the likelihood that frequency was involved in the differential encoding of order memory. These outcomes were consistent with positions arguing that phonological short-term traces were recovered from a short-term store and subjected to reconstruction based on long-term knowledge. Greater accessibility of HF than LF words in LTM, on which reconstruction was based, would facilitate the frequency effect. The effect of phonological similarity in contrast, acted on the degree of confusability of degraded traces in the short-term store, reflecting the dependence on a phonological code (Baddeley, 1986).

4.3.2 The reconstruction hypothesis

Poirier and Saint-Aubin (1996) extended this explanation to address differences in order memory across phonological similarity conditions. They contended that if degraded short-term traces were considered as retrieval cues, submitted to a reconstruction process that sought access to a candidate in LTM consistent with cue information, then the effects of phonological similarity (and frequency) could be more fully described. In the case where the cue is not unique, access to the wrong item in LTM might be attempted, and then accepted because of other factors, for example familiarity due to the presentation of the incorrect item in the same list. This is what would be expected to occur in the case of phonological similarity, where the loss of phonological features in the short-term trace could lead to a retrieval cue consistent with multiple items presented in that list. Word frequency, assumed to reflect the accessibility of items from LTM given a retrieval cue, would therefore influence probability of retrieval; HF items would be more likely to be successfully recalled than LF items. Additionally, Poirier and Saint-Aubin (1996) thought that familiarity judgements would also apply to the identification of items in lists varying in frequency. This account was to be later referred to as the retrieval-based hypothesis (Saint-Aubin & Poirier, 2000).

4.4 The word frequency by serial position interaction

Hulme et al. (1997) had interpreted the frequency effect in memory span as reflecting a late-stage reconstruction process that made use of whatever short-term phonological information had survived at the point of retrieval, and was influenced by the relative accessibility in phonological long-term representations of words. This process, and therefore the frequency effect, would occur for all serial positions.

4.4.1 The multinominal processing tree (MPT) model of Schweickert (1993)

Hulme et al. (1997) described Schweickert's (1993) multinomial processing tree (MPT) model of ISR as an example of a two-stage model providing quantitative predictions in serial recall (see Figure 4.1). According to the MPT, recall is a result of up to two mental processes sequenced in series. Each process is associated with a given likelihood that it will be successful. The outcomes of the MPT model are mapped onto the terminal nodes of the tree and the probability of each outcome is given as the product of the probabilities of the pathway leading to it.

Please see print copy for images

Figure 4.1 The multinomial processing tree model of Schweickert (1993). Adapted by Hulme et al. (1997) from "A multinomial processing tree model for degradation and redintegration in immediate recall", by R. Schweickert, 1993, *Memory, and Cognition*, *21*, p. 169. Copyright 1993 by Psychonomic Society, Inc. Reprinted with permission.

The first process in the MPT model involves the access of a short-term trace from the short-term store. If this trace is intact, with probability *I*, then it can be recalled directly from the short-term store without reference to a second process. If however, the trace is degraded, it cannot be recalled directly from STM and will require further processing in the form of redintegration in order for recall to be successful. The likelihood of this event is 1 - *I*. In the second stage, each frequency type *j*, has a probability that it will be successfully redintegrated, R_j. Therefore the probability, P_{ij}, that recall will be successful is the sum of the two possibilities leading to correct recall, namely

$$P_{ij} = I_{ij} + (1 - I)R_j \quad .$$

I_{ij} is the probability of intactness for a frequency type *j* and decreases with serial position *i*. This trend reflects the increased degradation of item traces with increasing serial position.

Better recall for early than late list items was assumed to arise because they would be more likely to be recovered directly from the short-term store. The extent to which items into the list were directly retrieved would be dependent on the speech rate of the material concerned. The degradation of item traces was assumed to be a function of elapsed time since list presentation, and lists of items with longer spoken duration, and therefore recall, would lead to a greater decrease in the probability of intactness across serial positions. As LF words have longer spoken duration than HF words (Wright, 1979), Hulme et al. (1997) argued that the decrease in the I_{ij} for HF lists across serial position should be less than the corresponding decrease for LF lists, contributing to the presence of a frequency by serial position interaction. Hulme et al. (1997) also inferred that the process of redintegration made an additional contribution to a frequency by position interaction, as later list items would be more likely to require reconstruction on account of increased degradation of the short-term traces. Accordingly, they anticipated that word frequency would be more important for later items in the list, and the frequency effect would be seen to increase across serial positions. This prediction stood in opposition to the form of frequency by serial position interaction observed by O.C. Watkins and Watkins (1977).

As a further consequence of the increasing importance of redintegration across list items, the MPT was argued to predict differences in the levels of omissions and intrusions in the recall of HF and LF lists. Redintegration for LF words would be more difficult than for HF words, and therefore it would be expected that more of these errors would occur in LF than HF lists. The MPT however, was not capable of providing any indications as to the nature of order errors across conditions.

4.4.2 Hulme et al. (1997) Experiment 2

Hulme et al. (1997) examined the frequency effect in serial recall. They identified that effects were present after controlling for differences in speech rate of HF and LF words. The serial recall data revealed effects of frequency and serial position, and a frequency by serial position interaction that manifested in the form predicted by the MPT.

4.4.2.1 Item analysis

There was no difference in order errors between frequency conditions, however these error rates had not been conditionalised (see Murdock, 1976; Poirier & Saint-Aubin, 1996). Omissions and intrusions occurred more often for the LF than HF lists. Lastly, HF lists did not induce extralist intrusions while the extralist intrusions for LF lists were often HF phonological neighbours (see Roodenrys et al., 2002).

The frequency effect in serial recall had been demonstrated in this experiment, and was shown to be independent of speech rate. Schweickert's (1993) MPT model was seen as a suitable working model for the effect in serial recall because it could predict the observed frequency by serial position interaction. This was based on the premise that the capability of redintegration to restore item identity was more important for later list items, and as short-term traces become more degraded, redintegration would be less efficient for LF than HF items, reflecting the relative accessibility of phonological LTM representations.

4.5 Word frequency and word length

Hulme et al. (1997) extended these findings in another serial recall task that included two levels of word length. After controlling for speech rate differences they found effects of frequency and word length, indicating that portions of these effects were not reliant on processes sensitive to the speed of articulation. Frequency and word length interacted with serial position independently, but there was no frequency by word length interaction, and no three-way interaction with serial position. Furthermore, the form of the frequency by serial position interaction was repeated across two levels of word length and demonstrated that the attenuated frequency effect for early serial positions was not due to a ceiling effect for short words.

4.6 Word frequency and backwards recall

4.6.1 Hulme et al. (1997) Experiment 4

Hulme et al., (1997) explored whether the increase in frequency effect across serial positions was due a 'range effect', namely an effect that increases with poorer performance. The method of backward recall was chosen to test this possibility as it had been shown to reduce overall STM performance (Cowan et al., 1992).

The results discounted a range effect explanation as there was no frequency effect for backwards recall, while the forward recall condition reproduced the pattern established by the earlier experiments. Furthermore, the absence of an effect for backwards recall suggested that the speech rate advantage for HF words is not translated into superior recall (see Cowan et al., 1992, for a contrast with word length).

The authors analysed the size of the frequency effect by trial in the forward recall data to examine whether the use of restricted item pools may have diminished the frequency effect as the experiment progressed. A reduction in effect would not be predicted by the redintegration explanation of recall, as the accessibility of long-term knowledge would not be altered to any meaningful extent during the course of the experiment. The results of this analysis identified that no reliable change of the frequency effect across trials in forward recall was evident. However, it should also be noted that this data had been collected from a procedure that included blocked backwards recall conditions counterbalanced with the blocks for forward recall. Therefore, half of the participants had already experienced repeated exposures to the stimuli prior to being tested on the forwards recall trials. Failure to separate the data by counterbalancing condition may have masked an interaction revealing a change of frequency effect across forward recall trials for participants who were initially allocated the forward recall condition, coupled with no difference in frequency effect across forward recall trials for participants who had been familiarised with the stimuli during the backwards recall trials.

4.6.2 The redintegration hypothesis for word frequency in serial recall

Across a number of experiments, Hulme et al. (1997) had demonstrated a frequency effect that was independent of speech rate. In addition, a frequency by serial position interaction consistent with predictions based on the MPT model of Schweickert (1993), and contrary to the prediction of O.C. Watkins and Watkins (1977) was repeatedly observed. The authors endorsed a two-stage model of serial recall in STM, where the retrieval of partially decayed traces from the short-term store was complemented by late-stage redintegration. This reconstruction process was sensitive to word frequency because there is better access to, or greater specification of, the phonological representations of HF than LF words.

In light of the preceding research, and following Schweickert (1993), Hulme et al. (1997) proposed that two parallel processes were responsible for the restoration of a degraded item in redintegration. These processes were similar to those involved in the correction of speech errors in speech production. The first process transforms a degraded phonological short-term trace into a word and thus takes a noisy signal and cleans it up, while the second converts the representation into a string of phonemes for output. The efficiency of these processes between conditions would determine any recall advantages observed.

4.7 Word frequency and set size

Word frequency had been established as a variable influencing serial recall in terms of a direct contribution from LTM, namely differences in the accessibility of phonological representations (Hulme et al., 1997). Roodenrys and Quinlan (2000) argued that a second LTM influence, though an indirect one, might result from the number of stimuli used in testing. In experiments where items from the same restricted pool formed the composition of many lists for a given condition (e.g. Hulme et al., 1997), the availability of LTM representations was likely to be enhanced through repeated exposure, and therefore improve the level of recall (Poirier & Saint-Aubin, 1996). Furthermore, a closed set of items might mask the true extent of a given effect because performance on the weaker condition would be offset by an improvement to recall due to the greater familiarity of items. However, in cases where open sets of stimuli were used to construct lists (e.g. Poirer & Saint-Aubin, 1996), increased familiarity would be avoided as different items would be used for each list.

Accordingly, it would be expected that the use of closed pools provide a relative recall advantage, and the small number of studies using open and closed sets have determined this (see for example La Pointe & Engle, 1990; Coltheart, 1993). Specifically, the increased familiarity of items lessened the number of extra-list intrusions and omissions by participants (see Coltheart, 1993). Roodenrys and Quinlan (2000) argued that if these outcomes were viewed as an effect of a restricted search set at recall, resulting from the knowledge that the same items appear in successive lists, then the effect of stimulus set size could be interpreted as an LTM influence.

The authors argued that the redintegration hypothesis of the frequency effect, involving the greater accessibility of LTM representations of HF than LF words (Hulme et al., 1997), also provided a potential explanation of the set size effect. Roodenrys and Quinlan (2000) saw redintegration as a competitive process in pattern completion; given a word fragment from a degraded short-term trace, long-term lexical candidates matching the partial information might compete for selection at output on the basis of accessibility. As HF items were considered to have more accessible representations than LF items, HF words would be naturally advantaged in this process. However, awareness of the contents of a stimulus set could limit the selection of potential output candidates underpinning redintegration. This in turn would lead to the greater likelihood of a correct response, and form a critical contribution in conditions where recall would otherwise suffer. In terms of word frequency, this would equate to a set size effect that is greater for LF than HF words. Alternatively, Roodenrys and Quinlan (2000) reasoned that if the nature of a set size effect was purely episodic, and consequently independent of short-term phonological processing, a constant effect size would be expected across all positions in the list and for all conditions examined.

These predictions were tested in two experiments. Experiment 1 tested the recall of HF and LF words with open and closed sets of stimuli. The standard frequency effect was found and recall was greater for closed than open sets. All two-way interactions were significant. The set size effect was smaller for HF than LF words, and specifically was significant for LF but not HF words. The other two-way interactions were interpreted in terms of the significant three-way interaction; the set size effect for LF words became larger across serial positions, but did not change for HF words.

However, in this experiment, the level of recall for HF lists was high, and possibly subject to a ceiling effect. Roodenrys and Quinlan (2000) ran a second experiment using longer words, to increase task difficulty and reduce the overall level of performance.

The frequency and set size effects were replicated. However, while there were interactions between frequency and set size, and frequency and serial position, the interaction between set size and serial position, and the three-way interaction was not significant. The effect of set size was found to be smaller for HF than LF items, and once more the set size effect for HF words was non-significant.

The results of both experiments cast doubt over the possibility that set size effects were the outcome of an episodic memory contribution to recall. The effect was clearly selective, not a general effect as would be expected if it were derived from a process that was insensitive to the contents of STM traces. The results were consistent with an account of frequency and set size as LTM variables influencing the operation of a second stage contribution to short-term recall. Redintegration was assumed to operate on the basis of competition between long-term phonological representations that contained matching information with degraded short-term traces. Roodenrys and Quinlan (2000) suggested that set size placed further constraints on this competitive process. It was argued that the repeated presentation of items would increase the activation levels of the phonological representations of those items, in effect priming them, and increase their competitive position in the redintegrative process. If the frequency effect was explained in terms of greater activation of LTM phonological representations for HF than LF words, reflecting a competitive advantage for HF words, an additional set size effect would be more likely to impact the activations of LF representations and lead to a greater improvement in the recall of LF than HF words.

4.8 Word frequency and age

Majerus and Van der Linden (2003) performed a replication study to determine whether the effects long-term lexical-semantic variables in serial recall, that is, lexicality, phonotactic frequency, word frequency and word imageability (or concreteness) would generalise across age groups. They compared the serial recall performance across five groups (6 years, 8 years, 10 years, 13-16 years, and 20-22 years) and analysed the data for each LTM variable separately. The analysis for both the strict and lenient scoring of the frequency condition identified that the frequency effect was invariant with age.

The authors argued that the results of word frequency, along with concreteness, lexicality and phonotactic frequency, demonstrated a developmental insensitivity of the role of the lexical-semantic properties of words and phonological language knowledge to short-term recall performance. Accordingly, redintegration (Hulme et al., 1997; Schweickert, 1993) could be considered functionally equivalent in adults and children.

4.9 Word frequency and phonological neighbourhood

4.9.1 Allen and Hulme (2006) Experiment 2

Allen and Hulme (2006) examined the performance of phonological neighbourhood size and word frequency in serial recall, speech perception tasks (auditory lexical decision speed, word identification in noise), and speech production tasks (definition naming, delayed repetition, and maximal articulation rate). Roodenrys et al. (2002) had shown that the phonological neighbourhood size effect in serial recall did not correspond to the effect in auditory perception (Luce & Pisoni, 1998). Allen and Hulme (2006) reviewed performance on stimuli varying in phonological neighbourhood size across speech processing tasks and found that phonological neighbourhood size consistently showed inhibitory effects in speech perception tasks (e.g., identification in noise - Luce & Pisoni, 1998; auditory naming - Vitevitch & Luce, 1998, 1999), that contrasted with facilitative effects in word production tasks (e.g. induced tip-of-the-tongue states - Harley & Bown, 1998; induced speech errors, induced tongue twisters, and object-naming, Vitevitch, 2002[1]).

Based on these patterns, Allen and Hulme (2006) predicted that serial recall might be more closely related to performances on speech production than speech perception tasks. They used the stimuli of Roodenrys et al. (2002, Experiment 1), a factorial manipulation of CVC words varying on phonological neighbourhood size and word frequency.

Participants performed six tasks (including speech perception tasks, speech production tasks and serial recall). Analysis of the serial recall data replicated the reported effects of frequency and phonological neighbourhood size (Hulme et al., 1997; Roodenrys et al., 2002). However, in this case the frequency by serial position interaction was not present, and frequency did not interact with phonological neighbourhood size.

[1] This paper also reported facilitative effects of word frequency on speech production when speech errors were induced.

4.9.1.1 Item analysis

The differences in order error rates were found to be significant for phonological neighbourhood size, but not for frequency, and the interaction between frequency and phonological neighbourhood size was not significant. More order errors occurred in small than large neighbourhood lists. Item error rates were shown to be dependent on the frequency and phonological neighbourhood size of items, as well as the interaction of these factors. Fewer item errors occurred for HF and large neighbourhood words, but the phonological neighbourhood size effect was greater for HF than LF conditions. Analysis of the levels of omissions across list sets found effects of frequency and neighbourhood size, but no interaction; LF and small neighbourhood words were more likely to be omitted in recall. Allen and Hulme (2006) also reported a greater tendency to incorrectly recall phonologically related items in the recall of lists of LF words, and this was especially so when words came from large phonological neighbourhoods (Roodenrys et al., 2002).

The authors investigated the relationships between serial recall and the speech perception and speech production tasks by running a series of repeated measures regressions (Lorch & Myers, 1990). Word frequency was found to influence performance in the perception tasks and two out of three production tasks in a consistent fashion; HF words facilitated faster and more accurate processing of speech input and speech output than did LF words. The exception to this pattern occurred with delayed repetition. This task was shown to be insensitive to word frequency and suggested that any advantage in the speed of processing in speech production to HF words occurs prior to an output buffer.

The results of the regression analyses identified the speech processing measures of definition naming accuracy and word identification in noise as predictors of serial recall performance. The second of these was a perception measure, and although a weak predictor it was shown to relate to serial recall in the direction consistent with effects of word frequency in the processing of speech input (and this contrasted with phonological neighbourhood size that produced opposing effects in speech perception and speech production). That is, better serial recall performance of HF than LF items was associated with earlier identification of HF than LF items in noise. Allen and Hulme (2006) noted that as a consequence, serial recall may be weakly reliant on speech perception processes. However, a positive association between serial recall and definition naming accuracy was also found. The direction of the effects of frequency and phonological neighbourhood size on this production measure were consistent with the effects observed in serial recall, and implied that the likelihood of producing the target item from a semantic representation is dependent on these stimulus properties.

The failure to observe relationships between the other speech production tasks (delayed repetition and maximal articulation rate) and serial recall was discussed by Allen and Hume (2006). These measures were considered to be estimates of articulation in the absence of either perceptual or semantic contributions of items. Consequently they concluded that articulation measures of this kind could not explain the observed differences in recall and were independent from the measures derived from the definition-naming task. This stands in contrast to many studies that have found associations between maximal articulation rate and recall performance (e.g. Hulme et al., 1997). Allen and Hulme (2006) suggested that differences in analysis (in this case the assessment of within-subjects variance in comparison to studies examining between-subjects variance), and the control over word length (and item complexity) through the use of monosyllabic stimuli in these experiments may have attenuated effects present in other studies. However, they qualified their position by referring to the work of Murray and Jones (2002) who found an effect of co-articulation consistent with recall performance; these authors argued that articulatory complexity at the boundaries of adjacent items in a list influenced how well items were co-articulated and recalled.

The relationship between definition naming and serial recall for concreteness (from a previous experiment), phonological neighbourhood size and frequency suggested that all of these variables contribute to how well semantic representations promote the accurate selection of output representations and in turn facilitate speech production. Allen and Hulme (2006) noted that the effect of picture naming, considered to be a task similar to the one conducted here, had been explained and modeled in terms of the interactions that take place at the interface between output phonological representations and the speech production system (e.g. Vitevitch, 2002). Similar explanations could therefore be adopted for definition naming and serial recall.

4.9.2 The psycholinguistic model of N. Martin and Saffran (1997)

The results of Allen and Hulme (2006) were argued to further inform the character of psycholinguistic models that assume serial recall is reliant on speech perception and speech production processes. They referred to the model of N. Martin and Saffran (1997) where the short-term maintenance and retrieval of items is dependent on the temporary activation of LTM representations of items at multiple levels in the language processor (refer to Figure 3.1 for an equivalent LTM structure). Serial order is maintained by connections between representations in each of these layers and a sequence placeholder, that are established as a string of words is processed (N. Martin & Saffran, 1997).

This model was developed from the interactive activation features of Dell and O'Seaghdha's (1992) speech production model and includes separate speech input and output pathways (N. Martin & Saffran, 1992, 1997). The processing of auditory input involves activation of the phonological representations while spread of activation then engages the lexical and semantic representations of items. In turn, this activation provokes feedback activation to each of the preceding levels in the system. Perpetuation of this activity maintains item information until retrieval. In contrast, the output pathway commences with the activation of the semantic features of a word that feed forward to the lexical and phonological representations. Feedback from these layers produces a cyclic activation pattern that, when stabilised, will select a candidate for phonological encoding.

Allen and Hulme (2006) considered this model to be compatible with the findings presented in their study; the model caters for separate speech input and speech output processes in terms of dedicated pathways, while observed effects in serial recall could be accounted for by differences in the activations of output phonological and semantic representations and the connections between them. Importantly, this interpretation abandoned the concept of LTM effects occurring at a late-stage and as part of the redintegration process. In contrast, the position of language-based models is that the LTM representations associated with an item are activated at encoding and are supported by the ongoing feedforward and feedback of activation within the LTM system throughout the retention period and up to the point of recall (R.C. Martin, et al., 1999; N. Martin & Saffran, 1997). Therefore, Allen and Hulme (2006) elected to replace the two-component view of STM (Hulme et al., 1997; Schweickert, 1993) with an approach emphasising, at least, the item memory component of STM as the activated portion of the language processor.

4.10 Summary and conclusion

As with memory span for pure lists, the nature of the frequency effect in the serial recall of pure supraspan lists was found to be unequivocal; in this context HF words are better recalled than LF words (Allen & Hulme, 2006; Hulme et al., 1997; Majerus & Van der Linden, 2003; Poirier & Saint-Aubin, 1996; Roodenrys & Quinlan, 2000; O.C. Watkins & Watkins, 1977). The effect was shown to be invariant across age groups from young children to early adults (Majerus & Van der Linden, 2003) and was not affected by the modality of presentation (O.C. Watkins & Watkins, 1977). Variables that directly affected the phonology of words, that is word length and phonological similarity, did not interact with frequency (Hulme et al., 1997; Poirier & Saint-Aubin, 1996). Set-size, argued to be an LTM variable, determined the size of frequency effect as open word pools produced larger differences between HF and LF lists (Roodenrys & Quinlan, 2000), however, phonological neighbourhood size, also an LTM variable, did not (Allen & Hulme, 2006).

Conflicting forms of the frequency by serial position interaction were observed. O.C. Watkins and Watkins (1977) found that the frequency effect was absent for the final serial positions while Hulme et al. (1997) determined that the effect was present and increasing in size for all positions but the last. Respectively, this effect was argued to reflect differences in episodic encoding as a function of frequency (O.C. Watkins & Watkins, 1977), or resulted from a redintegrative process that would be more successful for HF than LF words, and more effective for HF than LF words as the degradation of short-term traces increased (Hulme et al., 1997; Schweickert, 1993).

Item analysis, when performed, suggested that frequency impacted the degree to which item identity was preserved (Allen & Hulme, 2006; Hulme et al., 1997; Poirier & Saint-Aubin, 1996). More omissions and intrusions occurred for LF than HF items. In cases where order error rates were conditionalised (Allen & Hulme, 2006; Poirier & Saint-Aubin, 1996), the results suggested that frequency does not contribute to memory for order. Within these constraints, the frequency effect was conceived as a late-stage influence that manifested after the point of short-term trace selection

(Hulme et al., 1997; Poirier & Saint-Aubin, 1996), although Allen and Hulme (2006) considered the possibility that frequency may play a role in recall at earlier stages; frequency could contribute to recall performance in terms of differences in speech input processing. Furthermore, their selection of a psycholinguistic model (N. Martin & Saffran, 1997) to explain speech processing and the retention of items in STM placed the involvement of LTM variables much earlier than variants of the redintegration hypothesis (Hulme et al., 1997; Poirier & Saint-Aubin, 1996).

Chapter 5

The Frequency Effect in the Serial Recall of Lists: The Case for Inter-item Associativity

5.1 Introduction

Based on the study of serial recall using pure lists, the frequency effect had been argued to reflect an item-specific contribution to recall (e.g. Roodenrys et al., 1994; Hulme et al., 1997; Poirier & Saint-Aubin, 1996), where the likelihood of an item's recall was determined by its frequency alone. However, it was also possible that lists constructed from words of the same frequency had confounded item-specific with list-level contributions, should either (or both) exist. Experiments involving the use of mixed lists would better specify the nature of the frequency effect with respect to these possibilities (Hulme et al., 2003).

The studies presented in this chapter challenged an item-specific interpretation of the frequency effect. In these experiments, HF words were not always better recalled than LF words, suggesting that the mechanisms involved in recall included influences drawn from the list composition. The source of this wider influence was thought to derive from associations between presented items, as a function of natural language use and was argued to be a key factor in recall performance.

5.2 Co-occurrence/familiarisation effects on word frequency

5.2.1 Familiarisation by pairwise association

Stuart and Hulme (2000) investigated whether the manipulation of pre-experimental association between HF and LF items might affect recall performance. Their study did not 'mix' HF and LF items within the same list, however they compared recall between lists of items that had been familiarised by pair-wise association during a training period, with lists of items familiarised separately. It was hypothesised that the creation of inter-item links by pair-wise familiarisation would facilitate better recall, and that the effects would be greater for LF words, given on average, HF words already share higher inter-item association (Deese, 1960). Should the results conform to these predictions, the experiment would offer evidence that the short-term recall of words depended, to some degree, on the extent to which words in the list co-occurred in language use. This experiment included a control condition where list items were not familiarised, that is, associations between items were derived from their natural co-occurrence.

The motivation for this study had been drawn from the free recall literature. While the serial recall of pure lists consistently demonstrated an advantage to HF words, Stuart and Hulme (2000) observed that in free recall, those studies where mixed lists had witnessed the reversal or elimination of the effect found with pure lists, demonstrating that performance of HF items was not universally superior to LF items, and that mechanisms involving the relationships between words might be an important factor that determined how well a word is remembered.

5.2.2 Association-based accounts in free recall (Deese, 1959, 1960)

Stuart and Hulme (2000) reviewed the work of Deese (1959, 1960) who, based on studies in long-term free recall, suggested that the frequency of co-occurrence between items was the more influential measure of memory performance and not the specific frequency counts for the items themselves. He manipulated lists based on their mean inter-item association, formed from pair-wise associations that were sourced from word association norms. When the mean inter-item association for a mixed list was zero, the frequency effect in free recall disappeared. Deese (1960) reasoned that the typical advantage seen for HF words was not due to the higher frequency of occurrence of each item, but because, as a function of higher usage, these items were more likely to share common contexts in natural language, and consequently possess stronger associative links.

5.2.3 The Stuart and Hulme (2000) experiment

The design was between groups with two factors, frequency (HF versus LF) and familiarisation (pairwise familiarised stimuli versus no familiarisation). For each participant, a twelve-word set (comprising either HF or LF words) was randomly halved into two subsets. Participants allocated to the no familiarisation condition were not exposed to the stimuli before the serial recall task. Participants assigned to the familiarisation condition underwent pair-wise familiarisation of the thirty possible pair-wise combinations from each six-item subset. Pairs of words were presented visually and the participants were required to read these aloud. All pairs for one subset were presented, followed by all pairs of the other subset. This process was repeated 10 times.

The subsequent serial recall task involved the presentation of lists that included pure lists made from the items within each subset, and alternating lists that were composed of alternate items from each subset. Half of the alternating lists began with an item from one subset and half of the lists began with an item from the other list.

The recall data identified effects of frequency and list type. High frequency words were better recalled than LF words, and pure lists were better recalled than mixed lists. Familiarisation across frequency, list type and serial position was not significant. There was however a familiarisation by list type interaction revealing a performance difference between pure and mixed lists for familiarised word sets but not for unfamiliarised word sets. A three-way interaction involving list type, frequency and familiarity was also present in the data. This can be summarised as a benefit to the recall of pure LF lists after familiarisation that did not occur for pure HF lists. A frequency by serial position interaction was identified revealing that recall for LF words declined more rapidly across serial positions than for HF words as Hulme et al. (1997) had earlier found.

Examination of the HF and LF data separately explored the interaction between frequency, list type and familiarisation. When HF data was analysed alone, serial position was found to be the only significant effect. The data for LF conditions showed that serial position and list type were factors influencing recall. Familiarisation, collapsed across list type and serial position did not affect recall performance; familiarisation *per se* did not result in better recall. However, lists constructed from the same subset of items, that is, lists containing pair-wise familiarised items, were better recalled than the alternating lists where adjacent items were not from the same familiarisation pool.

Furthermore, when the recall performances on HF and LF pure lists in the familiarised conditions were compared there was no difference. Stuart and Hulme (2000) argued that the associative links built during the exposure phase were responsible for the abolition of the frequency effect in this instance.

5.2.3.1 Item analysis

Error analysis investigated whether any of the manipulated factors contributed to observed patterns within error categories. Conditionalised order errors were not impacted by frequency, list type, familiarisation, or any of the interactions of these variables. Item errors however, were influenced by list type as more item errors occurred for mixed lists than pure lists. Frequency also contributed to the number of item errors; LF lists produced more errors than HF lists. An interaction between list type and familiarisation showed that under the familiarisation condition, the number of item errors for pure lists was less than for mixed lists. A three-way interaction between familiarisation, list-type, and frequency detailed that the reduction in errors for pure lists after familiarisation was larger for LF words. Stuart and Hulme (2000) noted that the pattern of errors replicated those for strict serial recall, and concluded that the effect inter-item associations have on recall is functionally located in memory for items.

Participants who were familiarised with the stimuli rarely made extra-set intrusions when recalling lists. Participants unfamiliar with the items tended to recall words from outside the stimulus pool, typically when the lists were LF. Intra-set intrusions occurred in all conditions, but were more likely in alternating lists relative to pure lists. It was noted that this difference was a reflection in the difference of the set sizes between pure and mixed lists (six and twelve items respectively). However, the magnitude of the effect was regarded as small.

Omissions constituted the largest proportion of item errors. Analysis of this data showed that frequency and lists type were significant main effects. A three-way interaction between frequency, list-type and familiarisation indicated that for the familiarisation condition with LF words, the difference in the number of omissions between mixed and pure lists was marked. As this pattern repeated that found for serial recall, it was argued that the benefit of co-occurrence resulted in greater protection from omissions in recall.

Stuart and Hulme (2000) argued that the equivalence of recall for pure LF lists after familiarisation and pure HF lists, and the difference in performance between lists of familiarised co-occurring items and lists of alternating familiarised items for LF lists, suggested the frequency effect could be explained entirely in terms of associative links between items. The authors conceded that due to a non-significant advantage of familiarised LF lists over unfamiliarised lists, and the observation that extraset intrusions were appreciable only in unfamiliarised LF conditions, it was possible that familiarity might have contributed to the improved recall of LF items. However this effect was considered to be small in relation to the co-occurrence effect.

5.2.4 The associative links/item co-occurrence hypothesis

The results of this experiment motivated Stuart and Hulme (2000) to modify the redintegration-based explanation. Rather than the accessibility of individual item representations in LTM determining the likelihood of item reconstruction, Stuart and Hulme (2000) suggested that the frequency effect might be an outcome of the pre-experimental associations formed from the co-occurrence of items in natural language. The representations of items would form a ‘mutually supporting network of item nodes’ (p. 801) facilitating the accessibility of each item’s long-term trace in the retrieval process; that is, the associative links would mutually excite connected list members and determine the accessibility of LTM representations (Saint-Aubin & Poirier, 2005).

5.3 The word frequency effect in randomly constructed lists

5.3.1 Roodenrys et al. (2002) Experiment 4

An experiment reported by Roodenrys et al. (2002) involved the presentation of lists where the frequency of items varied randomly across serial positions. They used a regression approach based on presenting items in all serial positions across the list set, and regressed the properties of each item onto its performance level by individual participant in the recall task (Lorch & Myers, 1990). The item pool comprised words ranging across frequency, neighbourhood size, and neighbourhood frequency values. It was argued that control for extraneous factors (e.g. speech rate) was unnecessary, as over a large sample every item would be presented equally often in lists that were combinations of all other items. Thus the likelihood of correct recall was not dependent on the articulation rate of the other items in the list. Additionally, as each item was repeated only six times, it was anticipated that the masking of effects associated with lexical variables by the repetition of items should be at a minimum.

Analysis followed the within-subjects procedure for repeated measures regression as outlined by Lorch and Myers (1990). Effects of frequency, neighbour size and neighbourhood frequency were present.

5.3.1.1 Item analysis

Words with more neighbours were less likely to be omitted, and words of lower frequency were more likely to be replaced by a higher frequency neighbour. Consistent with this observation, words of lower frequency were also likely to be replaced if they had a large neighbourhood, or a high neighbourhood frequency.

Both word frequency and neighbourhood size were variables that impacted on order memory, as HF words and words from larger neighbourhoods were more likely to be recalled in an incorrect position. Roodenrys et al. (2002) noted that some models of serial recall proposing separate mechanisms for the encoding and maintenance of short-term traces, and the second-stage restoration using long-term knowledge (e.g. Henson, Norris, Page & Baddeley, 1996; Page & Norris, 1998), would have difficulty explaining order effects associated with lexical variables. Other models however did allow for lexical effects to impact on order memory, although in a minor capacity. For example, the Burgess and Hitch (1999) model was argued by Roodenrys et al. (2002) to have the flexibility to accommodate small effects of order. In this model, long-term activation from HF items could combine with partial activation from the context signal to create sufficient competition to the correct item to jump the cue in a winner-take-all selection, resulting in early output.

5.3.1.2 The frequency effect in random lists in terms of associative links

Roodenrys et al. (2002) argued that their results were consistent with the Stuart and Hulme (2000) explanation of the frequency effect. This is not strictly true, as Stuart and Hulme (2000) argued that inter-item associations might operate in a non-directional way across the set of items making up the list. If this were the case then the last experiment should not have found a frequency effect; in this experiment HF items were mixed with LF items, and consequently the set-level activation across all items, as a function of co-occurrence, would have been at an intermediate level and differed randomly between lists. Nevertheless, if item co-occurrence operated in a directional fashion on successive pairs of items across the list, then the presence of a substantial frequency effect in this experiment could be accommodated by an associative account of serial recall.

5.3.2 Miller (2004)

A second random, mixed-list experiment was conducted by Miller (2004), who examined whether metrics of the semantic neighbourhood of items and word frequency influenced recall performance. Lists of six monosyllabic words were drawn from a pool of 87 items, and used the regression technique of Lorch and Myers (1990). An attenuated frequency effect was found, an order of magnitude smaller than that found by Roodenrys et al. (2002). Subsidiary analyses failed to find any confound of frequency with other variables that might influence serial recall performance; phonological similarity, phonological neighbourhood size, number of letters, number of phonemes or word concreteness were all matched across frequency values. However, all items were highly concrete (see Chapter 9 for a review of word concreteness in serial recall).

The result of this experiment is more consistent with the associative links proposal of Hulme et al. (2003; Stuart & Hulme, 2000) than the results reported by Roodenrys et al. (2002); in this case, the mixing of HF and LF items in the same lists produced a diminished frequency effect. A possible explanation of this effect size would argue that LF words benefited from their co-presentation with HF words, while the performance on HF words suffered because of the presence of the LF items in the lists. It is feasible that the supportive activation from inter-item associations improved the recall of LF items, at the cost of HF items, yielding a much reduced effect size.

5.4 The word frequency effect in alternating lists

Hulme et al. (2003) observed that the frequency effect had been more thoroughly investigated in free than serial recall, and until Stuart and Hulme (2000) the explanations of the frequency effect in these paradigms remained separate. Hulme et al. (2003) investigated the serial recall of lists of pure and alternating frequency, and proposed that the results from their study could go someway to forging a link between free and serial recall, implicating common, basic mechanisms of memory.

Furthermore, this would shift the basis of interpretation for serial recall from an item-based effect to one involving the nature of relationships between list items.

5.4.1 The frequency effect for mixed lists in free recall

The typical frequency effect in the free recall of pure lists was reported as a performance advantage for HF words (e.g. Deese, 1960; DeLosh & McDaniel, 1996; Tan & Ward, 2000; Ward, Woodward, Stevens, & Stinson, 2003; M.J. Watkins et al., 2000). However, when lists were mixed the nature of effect was varied. In some instances HF words maintained superior recall (e.g. Balota & Neely, 1980), while in others there was no difference in recall (e.g. M.J. Watkins et al., 2000, Experiment 1), yet on some occasions LF words were better recalled than HF words (e.g. DeLosh & McDaniel, 1996, Experiment 2a). Across these outcomes, it was acknowledged, that when compared to performance with pure lists, mixed lists result in reduced performance for HF words and enhanced performance for LF items (Hulme et al., 2003).

5.4.1.2 The order-encoding hypothesis

Differential processing and consequent resource constraint formed the basis of one explanation of the elimination or reversal of the frequency effect in free recall. The order-encoding hypothesis (DeLosh & McDaniel, 1996) argued that the extra attention required to encode LF items would limit the resources available to encode item order. In mixed lists this would equate order encoding between HF and LF items, however the LF items would also maintain an advantage due to the additional item encoding they receive. This would form the basis of the enhanced LF recall in mixed lists. Furthermore, M.J. Watkins et al. (2000) identified that the better free recall of LF items in mixed lists was the outcome of a conscious strategy to differentially encode these items when the recall task was explicit. Hulme et al. (2003) noted that the slower presentation rates more typical of free recall experiments might promote the use of this strategy whereas the presentation of items in ISR (usually about one item per second) was less likely to encourage differential item encoding in this task.

In addition, the proposed relationship between differential attention to items and free recall performance might be mediated to some extent by the differential rehearsal of items (see Ward et al., 2003). In the free recall of pure lists, the superior retention of HF words was coupled with more rehearsals and rehearsals that occurred later than those for LF words (Tan & Ward, 2000). In mixed lists however, the differences in both recall and rehearsal became non-significant with the descriptive difference just favouring HF words. When rehearsals were equated for both frequency and recency a significant, though minor difference was found, and this was suggested by Ward et al., (2003) to be the influence of inter-item associations which was argued, on the basis of recall order, to be greater for HF than LF words. Because free recall placed no requirements on the order of recalled items, it was assumed by Hulme et al. (2003) that the order of responses reflected, at least in part, pre-existing connections between words that would act as useful cues/prompts to subsequent items.

5.4.2 Predictions for alternating lists

The mixed lists of Hulme et al. (2003) were constructed so that HF and LF items alternated across serial positions. The authors identified three classes of predictions for the recall performance in mixed lists. Item-based accounts (e.g. Hulme et al., 1991; Hulme et al., 1997; Schweickert, 1993) would anticipate a saw-tooth response for mixed lists, where recall for respective items would match the levels of recall in pure lists, as each item’s likelihood of successful redintegration would be determined by its frequency. Association-based accounts, derived from the free recall literature (e.g. Deese, 1959, 1960) and explored more recently in serial recall by Stuart and Hulme (2000), predicted equivalent and intermediate recall for HF and LF items in mixed lists; the level of recall for a set of items would be determined by the inter-item associations between list members. High frequency words are more likely to co-occur in natural language than LF words, and therefore the recall of pure HF lists would be expected to be superior to the recall of pure LF lists. In mixed lists the level of co-occurrence between HF and LF items would be at an intermediate level given that HF items would be expected to co-occur with LF items to a moderate degree. Finally, resource-sharing accounts (e.g. DeLosh & McDaniel, 1996) would predict intermediate levels of performance for mixed lists because they assume constraints on the available resources for the processes enabling recall, for example, item encoding or item maintenance.

5.4.3 Serial recall of alternating lists with closed word pools

The first experiment reported by Hulme et al. (2003) was a test of immediate serial recall for pure and mixed lists using a closed pool of 8 items (the medium length items from Hulme et al., 1997). A brief familiarisation procedure and speech rate test for the items (the time taken to repeat each item 10 times) preceded the memory task. Because a small difference in the speech rate data was detected (HF words were articulated faster than LF words), a covariate analysis was performed to determine whether effects were present after speech rate had been controlled for. This analysis identified that the pure list results replicated those found by Hulme et al. (1997), while the mixed lists revealed an effect of serial position, but no effect of list, and no interaction. Thus, in mixed lists, HF items were recalled no better than LF items.

5.4.3.1 Item Analysis

Error analysis was conducted to determine any differences in processing between items and list contexts. The item levels of error analysis compared HF and LF words in pure and mixed lists respectively. Hulme et al. (2003) argued that this allowed for the examination of effects of item type in the context in which they were presented. In terms of item memory, HF and LF items were recalled equally well in mixed lists, HF items were recalled less well in mixed than pure lists, while the reversed pattern occurred for LF items. Conditionalised order errors showed that while there were fewer order errors for HF items, this difference was small and was not influenced by list type. Therefore HF words demonstrated a small advantage for order, however, differences were too small to reconcile with the differences in serial recall, and incompatible with the difference between pure and mixed list performances. Accordingly, the contribution of frequency to serial recall was located in memory for items, as others had suggested (e.g., Whiteman et al., 1994).

The patterns of item errors replicated the serial recall data. The majority of errors for all conditions were made as omissions. In pure lists, HF items in pure lists were less likely to be omitted than LF items, but omissions for HF and LF words in mixed lists were equated. Furthermore the level of omissions in mixed lists was intermediate with respect to the pure list envelope.

The results of the first experiment were seen by Hulme et al. (2003) to favour an association-based view of short-term recall. Specifically, the pre-experimental associations between list items in permanent memory were argued to facilitate recall performance. Stronger inter-item associations would occur for HF than LF words, and these would produce the HF advantage in pure lists. In contrast inter-item associations in the mixed lists would be of intermediate strength and result in the recall of HF and LF at equivalent and intermediate levels.

5.4.3.2 Latent Semantic Analysis (LSA)

Hulme et al. (2003) used Latent Semantic Analysis (LSA, Landauer & Dumais, 1997) to examine the inter-item associativity between words in each item pool. LSA is a statistical decomposition method that identifies the location of words in multi-dimensional psychological space. The resultant cosine metric is considered to be a measure of semantic similarity and by virtue of the method used in LSA is, in effect, an assessment of the commonality of context for two words. The dimensions of psychological space are determined from a selected text corpus. Using the ‘literature with idioms (528 factors)’ data set Hulme et al. (2003) determined that the HF set had higher pair-wise associations than the LF set, and that average of all pair-wise associations across all items was intermediate to the pure sets. Thus, LSA was seen to ratify the assertion that inter-item association was responsible for the nature of the frequency effect as observed in their first experiment.

5.4.4 The serial recall of alternating lists with open word pools

As the first experiment used closed word pools the size of the frequency effect may have been masked because LF items would have increased in familiarity across the life of the experiment and item information for LF words would be made more available (Roodenrys & Quinlan, 2000). However, as Hulme et al. (2003) noted, the pattern of item errors did not suggest that LF items were as available as HF items (although early and late trials were not compared to fully support this claim).

A second experiment replicated the method of the first with open stimulus sets. The memory task was administered first and was followed by a speech rate check on one of 6 randomly constructed subsets of stimuli.

This experiment replicated the results found in Experiment 1 except that the overall recall level was less in this case. This reduction was attributed to the open sets of stimuli used (Roodenrys & Quinlan, 2000). A reliable difference in speech rate between HF and LF frequency sets was observed, and so effects of speech rate were statistically removed by covariate analysis. This identified an effect of frequency and a significant frequency by position interaction for the pure lists (Hulme et al., 1997). Alternating lists also produced effects for frequency and a significant frequency by position interaction. Mixed lists exhibited intermediate levels of recall with respect to pure lists, however if they began with an HF word they were slightly better recalled than if they commenced with an LF word. The interaction was driven by greater recall for lists beginning with a HF word some serial positions, although simple effects analysis showed none of these to be significant.

5.4.4.1 Item Analysis

The error data identified that in mixed lists, item recall was less for HF than LF words, although this difference was not reliable. When the performance of HF and LF items in mixed lists was compared with the levels of correct recall for pure lists, the cost and benefit to HF and LF items respectively was apparent. The level of order errors for HF and LF items across conditions was comparable although Hulme et al. (2003) did not report the reliability of any differences. They did however report that transposition gradients suggested some differences with order errors, namely that transpositions in mixed lists tended to be less than in pure lists, and there was a tendency for items of the same frequency to transpose. (Note that this order analysis maintained items within list context in contrast to the item analysis described above).

The pattern of item errors mirrored the serial recall performance across the list types. In this experiment, the pure list advantage for HF items was unreliably reversed in mixed lists. Omissions again formed the greatest component of item errors, and repeated the pattern of item errors most closely; fewer HF omissions in pure lists contrasted with greater HF omissions for alternating lists.

Hulme et al. (2003) promoted the results of Experiment 2 as further evidence against item-based accounts of the frequency effect in serial recall, given the generality of the findings across two sets of stimuli. The use of open pools of items did not alter the main findings observed in Experiment 1.

5.4.5 Serial recall with alternating lists of words and nonwords

Hulme et al. (2003) conducted a final experiment to further explore the associative-based account by examining the recall of alternating lists of words and nonwords. In pure lists words are better recalled than nonwords (Hulme et al., 1991). This effect has been argued to reflect differences in redintegration between words and nonwords; words possess long-term phonological representations while nonwords, by definition do not have lexical representations. The results of Experiments 1 and 2 demonstrated that item-based accounts were unable to accommodate the nature of the frequency effect across list composition. This final experiment sought to investigate whether an item-based explanation was still compatible with nonword recall.

A large difference in the recall of words and nonwords occurred in the pure list conditions, replicating findings from other studies (e.g. Hulme et al., 1991). A lexicality by serial position interaction was also evident in the data. A saw-tooth pattern for the mixed lists was found, where words were recalled as well as words in the pure list condition, while nonwords were recalled less well than words but substantially better than nonwords in pure lists. Nonwords were articulated faster than words in this experiment, so differences in speech rate could not account for the direction or significance of results. The pure list data revealed a significant effect of list type, as well as a significant interaction of list type with serial position; the size of the lexicality effect increased across serial positions. The mixed list data produced a list type by serial position interaction, but no main effect of list type; recall for mixed lists was equivalent across all serial positions. The interaction was driven by the presence of a sawtooth pattern in the data.

5.4.5.1 Item analysis

The differences between the recall of words and nonwords were attributed to the pattern of item errors formed by examining each stimulus type in pure and mixed conditions, respectively. Nonwords incurred less item errors in mixed than pure lists, while more item errors occurred with words in mixed than pure lists. Omissions formed the largest proportion of item errors and followed the general pattern of results. Extraset intrusions were rarely made when recalling words, but were far more common when recalling nonwords. The number of intraset errors was small for all cases. With respect to order errors, words were better recalled than nonwords, but the order error rate improved for nonwords when presented in mixed lists. The latter observation was noted as consistent with the division of mixed lists into word and nonword sublists, thus limiting the opportunities to recall nonwords in the wrong position.

In terms of pure lists, there was a parallel between the results found for HF and LF words, and words and nonwords respectively; if nonwords are considered to be the limiting case of low familiarity, then better recall occurred with more than less familiar items. In contrast, the performance on mixed lists varied markedly depending on whether alternating lists comprised HF and LF words or words and nonwords. Mixed lists of HF and LF words generated curves that were smooth across serial positions, while alternating lists of words and nonwords resulted in sawtooth patterns. In the latter case, words were recalled as well as words in pure lists, implying the recall of words is an item-based phenomenon although alternatively, this may have been an artefact of the lack of association nonwords share with words. Accordingly, the recall of words might be supported by other words in the list, making the circumstance similar, from a redintegration perspective, to the recall of pure words.

The more difficult outcome to account for was the better recall of nonwords in mixed lists when compared to pure lists. Hulme et al. (2003) argued that this might be the result of a greater capacity to keep active phonological information when the number of nonwords in the list was reduced, or perhaps was a reflection of a limit on the amount of information that can be redintegrated in combination with spoken output, especially when LTM lexical/semantic representations were non-existent.

5.4.6 Associative and item-based processes in serial recall

The results for mixed lists of HF and LF items suggested that associative mechanisms operated in the recall of items. However, this did not translate to mixed lists of words and nonwords; the evidence presented suggested that the recall of words in these instances operated on distinct, item-based processes.

In summary, the results of the experiments investigating the nature of the frequency effect in alternating lists presented a serious threat to explanations of serial recall that viewed performance as a function of the properties of the item alone (e.g. Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999). These results did not undermine the proposition that redintegration is an integral feature of recall, but they did demand an explanation of how influence from the set of items might transfer onto the retrieval of each word.

Hulme et al. (2003) argued that the frequency effect in serial recall was dependent on attributes of the items that constituted the retrieval set for a presented list. This was assumed to include the just-presented list items and possibly some items from former lists, but was constrained in size. These representations would be maintained in a state enabling retrieval at the point of recall. Furthermore, the idea of a restricted, activated retrieval set had a precedent in some models of memory (e.g. the Feature model, Nairne, 1990). The authors argued that the pattern of recall errors was consistent with the existence of a restricted search set (relatively few extra-set intrusions, some intra-set intrusions, but mostly omissions). It was more likely that a word would be omitted, than that it would be substituted with an item from outside the list.

Furthermore, the relationships between words in the retrieval set appeared to be central in determining the level of recall performance that could be achieved for an item of a given frequency. The key characteristic of these relationships was argued to be the pre-existing association strength of item traces in lexical memory (Stuart & Hulme, 2000), a property correlated with the frequency count of items (Deese, 1960). Furthermore, Hulme et al. (2003) argued that the premise of varying strength of association, according to the properties of the list items, was supported by the results from LSA. The levels of semantic similarity, derived from this model of psychological space, corresponded to the posited theoretical association strengths for each condition.

The smooth curves found in the mixed lists for the frequency-based experiments, contrasted with the sawtooth curves for mixed lists of words and nonwords where redintegration was presumed to operate for words as they do in pure lists. This difference, it was reasoned, implicated the operation of an associative mechanism in isolation from the process of redintegration. It is noteworthy that this proposal, pre-empted by Stuart and Hulme (2000) implied a relaxation of the separation between short-term and long-term stores, as the creation of an associative network on which recall would be based required access to LTM before the point of recall.

When the results of Experiments 1 and 2 were compared with the findings from free recall, a common theme emerged. In mixed list experiments, the recall of LF words was advantaged when compared to the recall of these items in pure lists, while the recall of HF words in mixed lists was harmed relative to recall in pure lists. Hulme et al. (2003) argued that this similarity between free and serial recall supported the claim that common mechanisms operated in these recall tasks.

5.4.7 The Temporal Context Model (TCM, Howard & Kahana, 2002a, 2002b)

5.4.7.1 The TCM as an explanation of the effects of pre-experimental associations on recall

A model of episodic memory, the Temporal Context Model (TCM, Howard & Kahana, 2002a), founded on observations from free recall, was argued by Hulme et al. (2003) to offer a basis to the serial recall patterns of pure and mixed frequency lists. This model associates the semantic representations of presented items with evolving temporal context representations. Recall for an item in a list is dependent on reinstating the temporal context of that item, and the recall of the item is used as a cue to retrieve the next temporal context, and so on. Thus it is argued that recall is dependent on temporally formed inter-item associations, a proposal argued by Howard and Kahana (2002a) to be supported in free recall by observed relationships between the order of recalled items and the relative order of presentation (lag-conditional response probability (lag-CRP), Howard & Kahana, 1999; Kahana, 1996). Furthermore, this tendency is biased towards the recall of items in a forward order; the retrieval of item N in a list is most likely to cue the retrieval of item N+1. A second influence in TCM is sourced from the pre-existing semantic associations between items. Howard and Kahana (2002b) demonstrated in free recall that the likelihood items will be outputted in nearby positions is influenced by the strength of semantic association, but that this effect is reduced as temporal distance between associates increases.

Hulme et al. (2003) argued that the TCM was a lens through which their results could be interpreted. They demonstrated in their first experiment that the pattern of LSA association strengths mirrored the level of serial recall for pure HF, pure LF, and mixed lists, and suggested that inter-item associations based on pre-experimental associations in natural language were likely to be responsible for the varying performance of HF and LF items across list formats.

The TCM states formally that inter-item associations drive the forward order of recall, and the stronger the semantic association between successive items, the greater the likelihood that the first item in a pair will facilitate the retrieval of the second. This suggests that effects of inter-item semantic association should be localised to the associative strengths of consecutive pairs of list items. Alternatively, Stuart and Hulme (2000) proposed that the support of inter-item associations in serial recall operated in a mutually supporting and non-directional manner, facilitating the retrieval of all list items to a similar degree. The alternating lists used by Hulme et al. (2003) did not provide an adequate test to discriminate between these positions, as frequency-wise alternation would yield intermediate levels of associative strength between consecutive pairs of items across the list (HF item to LF item, or vice versa), as well as an intermediate association strength at the list level (each mixed list contained 3 HF and 3 LF items). Therefore, the question of whether associative effects are directional or non-directional, remained unresolved.

5.4.7.2 A critique of the reported instantiations of TCM and data analysis underpinning its explanatory capacity

Recently, Farrell and Lewandowsky (2008) have reported a number of shortcomings with the TCM as described and modelled in the literature (Kahana, 1996; Howard & Kahana, 1999, 2002a, 2002b). One concern centres on the accuracy of the lag-CRP function that describes the likelihood of recall of the next item, according to its relative distance from the just-recalled item, as this forms a key descriptive feature of the model. In the published work to date these have been generated by analyses that have excluded data, that is transitions, at larger lags. Full analyses reveal the tendency for individuals in free recall to make large transitions to ends-of-list items, the primacy and recency portions of the list, for a substantial proportion of all transitions. While immediate transitions to adjacent items remain the most likely, the form of this function is not monotonically decreasing for forwards and backwards lags, as originally described (Kahana, 1996). A second concern relates to the demonstrated incapability of the model to reproduce these functions, even when based on partial data, for immediate free recall. Furthermore, in general, across a number of experimental conditions (delayed and immediate), the model cannot replicate these functions, and functions describing the probability of first response by serial position, when they have been formed on the full set of item-to-item transitions. Another problem refers to the simplifying assumptions made in the original instantiation of the model. While these reproduced acceptable first response probability and item transition probability functions, they were mutually incompatible. That is, the model's ability to describe these features simultaneously was founded on divergent assumptions regarding the state of temporal context at the commencement of recall that could not, in reality, co-exist.

Lastly, Farrell and Lewandowsky (2008) determined that a modified version of the TCM, altered to remove this incompatibility, was still incapable of describing item transition probabilities for immediate recall when they were not aggregated across serial positions, as lag-CRP functions vary when analysed separately. These authors did not disagree with all of the basic premises embodied in the model, for example, they noted the greater likelihood of transitions to adjacent items in the list as an indication of localised associations that would be either direct item-to-item links, or indirect links through temporal context. They also noted that the evolution of context as items are encoded and retrieved is appealing. However, the TCM, as currently described, could not be viewed as a viable computational model of recall.

5.4.8 A note regarding the use of corpora in LSA

Hulme et al. (2003) had reported that the average inter-item association of each list type in Experiment 1, as determined by LSA cosine values, corresponded to the levels of semantic association predicted by an associative explanation of the frequency effect. The HF stimuli were more strongly associated than LF stimuli, and the associations between HF and LF were intermediate and reliably different to the pure sets. The authors had used the 'Literature with idioms (528 factors)' as the corpus on which psychological space was defined in LSA[1]. However, if the 'General reading up to first year college (300 factors)' corpus is substituted with these stimuli, a different pattern of results emerges. It should be noted that one LF item, *bequest*, is absent from this space, however, while cosine values for HF words (mean .10, range

[1] This corpus is no longer available on the LSA website (http://lsa.colorado.edu/)

.03 - .28) were reliably greater than cosine values for LF words (mean .02, range -.12 - .14), $U = 104$, $n = 49$, $p < .001$, and reliably greater than the cosine values for the mixed set (mean .03, range -.04 - .18), $U = 236$, $n = 80$, $p < .001$, the distributions of LF and mixed set values were not reliably different, $U = 519$, $n = 73$, $p = .740$. This anomaly may be specific to the small stimulus sets used in this case, however, given the variability in the results as a function of corpora these measures should be used carefully.

5.5 Word frequency in alternating lists with incidental learning

5.5.1 Conscious manipulation of memory processes

M.J. Watkins et al. (2000) found the abolition of the frequency effect in the free recall of mixed lists with intentional but not incidental learning, and argued that participants make conscious use of the properties of stimuli that will maximise their performance in memory tasks. Morin et al. (2006) sought to determine whether, in serial recall, the elimination of the effect in mixed lists was a function of conscious strategy, or a more basic property of the short-term recall process. In addition, they investigated whether the pattern of results would alter with the introduction of a retention interval between stimulus presentation and recall, comparing short-term recall with recall under conditions that more closely resembled those used by M.J. Watkins et al. (2000). This would inform how generalised the effect might be.

5.5.2 Morin et al. (2006)

A first experiment involved the presentation of a single alternating frequency list (HL or LH) to each of four groups that recalled it under incidental or intentional learning conditions. In the incidental condition the lists were introduced as a distractor activity to a spatial STM task. On the critical trial, instead of performing the spatial memory recall, participants were asked to write down in order the list of words they has just seen. The intentional condition followed the normal procedure for an STM task; items were presented in a list and participants were instructed to recall them in order prior to list presentation.

The results indicated that while better recall was produced under intentional conditions (in both strict and lenient scoring conditions), there was no effect of frequency and no interaction between frequency and task. Order error data showed an effect of task, as the incidental task induced more order errors than the intentional task. However, no effect of frequency, and no interaction between frequency and task were observed for order memory.

While the results of the first experiment did not support the idea that participants consciously controlled the level of attention paid to unfamiliar items, it was also possible that a frequency effect for pure lists constructed from the word sets used may not have eventuated in serial recall. As pure lists were not tested in the first experiment, a second experiment examined the incidental STM performance for pure and mixed lists, where the serial positions of the items in the pure lists were replicated in the mixed lists. The procedure followed a similar procedure to the initial experiment, and any participants who reported that they anticipated a memory test were eliminated from the data set.

No effect of list type, or word frequency was observed but there was a list type by frequency interaction and this pattern existed for strict and lenient scoring of recall. There was a HF advantage in pure but not mixed lists. The order error data did not reveal any reliable effects. Therefore, incidental recall was found to produce a frequency effect in pure, but not mixed lists.

A final experiment examined whether an influence of participant-controlled strategy might be present after a retention period. This experiment replicated the second, however the spatial STM distractor activity included in previous trials was present in the critical trial and formed a 30 second, activity-filled delay in the task. After completion of the distractor activity instructions to recall the list were given.

For both strict and lenient scoring there was an effect of word frequency, no effect of list type, but a significant interaction. High frequency words were advantaged in pure but not mixed lists. Analysis of order errors revealed that more order errors occurred with LF words in the pure LF condition. The pattern of performance for pure and mixed lists was not altered by the presence of a retention interval.

5.5.3 Basic memory processes as a source of the mixed list paradox

The experiments by Morin et al. (2006) identified that the abolition of the frequency effect for mixed lists was not an outcome of task awareness, as this occurred when either incidental or intentional learning took place. Furthermore, this pattern was observable in both immediate and delayed recall. Given the effect was present under incidental learning conditions, Morin et al. (2006) argued it was reasonable to assume that the STM processes responsible for the nature of the frequency effect are not consciously mediated.

These results contrasted with work by M.J. Watkins et al. (2000) who found a standard frequency effect for pure and mixed lists under incidental but not intentional learning conditions, and argued this reflected more effortful encoding of LF items when list recall was anticipated. Alternatively, Dewhurst, Brandt, and Sharp (2004), taking a similar approach, found that the frequency effect was eliminated with incidental learning but that an LF advantage occurred for an intentional learning condition. These results were argued to be more like those determined by Morin et al. (2006), in that they did not support a conscious encoding strategy explanation of the frequency effect in mixed lists.

Morin et al. (2006) affirmed that the most common explanation of the frequency effect was the redintegration hypothesis (Hulme et al., 1997). However, this explanation could not account for the memory performance with alternating lists, as the observed patterns in these conditions suggested the recall of a list item was not wholly dependent on the properties of that item alone (e.g. LTM trace accessibility) but relied in some way on the list context. Therefore, Hulme et al.'s (2003) account of the redintegration hypothesis arguing that the accessibility of LTM representations would be governed by the co-occurrence of list items was preferred as an explanation.

5.5.4 The item-order hypothesis

Morin et al. (2006) suggested that the item-order hypothesis (Nairne, Riegler, & Serra, 1991; Serra & Nairne, 1993), derived to explain the dissociation of long-term item and order information in the generation effect (and a forerunner to the order-encoding hypothesis - DeLosh & McDaniel, 1996), could also account for these findings. This position argues that items are encoded in terms of item information and ‘relational’ or order information, and the degree to which item information assists recall is dependent on how well order information can support an output strategy. Morin et al. (2006) suggested the mixed list paradox would result from the trade-off between the processing of these forms of information between words of different frequency classes. Low frequency items would take more resources to successfully encode item information, leaving less available for order information encoding. Assuming that recall is driven by order information, they argued that the effect in pure lists would result from an order information advantage to HF words. In mixed lists, the order memory processing for HF would be penalized relative to pure lists while LF words would benefit in terms of order-information encoding. The better item encoding for LF words would facilitate the greater recall of these items, nullifying or even reversing the frequency effect.

Morin et al. (2006) proposed that a full test of this account would need to explore performance on separate tasks more focused on item and order memory respectively (e.g. recognition versus serial recall/serial reconstruction of order). They argued, according to the strict scoring used in these experiments, considered to be a measure of order memory (Saint-Aubin & Poirier, 1999), that their results were compatible with the item-order hypothesis. However, in the immediate recall of pure lists, conditionalised order errors did not reflect contributions from word frequency, while differences in item memory were still apparent. That is, evidence for a difference in item encoding existed but the corresponding evidence for a difference in order information encoding was absent. Therefore it is difficult to see how this position, and more specifically the item-order hypothesis could be maintained as a general influence in serial recall; Hulme et al., (2003) rejected the order-encoding hypothesis as a valid explanation of the frequency effect in serial recall on the grounds that variations in order memory by item type were small or nonexistent.

5.6 Summary

This chapter presented research that has been used to argue for an item co-occurrence (item associativity) interpretation of the frequency effect in serial recall. The importance of pair-wise associations between items was emphasised by the improved recall of LF words after familiarisation (Stuart & Hulme, 2000). Recall in this case matched that found for HF items, and contrasted with the recall of alternating LF lists constructed so that adjacent items came from different item sets. This list type arguably reflected effects arising from familiarisation alone, as alternation would result in the decoupling of learned associations between pre-exposed words. That such an effect was found for LF and not HF items was a function of the relative paucity of associations LF words possess with other LF items, and their corresponding sensitivity to the facilitation of new associative links. However, it should be noted that the type of learning established during familiarisation would have been minimal in terms of semantic association, making this a weak approximation to the types of association tested by Deese (1959, 1960), and those more typical of the co-occurrence of items in natural language.

Experiments using the random placement of HF and LF items within a list demonstrated the variability with which the frequency effect could be expressed (Miller, 2004; Roodenrys et al., 2002). The presence of a sizeable effect in the Roodenrys et al. (2002) study alluded to the possibility that the effect might not be non-directional but sensitive to the order of items across the list. In contrast, the much smaller effect found by Miller (2004) was more compatible with a non-directional interpretation of co-occurrence, although this study may have been affected by other influences (a restricted concreteness domain).

The most compelling evidence for an inter-item associativity interpretation for the frequency effect was provided by Hulme et al. (2003). In two experiments they demonstrated that the HF advantage in pure lists is abolished with the serial recall of lists of alternating frequency. Furthermore the recall patterns established in these experiments possessed distinct differences to the recall patterns for pure and alternating lists of words and nonwords. In particular, the smooth curves of mixed frequency lists contrasted with the sawtooth curves observed for lists alternating words and nonwords. The level of recall for words in alternating lists matched the recall for pure lists, while recall for nonwords in mixed lists was substantially better than for nonwords in pure lists. Hulme et al. (2003) interpreted these results collectively to be suggestive of a redintegrative process that operates at the item level, but includes a second component within redintegration sensitive to the pre-experimental associations between list items.

The results of Morin et al. (2006) suggested that the processes associated with word frequency in serial recall were not consciously accessed, and therefore could not be manipulated by participants to suit task demands. Furthermore, co-occurrence (inter-item associativity) effects could be viewed as a derivative of the organisation of items within LTM.

In accordance with the research conducted on pure lists, the frequency effect was shown to impact item memory, and in particular the level of omissions observed in a condition. Mixing lists resulted in greater levels of omissions for HF words, matched by a reduction in omissions for LF words, suggesting that inter-item associativity provides some protection against the failure to retrieve an item (Hulme et al., 2003). Stuart and Hulme (2000) also observed this pattern between pure and alternating familiarised lists with LF words. Order errors, when reported (Hulme et al., 2003; Roodenrys et al., 2002) were small and were not of a magnitude to suggest that frequency is a major determinant in the management of order information.

Chapter 6

Alternative Explanations for the Frequency Effect in the Serial Recall of Lists

6.1 Introduction

Like those of the last, the studies reviewed in this chapter demonstrated that the frequency effect is altered by list context. However, in contrast to the studies of the last chapter, different interpretations are placed on the results. In some cases, the results have been seen to support item-based redintegration (Saint-Aubin & LeBlanc, 2005; Saint-Aubin & Poirier, 2005), although this is qualified by influences from other sources (e.g. item distinctiveness or familiarity). Alternatively, an interpretation of lexical-semantic effects, including frequency, in the recall of mixed lists of words and nonwords has been developed from a psycholinguistic perspective, and more specifically from observations of the recall performance of patients with semantic dementia and the nonword recall of normal individuals (Jefferies et al., 2006a; Patterson, Graham, & Hodges, 1994). Consistent with other language-based accounts, this hypothesis argues that the influence of lexical-semantic variables proceeds from the point of encoding onwards, and is not a late-stage contribution to recall.

6.2 Item co-occurrence effects and word frequency revisited

6.2.1 Familiarity confounded with set size

Saint-Aubin and Poirier (2005) tested whether the effect attributed to item co-occurrence by Stuart and Hulme (2000) could be replicated by an item familiarity condition that was procedurally independent from pair-wise familiarisation. These authors argued that the familiarisation procedure used by Stuart and Hulme (2000) confounded familiarity and set-size in the recall of pair-wise familiarised and item familiarised lists, as pure lists were constructed of items from within each pair-wise familiarised set, but alternating lists were assembled from items drawn from both sets. While the familiarisation of all items had been controlled – namely all items were presented an equivalent number of times – the stimulus sets for the alternating lists contained twice as many items as those for the pure lists. Consequently, the difference in recall between pair-wise and individually familiarised items observed by Stuart and Hulme (2000) may have been a function of set size; superior recall for the familiarised pure LF lists may have resulted from the knowledge that list items 'belonged together as a set' and not from induced associations between items. If this were the case, a recall advantage might also be achievable in a familiarisation condition that did not promote pair-wise association.

6.2.2 Saint-Aubin and Poirier (2005)

A between groups design included the factors of frequency (HF and LF) and familiarisation (no familiarisation, item familiarisation, or pair-wise familiarisation). Pair-wise familiarisation followed the procedure used by Stuart and Hulme (2000). Item familiarisation involved the random allocation of the set of twelve items into two subsets, followed by the presentation of the first set of six items in the same order 10 times, and then the second set. The presentation time for the familiarisation of individual items was half that for the pairs in the pair-wise familiarisation procedure, to ensure that exposure times for each of the active familiarisation conditions were equivalent. The serial recall task followed immediately after familiarisation. The item familiarity condition had further constraints imposed on the sequencing of items within the recall trials to eliminate any order or position-related effects acquired during familiarisation.

The experiment produced a significant frequency effect; HF words were better recalled than LF words. However, the effect of familiarisation and the interaction between frequency and familiarisation were not significant. Saint-Aubin and Poirier (2005) attributed the absence of these effects, consistent with the report of Stuart and Hulme (2000), to the use of closed pools of items. They argued that item familiarity had developed during the course of the recall task and masked effects from pre-exposure.

To examine this more closely, performance on blocks of the first and last 10 trials were analysed. The effects of block, serial position and frequency were significant. Additionally, interactions between block and serial position, block and familiarisation, and frequency and familiarisation were found. The first of these related to an improvement in recall between the first and last blocks that was greater for later serial positions. The interaction between block and familiarisation revealed that the change across blocks was larger for the unfamiliarised condition than either item familiarisation or pair-wise familiarisation. The unfamiliarised condition demonstrated greater recall improvement across blocks than either of the familiarised conditions.

The interaction between frequency and familiarisation was due to the presence of a frequency effect for the unfamiliarised and item familiarised conditions but not for pair-wise familiarisation. Saint-Aubin and Poirier (2005) argued, despite the apparent support for the co-occurrence proposal of Stuart and Hulme (2000), strict serial recall scoring was not the most appropriate test for item-based effects (Saint-Aubin & Poirier, 1999); Stuart and Hulme (2000) failed to find a frequency by familiarisation interaction with strict scoring, but found a significant interaction in the analysis of item memory measures. However, only the extra-set intrusion category in Stuart and Hulme's (2000) data revealed this interaction, as most extra-set intrusions occurred in conditions with unfamiliarised LF words. Furthermore, differences in design between these experiments implicated different interactions as key features of analysis. Because Stuart and Hulme (2000) had tested 'item' familiarity in terms of the contrast between pure and alternating lists, and this was a within-subjects variable, the three-way interaction between frequency, list type and familiarisation would be most informative regarding differences between pair-wise and item familiarisation conditions. This pattern was realised by Stuart and Hulme (2000) in strict serial recall, overall item errors, and omissions and reflected the difference between recall for pure and alternating LF lists after familiarisation had occurred.

6.2.2.1 Item analysis

Analysis of order error data revealed that familiarisation and frequency were not influential in altering order error rates, and the interaction between frequency and familiarisation was also not significant. The final lists produced less order errors than the initial lists and a higher number of order errors for pair-wise familiarised than unfamiliarised HF words occurred in the final block. The general insensitivity of order error rates to frequency agreed with previous research examining memory for order (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Stuart & Hulme, 2000).

With respect to item errors, an effect of frequency was found, but familiarisation did not contribute to differences in item error occurrence, and there was no interaction between frequency and familiarisation. High frequency lists produced less item errors than LF lists. While the omnibus test for the interaction failed, planned contrasts showed that familiarisation reduced the number of item errors for LF lists but not HF lists, and post hoc testing revealed that although more item errors occurred with the unfamiliarised condition, both forms of familiarization produced equivalent levels of item error. Although not acknowledged by the authors, the failure of the omnibus test to detect differences between levels of familiarization, suggested that the data from this experiment was highly variable.

Examination of performance by first and last blocks showed that item errors decreased in the final block, while more item errors occurred with LF than HF conditions across blocks. There was also an effect of familiarization, and the significant three-way interaction was accompanied by significant interactions for frequency and familiarisation, and block and familiarisation.

The item error data for each block was analysed separately. The first block identified frequency and familiarisation effects, as well as an interaction. There was a familiarisation effect for LF but not for HF items. Once more post hoc testing revealed that more errors were produced for unfamiliarised than familiarised conditions but that familiarised conditions did not differ.

Saint-Aubin and Poirier (2005) also conducted an analysis of item error categories. In summary, according to item memory metrics, there was the usual effect of fewer errors for HF items coupled with a general reduction in item errors across blocks. Familiarisation encouraged greater levels of intra-set intrusions due to the pre-exposure of items, while more extra-set intrusions were recalled for unfamiliarised LF words. Importantly, both forms of familiarisation resulted in fewer errors when compared to a control condition, but neither familiarisation condition was superior to the other in this regard.

6.2.3 Familiarisation mistaken for associative links/item co-occurrence

The authors proposed that there were two alternative explanations for this equivalence; that either both forms of familiarisation produced inter-item associations to a similar degree, or that the familiarisations produced equal familiarity in the items, and this produced superior recall. The method of pre-exposure for pair-wise association was designed to produce stronger links between items than might arise from the presentation of individual items; the failure of pair-wise familiarisation to produce better item recall than pre-exposure to individual items was argued to challenge the validity of the associative links hypothesis. Instead, the improvement to recall was argued to be due to the increased familiarity of items from the pre-exposure phase. Familiarisation allowed items to be learned to some degree, which in turn contributed to the superior recall observed under these conditions.

6.2.4 A comment on the studies of Stuart and Hulme (2000) and Saint-Aubin and Poirier (2005)

There are some issues to be noted about the studies under consideration here. It is possible that both experiments failed to detect some effects because of reduced power. Saint-Aubin and Poirier (2005) used 15 participants per condition, while Stuart and Hulme (2000) used 22 participants per condition. The pair-wise familiarisation and item familiarisation comparisons were between-subjects in the Saint-Aubin and Poirier (2005) study while they were within-subjects for Stuart and Hulme (2000). Furthermore, the failure by Saint-Aubin and Poirier (2005) to detect some interactions in their analysis implies a high level of variability in their data. Consequently, the non-significant trend of a smaller frequency effect in the item memory scores for pair-wise familiarisation against item familiarisation may have been a result of insufficient power. Although not abolished, a reduction in frequency effect for pair-wise familiarisation would be evidence for some role of associative links in the recall of presented items. On the other hand, Stuart and Hulme (2000) failed to find a difference between item familiarised (alternating) and unfamiliarised lists for LF words, despite a trend towards better item memory for item familiarisation when compared to unfamiliarised stimuli. The lack of difference in this case may have also been a product of insufficient power (although an effect, if it did exist, could also be argued to reflect a weakened co-occurrence effect as found in the recall of alternating frequency lists - Hulme et al., 2003). It is conceivable, given the nature of the procedures followed in these experiments that item familiarisation and co-occurrence effects were generated in each case, but were not differentiated because of the variability of the data sets, and because these effects were tested as between-subjects comparisons.

Nevertheless, Saint-Aubin and Poirier (2005) had made a counter claim to the proposition that the standard HF advantage in serial recall was a product of the greater associative links between items (Stuart & Hulme, 2000). They argued that the item-specific account of reconstruction (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999) was the explanation more capable of accommodating their results.

6.3 Item distinctiveness as a factor in the frequency effect in serial recall

6.3.1 Saint-Aubin and LeBlanc (2005)

Saint-Aubin and LeBlanc (2005) argued that alternative interpretations to the abolition of the frequency effect in alternating lists, other than the associative links hypothesis, existed if item distinctiveness was considered a key element in recall. They nominated two classes of account, distinctiveness-only and distinctiveness plus item processing, based on distinctiveness arguments proposed to explain the abolition of the wordlength effect in alternating lists of short and long words (Hulme, Surprenant, Bireta, Stuart, & Neath, 2004). In these experiments, mixed lists were also found to be recalled equally well to pure lists of short words. A distinctiveness-only explanation argued that the likelihood of recall would be affected by how distinctive an item was from its near neighbours in psychological space. Short words are considered more distinctive from short neighbours than longer words are from their neighbours (Hulme et al., 2004; Neath & Nairne, 1995), accounting for the wordlength effect in pure lists. The mixing of wordlengths would result in the enhanced distinctiveness of all items in the list and support recall performance that matched that for pure lists of short words.

Saint-Aubin and LeBlanc (2005) extended this argument to the word frequency effect and based their reasoning on Neath (1994, as cited in Neath & Nairne, 1995) who used the Feature model to successfully reproduce the concreteness effect by assuming concrete words to be more distinctive than abstract words. Saint-Aubin and LeBlanc (2005) argued that due to the greater familiarity of HF than LF words, the recall of contexts in which words arise and the number of exemplars recalled should be greater for HF than for LF words, and item definitions of HF items should be better produced from long-term knowledge than LF words. These factors would contribute to the greater distinctiveness of HF items in psychological space.

The authors also offered a modified view of this explanation that included elements of both distinctiveness and item-based reconstruction. This position was derived from the work of Cowan, Baddeley, Elliott, and Norris (2003) who suggested that a combination of effects might be operating to eliminate or reverse the wordlength effect in mixed lists[1]. They argued that the representativeness of list items by type might influence what items would be emphasised in processing; specifically, increased recall performance associated with underrepresentation might reflect selective mnemonic activity. In such circumstances differential encoding and maintenance might be given to items in mixed lists that otherwise would be difficult to recall.

Research into free recall with mixed lists had shown that HF words are better recalled than LF words in incidental learning conditions but that the reverse occurred when participants had knowledge of an upcoming memory test (Cowan et al., 2003; Gregg, Montgomery, & Castano, 1980; M.J. Watkins et al., 2000). Saint-Aubin and LeBlanc (2005) argued that these results are consistent with the view that individuals adopted a conscious strategy to differentially process items on the basis of their relative distinctiveness (but see section 5.5, Morin et al., 2006).

6.3.2 Predictions

The authors used a serial recall task where trials contained frequency-based isolates embedded into an otherwise pure list to test the validity of each of these hypotheses (co-occurrence/associative links versus distinctiveness-only versus item-reconstruction plus distinctiveness). They reasoned that the associative links hypothesis would offer a number of predictions. Firstly, an LF isolate would be better recalled than an LF item in a pure list because of the greater strength of associative links between the LF isolate and the HF items in the list. However, the recall of this

[1] Note that Cowan et al. (2003), in contrast to Hulme et al. (2004), had retained a wordlength effect when the number of short and long words were equated in a list. However, subsequent research affirmed the abolition of the word length effect in mixed lists (Bireta, Neath, & Surprenant, 2006; Hulme et al., 2006) and Cowan et al.'s (2003) stimuli were shown to possess a confound between the wordlength of stimuli and other attributes known to influence recall performance (e.g. imageability, Bireta et al., 2006).

LF item should be lower than for an HF item in an HF pure list, as the associative links between items in pure HF lists should be the greatest of all conditions. The HF isolate in the LF list should be marginally better recalled, or perhaps recalled at a similar level to the LF item in the LF list, because the associative links the HF item has with the remaining LF items in the list would be weak.

In contrast, a distinctiveness-only account would propose that the LF isolate in the HF pure list would be better recalled than the HF item in the HF list, because the LF item would be more distinctive than its surround. Correspondingly, an HF isolate in a list of LF items would be, by this account, equally distinctive against a background of LF items, and therefore the relative superiority of the recall for the HF isolates should match that found for the LF isolates.

The distinctiveness plus reconstruction account would offer a modification on the latter set of predictions; with respect to the recall of pure lists, HF words should be better recalled than LF words, due to the greater accessibility of their long-term representations during item retrieval. However, in circumstances where an LF isolate is present in an otherwise HF list, the usual HF superiority is masked by a distinctiveness effect operating on the isolate. Additionally, the LF item may, according to this explanation, undergo greater processing during study because of its frequency-based uniqueness within the list context. Therefore the LF item is expected to match or surpass the recall of a corresponding HF item in a pure list. When an HF isolate is inserted into a pure LF list, distinctiveness should operate to enhance recall of the item when compared to that for a LF item in a pure list. However, as the surrounding items are LF, and may therefore have greater processing demands relative to HF items, recall of the HF isolate may be impaired due to resource limitations when compared to the recall of an HF item in a pure list, as has been observed in free recall (e.g., May & Tryk, 1970).

6.3.3 The experiment

The experiment involved the serial recall of pure and mixed (isolate) lists. Lists designated as mixed lists were formed by interchanging items in HF and LF pure lists at a given serial position.

Results of an analysis across serial positions where isolates had been inserted revealed effects of frequency, list type (pure, HF background, LF background) and isolate position (2-5). A frequency by serial position interaction was significant and driven by a difference between frequency types for all medial positions except the last (position 5). A frequency by list type interaction was also present, reflecting different effect sizes for the frequency effect in each of the pure, HF background and LF background lists. All list types were shown to possess significant frequency effects.

These results confirmed the basic finding that HF words are better recalled than LF words in pure lists. The authors argued that predictions derived from the associative link hypothesis received only partial support from the isolate conditions but that, based on previous research identifying the locus of the frequency effect in the memory for items, the analysis of item data would be more telling.

6.3.3.1 Item analysis

Analysing the item error data revealed a frequency effect in item memory for pure lists, and lists where the background frequency was low, although the effect was much smaller in this case; memory for HF words was better than LF words in these instances. However, when the background frequency was high, there was no difference in item memory for HF and LF items. In short, when an LF item was inserted into a HF list, its item memory performance mimicked that for the other HF items in the list. When a HF isolate was inserted into a LF list, performance for item memory was hindered by the presence of the LF surround, and although item memory for the isolate remained better than memory for the background items, this effect was small.

6.3.4 Theoretical interpretations of the data

Saint-Aubin and LeBlanc (2005) stated that the associative links hypothesis could accommodate the difference observed in the pure list condition, and the smaller difference witnessed in the HF isolate condition, but had difficulty with the LF isolate condition abolishing the frequency effect. According to the authors, the associative links hypothesis would predict that the stronger inter-tem associations between HF than LF words in pure lists would yield greater supporting activation, in terms of mutually excitatory connections, and maintain long-term representations in a more retrievable condition. However, when an HF isolate was embedded into an LF list, the

HF item would not benefit from strong supporting activation, as the associative links between items are much weaker; memory performance for the isolate should be poorer than in the pure condition, and in fact at a similar level to a LF item in a pure list. The authors conceded however, that HF words might be more likely to have co-occurred with LF items, than LF items with each other, even if LF have a very low occurrence. Accordingly, a small effect for the isolate was consistent with the hypothesis.

Saint-Aubin and LeBlanc (2005) argued that a greater challenge to the associative link hypothesis was the finding that an LF isolate in an HF list was not recalled less often than an HF item in a pure list. If item co-occurrence was important in determining the strength of the supporting activation among list items, then the presence of an LF item, which is rare by its own account, should lessen the strength of association, and therefore the availability of long-term representations during reconstruction. Yet, no difference in item memory was observed.

In contrast, Saint-Aubin and LeBlanc (2005) asserted that the data was consistent with the modified distinctiveness-based explanation outlined earlier. Distinctiveness alone could account for the observations that an HF isolate is better recalled than the LF background items while an LF isolate recalled at least as well as the background items, as in the latter case the LF item is also ‘distinctive’ by virtue of its isolation. However, as HF isolates were not better recalled than HF items in pure lists, as a distinctiveness account would predict, Saint-Aubin and LeBlanc (2005) reasoned that an additional mechanism must be present.

The authors argued that combining distinctiveness with a reconstruction element yielded a more complete account of the data. According to this position, HF words are better recalled than LF words in pure lists because the LTM representations of HF items are more accessible and more distinctive than representations for LF words. For the HF isolate condition, the effect of distinctiveness would be offset to some degree by a reduction in the effectiveness of HF item reconstruction that would arise due to the greater processing costs associated with the background LF items (M.J. Watkins et al., 2000). Finally for the LF isolate condition the effect of the distinctiveness of the isolate against its background would overcome the relative inefficiency in reconstructing the item at retrieval.

6.3.5 Processing demands from item-specific redintegration combined with item distinctiveness

Saint-Aubin and LeBlanc (2005) stated that their preference for retaining the item-based reconstruction account with the distinctiveness principle, rather than merge distinctiveness with the associative link account (as this would also explain the findings), was that the item-based account appeared to explain a broader range of phenomena. For example, the lexicality effect was argued to be an item-based effect that cannot be explained using an associative links hypothesis (Hulme et al., 2003). As lexicality could be viewed as the limiting case of word frequency (as a nonword is a phonological form of zero frequency) it was argued that retaining the item based explanation, in an overall sense, would provide greater parsimony.

Therefore, Saint-Aubin and LeBlanc (2005) contended that the list, or context-based, elements determining performance levels in the recall of HF or LF words in pure or isolate conditions respectively, refer to differences in the distinctiveness of items in psychological space (Hulme et al., 2004; Neath & Nairne, 1995), and the overall efficiency of processing for list items as a function of the resources they demand (Cowan et al., 2003; M.J. Watkins et al., 2000). However, these authors did not discuss in detail how their approach might accommodate the finding of an abolished frequency effect with alternating lists, or if lenient scoring is considered, a small advantage to HF words (Hulme et al., 2003). In addition, their assertion that the greater threat to the associative links hypothesis was the failure to determine a performance decrement between LF isolates and HF items in pure lists, and not the presence of a clear difference between HF isolates and LF words in pure lists, suggests that Saint-Aubin and LeBlanc (2005) interpreted the co-occurrence argument as a directional mechanism, not as a non-directional and non-specific effect across list items as originally proposed (Hulme et al., 2003; Stuart & Hulme, 2000).

6.3.5 A second isolate study

Roodenrys (unpublished) conducted a study involving the insertion of HF or LF isolates within otherwise pure lists. The recall of pure lists of HF and LF was examined in addition to the recall of lists of LF or HF words containing an HF or LF isolate respectively. Each isolate would appear in position three or four in a list and was yoked to the same serial position in the pure lists. The dependent measure was the proportion of trials the isolate was recalled in each condition, collapsed across serial positions three and four. The pattern of results in this experiment replicated the results determined by Saint-Aubin and LeBlanc (2005). Specifically, an LF isolate was recalled as well as an HF word in a pure list, while an HF isolate was better recalled than an LF word in a pure list, but not as well as an HF word in a pure list. In defense of the associative links or item co-occurrence explanation of these results, the non-significant trends of the recall of LF isolates (see Saint-Aubin & LeBlanc, 2005) were consistent with the possibility of poorer recall when compared to an HF item in a pure list. It is possible that poorer recall for LF isolates would require greater power to be detected.

6.4 Word frequency effects on phoneme migration in mixed lists of words and nonwords

6.4.1 Lexical-semantic effects on phoneme migration in serial recall

Jefferies et al. (2006a) expanded on an unpublished study by Knott and Monsell manipulating the number of words to nonwords in a list and the frequency and imageability of words, by comparing the recall of mixed lists with pure lists of words or nonwords in two experiments. Knott and Monsell (as cited in Jefferies et al., 2006a) had found that the level of phoneme migrations in word recall was influenced by the ratio of words to nonwords in a list and the lexical-semantic properties of the words. Phoneme migration had also been observed in patients with semantic dementia (Patterson et al., 1994). These individuals display selective lexical-semantic contributions in recall performance; they produce greater recall for words where understanding is preserved than for words where meaning has been lost (Jefferies, Jones, Bateman, & Lambon Ralph, 2004, 2005; Knott, Patterson & Hodges, 1997, 2000; Majerus, Norris, & Patterson, 2007; Patterson et al., 1994). Semantically impaired patients without progressive disease, for example patients whose impairments were acquired after cerebrovascular accident (Forde & Humphreys, 2002) and herpes simplex encephalitis (Caza, Belleville, & Gilbert, 2002), have also demonstrated this pattern. Additionally, these patients are more likely to make phoneme migration errors for words they do not understand than those they do. That is, errors patterns indicate that the identity of words can be disrupted by the intrusion of phonemes from other list items.

A second context where phoneme migrations are reported involves the recall of nonwords by normal participants (Treiman & Danis, 1988 – see also Ellis, 1980). This contrasts with the recall of words, where healthy participants exhibit order errors involving the exchange of item position of whole words, but not word fragments. Serial recall tasks that place emphasis on memory for order by using restricted stimulus sets in the construction of experimental trials (e.g. Gathercole, Pickering, Hall, & Peaker, 2001), and maximize familiarity of the items, report order errors to be a substantial proportion of all errors.

6.4.2 The semantic binding hypothesis

Patterson et al. (1994), on the basis of the comparison between recall of words by semantic dementia patients and the recall of nonwords by normal participants, proposed the semantic binding hypothesis. According to this position, phonological coherence, the intactness of phonological representations, is dependent not only on the links between sublexical elements resulting from the repeated activation in speech production and comprehension, but also on ‘constraints’ determined by the semantic representations co-activated with the phonological representations. The absence of semantic constraints in a list of items is argued to reduce the coherence of the phonological representations and induce phoneme migration errors. Patterson et al. (1994) adopted a psycholinguistic model that assumed the existence of semantic and phonological representations in LTM but not lexical representations; lexicality would be determined from differences in associations between frequently co-occurring phonemes and their links with semantic representations. This position therefore viewed STM as an emergent property of the language system.

6.4.3 Predictions

Jefferies et al. (2006a) argued that semantic binding and redintegration accounts (item-specific and item co-occurrence) were difficult to tease apart on an empirical basis. Both of these views argue for lexical-semantic influence on recall, and suggest that these influences operate on item rather than order information. The semantic binding hypothesis states that lexical and semantic factors influence phoneme migrations, and therefore impact the accuracy of the item information recalled. Redintegration assists the recall of an item and therefore the phonemes that comprise it, but not the order in which the item is recalled. However, these positions were argued to differ regarding predictions they make for the rates of phoneme migrations for differing levels of lexical factors. According to semantic binding, phoneme migrations should occur less for words than nonwords, because of the constraint lexicality imposes on the coherence of the phonological representation associated with a word.

In contrast, redintegration theory would not offer a prediction regarding the differences of phoneme migration between words and nonwords. Jefferies et al. (2006a) suggested that redintegration may be able to overrule phonemes that have moved across items when the study material are words, and mask evidence of such movement. The inability of redintegration to overrule migrations in nonwords might explain, from this perspective, the lexicality effect[2]. Despite this possibility, given the late-stage nature of redintegration, and the distinction between short-term phonological and long-term lexical-phonological representations, the redintegration hypotheses were argued by Jefferies et al. (2006a) to predict no difference in phoneme migration between words and nonwords.

In terms of overall recall, these authors argued that the presence of words and nonwords within the same list might lead to intermediate levels of recall for each, as according to the semantic binding hypothesis, the degree of phonological breakdown and recombination for items in a list would be a tradeoff between the capacity of lexical-semantic representations to hold phonological representations of words together, against the increased activity of migrating phonemes from nonwords in the list. This prediction contrasts with item-specific redintegration that would assume better recall for words than nonwords because the availability of LTM representations increases the likelihood of correct recall. In contrast, Hulme et al. (2003) showed that the lexicality effect was smaller for alternating lists of words and nonwords. Words in mixed lists were recalled as well as words in pure lists, however nonwords were better recalled in mixed than pure lists.

---

[2] However, this interpretation of redintegration differs from those of other researchers; for example, Thorn, Gathercole and Frankish (2005) assumed redintegration does not alter the remaining features of a degraded trace but builds on them.

6.4.4 Phoneme migration in mixed lists of words and nonwords

In a first experiment, normal participants were presented lists containing unpredictable ratios of words to nonwords, in unpredictable arrangements where items were drawn from open item pools. The frequency and imageability of words in these lists were factorially manipulated to test whether lexical-semantic properties made an impact on the level of phoneme migration in both words and nonwords.

In terms of serial recall, lexicality, frequency and imageability all made contributions to performance. Lexicality and frequency effects were greatest when more words than nonwords were presented in a list. In contrast, imageability effects were greater when the majority of list items were nonwords.

6.4.4.1 Item analysis

Order error measures, that is, whole item transpositions and phoneme migrations, were expressed in terms of percentages of total item recall and total phoneme recall, respectively. Across lexicality and frequency/imageability conditions, the number of whole item errors was much greater than whole order errors. There were however, more whole item transpositions for words than nonwords, while more item errors were made with nonwords. At the phoneme level nonwords were more likely than words to induce either order or identity errors. These observations implied a greater coherence for words over nonwords.

6.4.4.1.1 Frequency

Additionally, at the phoneme level, order and identity errors were less for words and nonwords when HF, rather than LF, words were included in the list. Furthermore, as phoneme migrations did not typically take place as whole item transpositions, frequency was seen to influence the degree of individual phoneme migrations. The phonemes of HF words were more likely to remain intact, and this resulted in fewer phoneme migrations in nonwords. Jefferies et al. (2006) interpreted these effects collectively to imply that word frequency influences the stability of the phonological trace of the entire list.

6.4.4.1.2 Imageability

Imageability produced smaller effects than word frequency. Imageability influenced the number of phoneme identity errors in words, where high imageability items were less likely than low imageability to be associated with the recall of incorrect phonemes, but phoneme order errors in words were not influenced by imageability. However, imageability did influence the numbers of phoneme identity and phoneme order errors for nonwords; fewer errors were made when high rather than low imageability words were presented with nonwords.

6.4.4.1.3 Lexicality

Fewer phoneme migrations took place for both words and nonwords when the ratio of words to nonwords was higher. The proportion of words in the list influenced the number of phoneme identity errors for words, but not nonwords; fewer errors occurred with higher word to nonword ratio lists. These results were seen as consistent with the proposition that the phonemes in words are bound together and resist the intrusion of migrating phonemes from nonwords, and that there would be less opportunity for phonemes from nonwords to reposition in the list when more items are words.

The migration of phonemes in nonword recall was impacted by the number of co-presented words in the list, and their frequency and imagability. Lists with more words, or words of greater frequency and/or imageability produced more stable nonword recall.

6.4.5 Phoneme migrations in pure lists of words and nonwords

A second experiment tested the recall of pure lists of words and nonwords and used the same stimuli as the first experiment. The data was, in turn, included in a between-groups analysis of recall comparing performance on mixed and pure lists. According to the semantic binding hypothesis, the recall of words and nonwords in mixed lists would be at a level intermediate to the recall of words and nonwords in pure lists. This averaging would occur because the presence of nonwords in the mixed lists destablises the phonological coherence of word representations when compared to pure lists of words, while the presence of words in mixed lists limits the opportunities for fragmented phonemes from nonwords to migrate to new positions when compared to pure lists of nonwords.

Recall for words was better in pure than mixed lists, while recall for nonwords was better for mixed than pure lists. The effects of the lexical-semantic factors of frequency and imageability did not change across list composition (i.e. pure versus mixed lists).

6.4.5.1 Item analysis

Once more words were more likely to be recalled in the incorrect list position than nonwords. This effect was the same for pure and mixed lists. However, item identity errors were greater for words in mixed than pure lists, while less for nonwords for mixed than pure lists. In pure lists, phonemes were more likely to migrate to new positions in nonword than word lists. The difference in phoneme migration between words and nonwords was greater for pure than mixed lists, and was due to the reduced phoneme migrations with words in pure, relative to mixed lists. Phoneme identity errors were less frequent for words in pure than mixed lists, and more frequent for nonwords in pure than mixed lists resulting in a lexicality by list type interaction. Word frequency influenced phoneme order errors and phoneme item errors in pure lists but these effects did not interact with list type. Imageability was found to alter the extent of phoneme item errors, but not phoneme order errors in pure lists, copying the pattern observed in mixed lists.

A final analysis looked at the proportion of items incorrectly recalled that were words, as an indication of sensitivity to a class of response. There was no difference in the proportions of incorrect real word responses corresponding to word or nonword positions in mixed lists. In pure lists however, the number of erroneously recalled words was far greater for words lists than nonwords lists. These results indicated that knowledge of the lexical status of items guided the type of responses made in pure but not mixed lists. Furthermore, no difference in the proportion of incorrect word responses was found amongst the mixed lists of differing ratios of words to nonwords, but both the frequency and imageability of the words in these lists did influence how likely words were to be used as responses; when high imageability and HF words were included in the list, both words and nonwords were more likely to be replaced by the ‘recall’ of intruding words.

The recall of words and nonwords in mixed lists was found to be intermediate when compared with the recall of pure lists of words or nonwords. The poorer recall for words in mixed lists contrasted with Hulme et al. (2003) who found equivalent performance for words in mixed and pure lists. Jefferies et al. (2006a) attributed this disparity partially to differences in the structures of the lists used in the respective experiments. Hulme et al. (2003) used alternating lists that were obvious in structure and responses may have been cued, in part, by knowledge of the lexical status of items across serial positions. In contrast, the positions of words and nonwords in the mixed lists were not predictable, limiting the usefulness of strategic knowledge regarding list structure. Jefferies et al. (2006a) also considered that phoneme migrations in lists where nonwords outnumbered words could also have contributed to the reduction in word recall in mixed lists.

6.4.6 Word frequency influences on phonological coherence

Therefore, lexical-semantic factors were shown to influence the rate of phoneme order and phoneme identity errors, and accordingly the coherence of phonological STM representations. Phonemes in words were more likely to be outputted together than the phonemes in nonwords; phonemes forming nonwords displayed a greater tendency to migrate to other positions in the list, or not be recalled at all. The nature of influence of lexical-semantic factors in these experiments identified that LTM knowledge contributed to the integrity of item information, as had been indicated by other studies (Gathercole et al., 2001; Hulme et al., 1997; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 1999, 2000). When entire items did migrate together, they were more likely to be words, producing a lexicality effect for order memory (Saint-Aubin & Poirier, 2000).

List composition was found to impact on the magnitude of phonological errors in the recall of words. These errors were found to occur at a higher rate in mixed than pure lists, and were higher in mixed lists when more rather than fewer nonwords were included. Therefore, the presence of nonwords was found to disrupt the phonological representations of the words, arguably through the migration of unbound phonemes from the nonwords to the words.

The recall of nonwords was also influenced by frequency, imageability and lexicality. In particular, nonwords were found to possess greater recall accuracy when mixed with words than when presented in pure lists. The authors suggested that the greater binding between phonemes of words in mixed lists limited the opportunities nonword phonemes had to reposition across list items.

That lexical-semantic factors were influential in determining the rate of phoneme migration in serial recall was interpreted as direct evidence of the semantic binding hypothesis (Patterson et al., 1994). Less phonologically coherent items (for example nonwords or LF words) were argued to induce greater levels of fragmentation and recombination than more strongly bound items (HF words) in recall. Accordingly, recall performance of a particular item would be dependent on the context in which it was presented and not determined solely from item-specific properties.

In comparison, the item-specific view of redintegration (Hulme et al., 1991; Hulme et al., 1997; Schweickert, 1993) did not explain the observed patterns of phoneme migrations across conditions, as it is limited to explaining the restoration of phoneme identity through reconstruction based on long-term representations of items. Jefferies et al. (2006a) claimed that in pure lists of words, missing or intruding phonemes could be replaced by redintegration, limiting the outputted phoneme migrations for this condition. However in mixed lists, redintegration would not be capable of reinstating the identity of nonwords, especially when phonemes from words intruded into their phonological representations. Furthermore, under redintegration, the identity of nonwords would not be sensitive to the level of lexical-semantic properties of co-presented words, as this information would not be used in item reconstruction. Lastly, the properties of list items would not alter across list contexts because each item would be redintegrated independently.

Co-occurrence versions of the redintegration hypothesis (Hulme et al., 2003; Stuart & Hulme, 2000) were seen by Jefferies et al. (2006a) to explain some results but not others. While the argument that item co-occurrence determines the success of redintegration through the availability of item representations was consistent with the intermediate recall of words in mixed lists, it could not explain why the frequency and imageability of words changed the recall of nonwords. Jefferies et al. (2006a) argued that the co-occurrence of words with nonwords, regardless of their frequency or imageability levels, should be similar, that is close to non-existent, and so effects on nonwords should be also be similar if item co-occurrence was the mechanism responsible for recall performance.

6.5 Associative and item-based frequency effects in serial recall

Tse and Altarriba (2007) argued that an orthogonal manipulation of word frequency with LSA-cosine (Landauer & Dumais, 1997), as a measure of inter-item associativity would provide a test of the associative-links hypothesis (Stuart & Hulme, 2000), and identify contributions from each variable to recall performance, should they exist. Based on the work of Deese (1960), if inter-item associativity was controlled, then the frequency effect should be abolished. As Hulme et al. (2003) had found a correspondence between the LSA-cosine measure with the serial recall performance (Experiment 1), Tse and Altarriba (2007) considered this measure to be an appropriate index of inter-item associativity.

The authors controlled pair-wise semantic association by limiting the presentation of items to dedicated lists. The LSA cosines were generated on the basis of the 'General reading up to 1st year college' corpus. The stimulus sets for each frequency and LSA-cosine condition were matched on word length, number of syllables, forward and backward association strength and context availability.

The experiment produced a standard frequency effect; HF lists were better recalled than LF lists. An effect of LSA-cosine was also found and additionally this manifested as an interaction between frequency conditions. Positive effects of LSA-cosine of similar magnitude were observed for HF and LF items for early list positions, while the effect of LSA-cosine was present only for LF lists in the later list positions. An order effect, where more errors occurred with low than high LSA-cosine lists and more errors in LF than HF lists was found.

Tse and Altarriba (2007) argued that the results contradicted those of Stuart and Hulme (2000) and the associative links hypothesis, as the contribution of LSA-cosine was dependent on the level of frequency of items; the authors claimed that if recall was dependent on the level of inter-item association, improvement in recall should have been observed for both high LSA-cosine conditions. Furthermore, Tse and Altarriba (2007) proposed that the presence of the frequency effect, regardless of the level of LSA-cosine, indicated it was dominated by an item-specific contribution that could be explained using item-based redintegration (Hulme et al., 1997; Saint-Aubin

& Poirier, 1999). The effect of inter-item association in terms of LSA-cosine would be of greater benefit to the recall of LF than HF words, because of the greater relative accessability of long-term representations of HF words in item-based redintegration. Finally, the authors claimed that the nature of the observed interaction, that is, an effect of LSA-cosine for both HF and LF words for early list items, with an effect of LSA-cosine only for LF words for later list items, was consistent with the impact of output delay and proactive interference in the second half of the list, although they offered no qualification as to why this would selectively impact inter-item associativity.

As previously noted, in terms of the application to serial recall, LSA-cosine measures are not without issue. In addition to word co-occurrence (Landauer, Foltz, & Laham, 1998), they quantify how likely words can substitute for one another in passages of similar context. While Hulme et al. (2003) considered that they could therefore be used as an index of the degree to which items are associated in memory they also conceded that the measure was more complex than other estimates of associativity.

The notion of LSA-cosine therefore reflects a measure of semantic relatedness more than co-occurrence *per se*. Words that co-occur in language will not always act as appropriate substitutes for each other in text passages, while those items that do might not always appear in close proximity to each other in natural language use. Therefore, at best it would appear that LSA-cosine is a partial measure of inter-item associativity, and one possibly better classified as a measure of semantic relatedness. Furthermore, as Section 5.4.8 demonstrated, the measures are corpus-dependent. Accordingly, it is likely that the measures, as used in serial recall, under-represent the extent of item co-occurrence in language and the item-item associativity between sets of words.

Therefore it is probable that the LSA-cosine measures did not match lists on inter-item associativity. As Hulme et al. (2003) demonstrated by the comparison between pure and alternating lists, the recall of pure lists is not diagnostic of the effects of inter-item association. In consequence, it is highly probable that the frequency effects observed in this experiment were not dominated by item-specific differences between HF and LF words as Tse and Altarriba (2007) claimed.

6.6 Conclusion

The studies reviewed in this chapter mounted arguments against item co-occurrence/associative links as the preferred explanation of the frequency effect in serial recall. However, these arguments have themselves been formed from differing theoretical perspectives. The first position argued for the retention of the item-specific redintegration account based on (i) the claim that Stuart and Hulme (2000) inadvertently captured an effect of familiarity moderated by set size and mislabeled this difference as item co-occurrence (Saint-Aubin & Poirier, 2005), (ii) the position that recall performance with isolates could be best explained by retaining the modal interpretation of redintegration and considering item distinctiveness as a second property active in recall (Hulme et al., 2004; Hulme et al., 2006, Saint-Aubin & LeBlanc, 2005) and (iii) the contention that associativity, as measured by LSA-cosine was selective in its effect. Alternatively, Jefferies et al. (2006a), on the basis of the micro-analysis of recall data identifying phoneme migration patterns and memory for identity and the item and phoneme levels, recommended the adoption of the semantic binding hypothesis and the concept of phonological coherence as the principle governing recall. Within this psycholinguistic framework, word frequency would modulate the phonological coherence of the list trace. The greater the frequency of list items, the less fragmentation and reordering of the phonological information in the list, resulting in fewer item errors at recall.

Finally, as a point of consistency, these studies, as with the vast majority of those reported in previous chapters demonstrated that the effect of frequency operates on item memory. Therefore, the effect, in whatever form it manifests, operates on an existing phonological trace, and is unlikely to possess any substantial role in its formation (Hulme et al., 2003).

Chapter 7

Order Effects and the Loci of Effects for LTM Variables in Serial Recall

7.1 Introduction

Word frequency has most often, in the context of serial recall, been considered a variable that exerts dissociable effects on item and order information (Nairne & Kelley, 2004). The general finding across frequency-based serial recall experiments is that word frequency impacts the level of item, but not order memory (Allen & Hulme, 2006; Hulme et al., 2003; Morin et al., 2006; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000). The purity of this distinction however has implications for how the frequency effect is theoretically described and the point(s) at which frequency effects are assumed to operate. It is possible that small order effects have gone unnoticed because of lack of power in previous studies (e.g. Hulme et al., 1997). This chapter reviews the evidence for the presence of order effects with word frequency, and more widely LTM variables. It also considers the related issue of the locus of the word frequency effect.

7.2 Order effects with LTM variables in serial recall

Order effects are usually reflected as differences in the number of whole-item transpositions, expressed as a proportion of all items recalled regardless of order (Murdock, 1976; Poirier & Saint-Aubin, 1996) across experimental conditions (but see Jefferies et al. (2006a) for phoneme-order analyses that relate to item memory). Evidence for order effects involving LTM variables is not always found and accordingly, those circumstances that produce order differences might reflect unknown idiosyncrasies with the choice of stimuli, or occur because of random variation in the item and order memory abilities in participant pools. Order effects have been determined for concreteness (Allen & Hulme, 2006; Tse & Altarriba, 2009), word pleasantness (Monnier & Syssau, 2008), word frequency (Hulme et al., 2003; Tse & Altarriba, 2007) and semantic association (Tse & Altarriba, 2007), and in each of these cases better order memory has occurred for stimuli that also facilitate better item memory. In contrast, order memory has been shown to be worse for semantically similar items (e.g. Saint-Aubin et al., 2005). In this instance, similarity has been argued to facilitate item memory while confusing the order of recall. Order effects have also been observed for lexicality (Hulme et al., 2003; Jefferies et al., 2006a; Saint-Aubin & Poirier, 2000), although the direction of effect in these experiments has not been consistent. In mixed lists, order effects have been observed with manipulations of word frequency and phonological neighbourhood size (Roodenrys et al., 2002), and lexicality (Jefferies et al., 2006a).

The order effects for LTM variables, when established are small, suggesting that they are unlikely to be a dominant influence on memory for serial order. However, the existence of these effects for LTM variables suggest that their influence in serial recall tasks might occur before the point of recall, in contrast to the arguments of late-stage redintegration theories (Monnier & Syssau, 2008; Saint-Aubin, Ouellette, & Poirier, 2005). Consequently, proponents of two-stage accounts of serial recall have conceded that short-term traces may encode semantic information as well as the phonological properties of items (e.g. Saint-Aubin et al., 2005).

7.3 Factors influencing measurement

A credible demonstration of the absence of an effect necessitates manipulation of key variables at a level that provide a reasonable opportunity for the effect to arise (Frick, 1995). In STM experiments, the strength of manipulation of item/order information can be modulated by such factors as the choice of stimuli and/or the selection of the task. Experimentation that uses a limited design can produce misleading outcomes. For example, a number of researchers have shown that tasks previously thought to produce null effects with LTM variables (e.g. matching span) may have been insufficiently sensitive due to the restricted nature of the stimuli used limiting the manipulation of item information (e.g. Jefferies et al., 2006b; Romani et al., 2008).

A second measurement concern relates to the interpretation of the nature of effects when they arise. One difficulty in establishing whether LTM variables affect order memory *per se* stems from the realisation that tasks presumed to focus on order information are not process pure (Nairne & Kelley, 2004; Neath, 1997). For example, although serial recall is a task that measures primarily the retention of order information it also requires the retention and use of item information (Saint-Aubin & Poirier, 1996; Murdock, 1976). In a similar vein, performance on tasks assumed to reflect pure order information may be influenced by unrecognised contributions from item information, and this needs to be taken into account when results are interpreted.

7.4 Tasks measuring serial order

7.4.1 Serial order reconstruction tasks

The serial order reconstruction task requires the rearrangement of represented material into the same sequence that was first encountered by a participant. As all items are available at test, and the emphasis is on placing them into their original order, this task has been treated as a pure measure of order memory (Whiteman et al., 1994).

With respect to variables such as word frequency and concreteness, the order reconstruction task has been used to examine order effects in LTM. Whiteman et al. (1994) examined LTM order reconstruction involving pure lists of HF and LF words, and failed to find any effect involving frequency across two experiments, although a non-significant difference for HF words was reported. In contrast, DeLosh and McDaniel (1996; Merritt, Delosh & McDaniel, 2006) determined a frequency effect in pure lists for long-term order reconstruction and no effect for mixed lists with equal numbers of HF and LF words. They replicated the null result of Whiteman et al. (1994) by presenting pure HF and LF lists in a within-subjects design and argued that the failure to detect an effect was due to stimulus cross-contamination within participants. They claimed that these results could be applied to the free recall of pure lists – that better order-encoding occurred between HF than between LF items and this facilitated greater performance for HF lists. However, in mixed lists, an order-encoding advantage for HF items was abolished by the presence of the LF words, and a free recall LF advantage in mixed lists would result from the superior encoding of item information with LF words.

Neath (1997) found, in an LTM order reconstruction task, that concrete words facilitated better performance than abstract words. He argued that because this task was sensitive to concreteness, and concreteness was a variable that influenced item memory, order reconstruction could not be viewed as a pure test of order memory. Alternatively, in STM serial order reconstruction has been used and acknowledged as a task maximising order memory retention when conducted under conditions that minimise the processing and maintenance of item information, for example when lists of digits comprise the test material (Majerus, Poncelet, Elsen, & Van der Linden, 2006; Majerus, Poncelet, Van der Linden, & Weekes, 2008).

Despite this ambiguity, serial order reconstruction is considered to be a task that can demonstrate whether LTM variables impact short-term memory traces prior to the point of retrieval of items (Thorn, Frankish, & Gathercole, 2009). Because there is no requirement to reproduce item information, variables that show little or no effect in serial order reconstruction but affect serial recall would be considered to operate on retrieval processes alone. In contrast, LTM variables that exhibit effects in serial order reconstruction would be argued to influence STM processes at an earlier point.

Thorn et al. (2009) reported on a study that compared the effect of lexicality on serial recall and serial order reconstruction performance. Phonological similarity was also manipulated in these experiments. While phonological similarity produced enhanced item recall and poorer order memory in the serial recall and order reconstruction tasks respectively, lexicality produced the standard effect for item memory in serial recall, but only a small effect in order reconstruction. The authors considered the residual nature of the lexicality effect in order reconstruction to be attributable to the encoding of semantic properties of items in short-term traces that would facilitate better order retention of items with words but not nonwords (Poirier & Saint-Aubin, 1995; Saint-Aubin et al., 2005).

The contrast between performance on serial recall and serial order reconstruction was extended by Thorn et al. (2009) in two further studies that examined word frequency and phonotactic frequency. The serial recall tasks produced the standard effects for word frequency and phonotactic frequency with nonwords. Performance on serial order reconstruction identified that, like the lexicality effect, the word frequency effect was reduced. However, the effect of phonotactic frequency was similar across tasks. Furthermore, these comparisons between serial recall and order reconstruction tasks for lexicality, frequency and phonotactic frequency were replicated in a second series of studies. The authors argued that the smaller effects in serial order reconstruction for lexicality and frequency were consistent with redintegration accounts of short-term recall that saw the effects of these variables as late-stage, while the persistence of a phonotactic frequency effect across tasks suggested that it is involved in STM activity before retrieval. It was proposed that phonotactic frequency might affect the strength or quality of the short-term traces of items, impacting on the resilience of item information to be retained for either serial recall or order reconstruction.

L. Clarkson and S. Roodenrys (personal communication, September 9, 2009) determined that STM order reconstruction was influenced by both word frequency and phonological neighbourhood size, a variable related to phonotactic frequency, in a design that contained a factorial manipulation of both variables. These variables interacted so that order reconstruction performance was greatest for HF, large neighbourhood words, while performance for conditions with small neighbourhood or LF words was similar. The stimuli used in this experiment were taken from Roodenrys et al. (2002; Experiment 1) and were closed sets (n = 16). Therefore, item information was not maximally tested in the experiment. This result supports the possibility that both frequency and phonological neighbourhood size are variables that are active in STM before item retrieval and can contribute to differences in order information retention.

7.4.2 Matching span and serial recognition tasks

In matching span and serial recognition tasks participants determine whether the second of two presented lists is the same or different from the first (Jefferies et al., 2006b). Differences between study and test lists involve the re-ordering of list items. List length varies for matching span tasks while serial recognition procedures test lists at a fixed length. These tasks have been used as a means of determining whether an effect arises prior to or after speech output (R.C. Martin et al., 1999; Walker & Hulme, 1999) but they have also been seen as measures reflecting differences in order memory as they involve judgments regarding the sequence of presented items (Gathercole et al., 2001). Despite this requirement, it is generally acknowledged that in serial recognition and matching span tasks order memory is likely to be confounded with item memory (Walker & Hulme, 1999).

Walker and Hulme (1999) used a matching span task to ascertain whether the effect of concreteness was present prior to an output stage. They reported a null result and argued that the locus of the concreteness effect was specifically late-stage; that is, the effect arose during speech output.

Other researchers have questioned the capability of matching span and serial recognition procedures to capture LTM effects under specific conditions, arguing that insufficiently sensitive tests falsely imply that the loci of effect for LTM variables are limited to a late stage (Jefferies et al., 2006b; Monnier & Syssau, 2008; Romani et al., 2008). For example, matching span tasks with minimal item information demands underestimate the influence of variables that impact item information levels in the stimuli (Jefferies et al., 2006b; Majerus, 2009). Romani et al. (2008) showed that with an open set of items a matching span procedure can produce a concreteness effect and word pleasantness has been found to produce an effect in serial recognition (Monnier & Syssau, 2008). These results stand in contrast to that of Walker and Hulme (1999) who used closed stimulus sets in their test of concreteness.

A lexicality effect in serial recognition has also been demonstrated (Gathercole et al., 2001), but in a pattern similar to serial order reconstruction this effect was found to be small when compared to the lexicality effect in serial recall. Gathercole et al. (2001; Thorn et al., 2009) interpreted the small effect for lexicality in serial recognition to be a function of semantic encoding in the short-term trace that was possible for words but not nonwords.

Lastly, in a novel form of serial recognition task Jefferies et al. (2006b) examined recognition performance in lists that varied in frequency and imageability when order information was changed at item or list levels. Item information was altered by swapping phonemes of list items to produce changes in their identities, while the transposition of whole items tested memory for item order within a list. The authors also compared performance for lists of words and nonwords. This task produced lexicality, frequency and imageability effects at both item and list levels, but effects were greater with item identity than item order change. These results suggest that these LTM effects operate on STM traces much earlier than the point at which items are overtly recalled, and that these effects relate to differences in the quality of item information.

7.5 Process dissociation with semantic similarity and word frequency

Nairne and Kelley (2004) used the method of process dissociation to separate any effects semantic similarity and word frequency had on item and order information respectively. Process dissociation involves the measurement of task performance on an inclusion task, that is, one presumed to depend on both item and order memory, and an exclusion task that tests recall performance when item and order information act in opposition.

The inclusion task for semantic similarity (where similarity was defined in terms of membership of a semantic category) was immediate serial recall and the exclusion task required the immediate recall of all items except the item from a nominated position. The latter task is argued to produce an error (a recall of the item presented in the to-be-omitted position) only when the item is known, but its position in the list has been forgotten. This method also assumes that item and order information are independent. Using this approach, Nairne and Kelley (2004) determined that there was a facilitative effect of semantic category for item information, but no impact on order information in immediate serial recall. Therefore, in STM semantic category might promote the recall of item identity, however the order information of the short-term trace is derived from another source.

In addition, the authors applied process dissociation to the recall of HF and LF words, however as recall in the dissociation tasks followed a filled delay this analysis is most applicable to the state of item and order information in LTM activities. The inclusion condition produced the typical pattern for the serial recall of pure HF and LF lists. In contrast, the exclusion condition was found to be sensitive to serial position but not frequency. Inspection of the errors identified that transpositions or order errors were of the same rate for either frequency condition across the tasks, while omissions were greater for LF words as has been witnessed in serial recall studies (e.g. Hulme et al., 1997; Poirier & Saint-Aubin, 1996). The item and order information estimates obtained from process dissociation revealed better item and order recall for HF than LF items, although the effect was smaller for order memory. Nairne and Kelley (2004) suggested that it may be possible for inter-item associations to be learned more readily with HF than LF words and these facilitate better order information. Furthermore, the small effect size is presumed to explain why LTM order reconstruction results are inconsistent (e.g. DeLosh & McDaniel, 1996; Whiteman et al., 1994).

Therefore, in the context of short-term recall the LTM variable semantic similarity did not produce an order effect, while word frequency, when tested using

LTM tasks, generated a small effect of order. These results would suggest that in STM, effects associated with word frequency are most likely to derive from the influence of item information.

7.6 Other research informing on the loci of LTM effects

7.6.1 Psychophysiological evidence

Evidence from ERP studies have suggested that activation of LTM representations may occur during the encoding and retention of items in STM tasks (Cameron, Haarmann, Grafman, & Ruchkin, 2005; Ruchkin et al., 1999; Ruchkin, Grafman, Cameron & Berndt, 2003). These studies have included tasks argued to elicit and isolate activation of semantic (Cameron et al., 2005) and lexical (Ruchkin et al., 1999) representations in STM contexts. Such observations are consistent with language-based models (e.g. R.C. Martin et al., 1999) that assume LTM representations are involved in STM tasks in an ongoing and integrated manner, and align with empirical studies that have revealed effects of LTM variables prior to the point of recall.

7.6.2 LTM effects on retrieval processes in serial recall

Thorn, Frankish and Gathercole (2009; Thorn et al., 2005) examined item recall errors, in particular the numbers of completely incorrect recall attempts made in a serial recall experiment manipulating word familiarity (comparing English monolingual with French-English bilingual participants), lexicality, frequency, and phonotactic frequency. They applied the logic inherent in a version of the MPT redintegration model of Schweickert (1993) to error data analysed at the phoneme level to establish at what point item information redintegration operates. This analysis, on the basis of inconsistent results across LTM variables, was argued to demonstrate differences in the loci of effects in serial recall.

Gathercole, Frankish, Pickering, and Peaker (1999) had extended Schweickert's (1993) MPT model of redintegration to include a third branch off the initial node. This branch was added to include the possibility that an STM trace might be lost altogether, a possibility that would render redintegration ineffective given there would be no partial information on which to base it. Thorn et al. (2005) reasoned, according to this model, that the number of intact short-term traces bypassing redintegration, and the number of completely incorrect recalls, presumably a result of guessing in the absence of any partial remaining information, should be equivalent (for different levels of LTM variables) if short-term traces are strictly phonological.

However, the possibility that additional completely incorrect recalls could result from redintegration was feasible if it was assumed that the minimum amount of partial information required by redintegration occurred at the sub-phonetic level (i.e. the voicing of phonemes or place of articulation). While it is assumed that redintegration does not alter or abolish existing short-term trace information, sub-phonetic features could be reconstructed to retrieve an item that did not share phonemes with the actual target item. In contrast, if a single phoneme was the minimum unit of information on which redintegration is based, failed attempts to retrieve the target item should contain that information in the response. In this case completely incorrect recalls could not eventuate from redintegration, and differences in the number of partially reconstructed items would reflect the respective contribution of LTM variables across levels. Lastly, interference effects that might alter the partial information on which redintegration operates, leading to ‘apparent’ completely incorrect recalls, partially correct recalls or completely correct recalls would be considered independent from LTM variables and affect all levels of variables equally.

Thorn et al. (2005) classified item recall responses from their serial recall task according to whether they were completely correct, partially correct (containing at least one target phoneme but not all) or completely incorrect recalls of the CVC stimuli presented in lists containing restricted phonemes. They found different recall profiles across the variables examined in this study, particularly with respect to the numbers of completely incorrect responses made. While there was no difference in the level of this error between LF words and high phonotactic frequency non-words (levels chosen to represent the lexicality effect), word frequency, nonword phonotactic frequency and language familiarity produced differences in the numbers of completely incorrect recall attempts made. High frequency words produced fewer completely incorrect responses than LF words, monolinguals produced fewer completely incorrect responses than bilinguals, and the monolingual group exhibited a phonotactic frequency effect such that more completely incorrect responses occurred with low rather than high phonotactic frequency stimuli.

Therefore, the phoneme error patterns for lexicality were consistent with a redintegration process that operated at the level of phonemes, but the other linguistic variables were not. If redintegration was to operate in the same manner across all variables, then the differences in completely incorrect recalls for word frequency, familiarity and phonotactic frequency would need to originate from some other point in serial recall. Thorn et al. (2009) proposed that differences in these errors reflected differences in the strength of the memory trace prior to redintegration as functions of these LTM variables.

Accordingly, Thorn et al. (2009; Thorn et al., 2005) have argued that different LTM variables are involved in multiple mechanisms operating at different stages of serial recall. Phonotactic frequency, word frequency and word familiarity influence the retention of short-term traces as well as the success of late-stage redintegration. Lexicality effects, in contrast, are derived solely from the late-stage reconstruction of partially degraded items.[1]

7.7 Distinct item and order processes

The literature reviewed in this chapter so far suggests that apparent order effects involving lexical-semantic variables reflect the confounding of item memory with order memory in the execution of an STM task, and that such effects arise because at least some item information is available at an early stage in the processing of verbal items. Language-based models naturally accommodate the majority of observations that relate to the encoding and maintenance of item information. However, they do not reconcile in an obvious way how item memory and order memory interact in short-term serial order activities.

7.7.1 STM, order and the language processor

Majerus (2009) has argued that, as specified, language-based models (either

[1] More recently Jefferies, Frankish and Noble (2009) have also suggested that serial recall involves separable processes that are strategic (depending on the lexical status of items) and automatic (not consciously controlled).

unitary or two-component) cannot fully account for the full range of observed STM phenomena, because they do not explain how the order of items is managed in STM. N. Martin and Saffran (1997) suggest that order is retained through a sequence placeholder mechanism (see also Romani et al., 2008) although how this occurs is not articulated, and R.C. Martin et al. (1999) assume that order will be managed within the STM buffers, but no details are provided on how this is achieved. Therefore, language-based accounts explain the management of item memory and implicitly assign order memory to an alternate mechanism.

That item and order information derive from distinct processes is not only supported by the small to nonexistent effects on order memory of lexical-semantic variables (e.g. Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999; 2000), but also the dissociated effects of phonological similarity for item and order memory (Fallon, Groves, & Tehan, 1999; Nairne & Kelley, 2004; Nimmo & Roodenrys, 2004). Additionally, Majerus and colleagues have amassed a substantial body of work that supports the item memory/order memory distinction (see Majerus, 2009, for a full review). For example, the association of new word learning in adults to performance on tasks maximising order but not item (i.e. phonological form) retention suggests that STM serial order capacities external to language representations might well be responsible for vocabulary development (Majerus, Poncelet, Elsen, & Van der Linden, 2006; see also Majerus, Poncelet, Van der Linden & Weekes, 2008). An equivalent study with young children demonstrated that the development of item and order retention capacities follow different maturational trajectories, and reinforced the likelihood that separate mechanisms manage item and order information (Majerus, Poncelet, Greffe, & Van der Linden, 2006). Data from neuropsychological patients, namely individuals with a specific genetic cause of velo-cardio-facial syndrome (22q11.2 chromosomal microdeletion) indicate that this group exhibits normal item memory with impaired order memory (Majerus, Van der Linden, Braissand, & Eliez, 2007), while semantic dementia patients produce the opposite pattern (Majerus, Norris, & Pattison, 2007), forming a double dissociation.

7.7.2 A neuroimaging basis to item and order mechanisms

Majerus, Poncelet, Van der Linden et al., (2006) examined fMRI while participants engaged in probe recognition tasks with visually presented words, testing either memory for item or order information. Both tasks activated the left intraparietal sulcus (IPS), while the order condition produced greater activation in the right IPS, right cerebellum, and bilateral premotor cortex. The item condition produced more activation in the superior temporal gyrus (sulcus) and left fusiform gyrus. These areas are thought to be recruited in the processing of phonological and orthographic information, respectively.

Functional connectivity analysis identified that during the order task the left intra-parietal sulcus (IPS) was functionally connected to the right IPS, the bilateral dorsal premotor cortex, the insula and the right cerebellum. The equivalent analysis for the item task found that the left IPS was connected to each superior temporal gyrus.

Therefore, while different brain regions were employed for each type of task, the left IPS was involved in both to a similar degree under conditions where task difficulty was matched. The study confirmed that language processing regions are recruited for tasks requiring the management of item information. The so-called parieto-fronto-cerebellar network identified in the serial order task was nominated by Majerus (2009) as a likely facilitator of the sub-processes responsible for the encoding, retention and recognition of serial order information. The right and left parietal sulci have been previously associated with the processing of magnitude in a number of tasks and it has been suggested that the IPS is the neural substrate responsible for controlling the common representation of number and order, while other research suggests that the right IPS is involved with retrieval of temporal order. Majerus, Poncelet, Van der Linden, et al. (2006) found that the prefrontal areas functionally connected to the left IPS during the order memory condition were dorsal premotor areas. Additionally, Cairo, Liddle, Woodward and Ngan (2004) found that bilateral dorsolateral premotor areas are active at encoding and during the retention of items in a verbal STM task.

Accordingly, Majerus (2009; Majerus, Poncelet, Van der Linden, et al., 2006) assert that the neural substrates identified for the processing of item and order information are distinct; item information is managed by the areas in the temporal lobe associated with language processing and order processing is conducted by the prefrontal and right parietal regions. The left IPS appears to possess more generalised functioning, and is involved in the processing of both item and order information. Specifically, it is proposed that the left IPS functions as an attentional modulator that controls the level of attentional resources dedicated to item and order processes in response to task requirements and resembles the focus of attention within the activation of LTM referred to by Cowan (1995, 2005). Furthermore, this region is activated in non-linguistic tasks and therefore appears to be involved across multiple memory activities.

7.7.4 A model of STM incorporating separate STM capacities and the language processor

Language-based models that assume the existence of separate STM buffers containing lexical-semantic information in a temporary form (R.C. Martin et al., 1999; Romani et al., 2008) have been motivated by neurological cases where STM impairment has been coincident with preserved language processing (R.C. Martin & Lesch, 1996). However, Majerus (2009) performed a review of the STM performance and single word processing of those neuropsychological cases available in the literature and found that cases with unaffected language processing had smaller STM deficits than those patients who could not perform single word tasks. He suggested that this relationship implied that the degree of STM impairment mirrored the extent to which temporary activations in LTM could be maintained; more severe impairment in terms of greater decay rates for activated representations would lead to greater STM impairment. Slower decay rates associated with better short-term recall might be sufficient to maintain representations in single word processing. Under this hypothesis, STM buffers would not be necessary to retain copies of item representations, and the management of item information could be contained within a unitary system.

Majerus (2009) considers that a hybrid model adopting the language processing features of interactive activation frameworks (e.g. N. Martin & Saffran, 1992, 1997) while retaining short term storage systems that manage serial order information (e.g. the context-timing signal of Burgess and Hitch (1999) model of the phonological loop or the phonological store facility within Gupta and MacWhinney's (1997) model of vocabulary acquisition and verbal STM) is a combination of mechanisms consistent with the accumulated evidence. At list presentation, the representations of items will be activated at sublexical, lexical and semantic levels, and this is maintained for a limited time period and until they experience decay (as per interactive activation models). Order information is encoded in a different system that interacts with the language representations and maintains the sequence of activated events in each of the levels in the language network. The degree to which capacity-limited attentional resources are allocated to item and order memory processes is governed by an attentional modulator that determines this breakdown in response to task demands.

7.8 Summary

Across a number of STM tasks thought to emphasis order memory, effects of LTM memory variables have been observed. The existence of these effects discounts the likelihood that initial processing of stimuli is restricted to a phonological code (Saint-Aubin et al., 2005). Redintegration approaches have resolved these irregularities by assuming that additional stimulus features form part of the short-term trace. These differences can be exploited in late-stage redintegration when short-term traces are matched against LTM representations.

Alternatively, language-based approaches assume that the activation of LTM representations is inherent to the creation and maintenance of items stored in STM, and predict pre-retrieval effects (R.C. Martin et al., 2008; Romani et al., 2008; Thorn et al., 2009) although Thorn et al. (2009) have obtained data that cannot be explained wholly within a language-based framework. Despite this, it would appear that, generally speaking, language representations play an important role in the management of item memory and are involved in mnemonic processes at an early stage. Furthermore, the mechanisms that govern the control of item and order information are argued to be separate (Majerus, 2009). A hybrid model that combines the language processor from language-based models with a generalised, possibly non-linguistic serial order mechanism has been proposed to account for the separation of item and order effects in STM tasks, and includes an attentional modulator that directs the proportion of resources dedicated to each mechanism as a function of task demands. This modulation would account for the variability of effect seen in some STM tasks when the relative levels of item and order information differ.

Chapter 8

Computational Models and the Frequency Effect in Serial Recall

8.1 Introduction

The majority of 'models' used to explain the frequency effect in serial recall are conceptual accounts (e.g. redintegration and pre-existing associations between items or language-based models of STM). This chapter provides an overview of computational models that have been considered in relation to the frequency effect.

8.2 Models of serial order

The primary goal of most computational models of STM is to explain phenomena associated with memory for serial order (e.g. G.D.A. Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Farrell & Lewandowsky, 2002; Henson, 1998; Lewandowsky & Murdock, 1989; Neath, 2000; Page & Norris, 1998). These models incorporate explicit mechanisms that determine how order memory is encoded and maintained, including activation gradients (Farrell & Lewandowsky, 2002; Henson et al., 1996; Lewandowsky, 1999; Lewandowsky & Farrell, 2008; Page & Norris, 1998), associations of items with complex temporal signals (G.D.A. Brown et al., 2000) or context signals (Burgess & Hitch, 1999), associations with positional markers (Henson, 1998), item-to-item associations (Lewandowsky & Murdock, 1989), or positional perturbations of encoded list items (Neath, 2000).

The frequency effect in pure lists has sometimes been included as a test of the explanatory capacity of such models (Lewandowsky & Farrell, 2000; Page & Norris, 1998), although this is more often a qualitative rather than quantitative description (see Table 1, Lewandowsky & Farrell, 2008). The problem of computationally explaining the frequency effect however has been treated as a secondary concern to other empirical benchmarks more central to the theoretical issues regarding the mechanisms of short-term forgetting (e.g. primacy and recency effects, modality effects, grouping effects, and phonological confusability and word length effects). In several models where the frequency effect has been considered, the locus of the frequency effect necessarily occurs after response selection, because the serial order mechanisms cannot incorporate earlier influence from frequency without making order memory sensitive to it (e.g. Page & Norris, 1998), and this is in contradiction to the general finding of, at best, small order effects in serial recall when frequency is manipulated (Hulme et al., 2003; Poirier & Saint-Aubin, 1996, although see Nairne & Kelley, 2004).

8.3 Models discussed in previous frequency-based serial recall research

8.3.1 Redintegration-based models

Hulme et al. (2003) provided a review of the existing computational models and their compatibilities with the characteristics of the frequency effect in serial recall. They concluded that the majority of existing STM models treated word frequency as an item-based influence.

8.3.1.1 Item-specific models

Item-specific accounts suggest that the frequency effect is driven by properties of individual items that influence performance at the point of recall. Some models assume this late-stage contribution of word frequency is consistent with the standard redintegration approach, that is, greater likelihood of pattern completion for HF than LF words (Hulme et al., 1997; Roodenrys & Miller, 2008; Schweickert, Chen, & Poirier, 1999), while other explanations involve the use of alternative mechanisms to facilitate the observed HF advantage in pure lists. Among these are differences in baseline activations of the representations for HF and LF words (Henson, 1998), variations in the levels of phonemic feedback dependent on item frequency (Burgess & Hitch, 1999), variable output thresholds for HF and LF words (Page & Norris, 1998) and differences in attractor networks that complete item retrieval as a function of lifetime exposure (Farrell & Lewandowsky, 2002; Lewandowsky & Farrell, 2000, 2008). Notably, in a number of these models it is clear that the frequency of each item will determine its redintegrative potential, that is, the frequency of an item will not contribute to (or detract from) the recall of other items in the list. Consequently, these approaches would predict the recall of each item in a mixed list to match the corresponding level of recall achieved in a pure list of the same frequency type (Hulme et al., 2003). At least one model (Burgess & Hitch, 1999) does predict that performance for HF and LF items in mixed lists is different to that achieved in pure lists, however it still predicts better recall for HF than LF words (Morin et al., 2006). Furthermore, recent derivatives of models have not been developed to account for the mixed list paradox with word frequency and it is not apparent whether they would be capable of accommodating such patterns (e.g. C-SOB, Lewandowsky & Farrell, 2008).

8.3.1.2 List-based redintegration – The Feature model (Nairne, 1990)

The Feature model (Nairne, 1990; Neath & Nairne, 1995) is an example of a redintegration-like framework that does not conform to the item-based definition (Hulme et al., 2003). Hulme et al. (2003) proposed that this model, where recall is dependent on the relative distinctiveness of items against other list members, might be more capable of describing list-based effects than item-specific accounts. A similar sentiment was expressed by Jefferies et al. (2006a). In contrast, Morin et al. (2006) implied that the Feature model was not capable of accommodating mixed list results, but provided no detail of the specific aspects of the model that would make it inappropriate. Therefore, as a model offering the possibility of greater flexibility than item-based models to explain frequency effects in serial recall, an examination of it is included here.

The Feature model assumes that the contents of memory traces can be represented as vectors in both primary and secondary memory (Nairne, 1990; Neath & Nairne, 1995). These reflect the attributes or features of the represented items; vector features acknowledge qualitative and quantitative differences between verbal items, and can vary according to type, value and number. The model encodes two qualitatively different forms of features. Physically based features of items are labeled *modality-dependent*, as these attributes are fundamentally a function of the presentation environment (Neath & Nairne, 1995). Additionally, a second set of features are encoded that reflect the long-term identification and categorical knowledge associated with the item. These features, referred to as *modality-independent*, are invariant with the physical aspects of the items in question and include attributes such as lexicality, frequency, and semantic properties (Nairne, 2001).

Forgetting in the Feature model arises from retroactive interference, in the form of trace-overwriting (Nairne, 1990; Neath & Nairne, 1995). When primary traces corresponding to adjacent events share the same feature there is a possibility this feature will be erased in the former trace. In the model's most simple form, consecutive list items will constitute adjacent traces, except that the final list item is also followed by an internally generated trace that is a copy of the modality-independent features of another list item (Nairne, 1990). This inclusion acknowledges the effect of rehearsal on the final list item.

The model assumes that degraded traces are selected in the order they were encoded (Nairne, 1990; Neath & Nairne, 1995). This order is stored in a vector that logs the sequence of events and is referred to during item retrieval. Subsequent to the selection of a trace, identification of the candidate item occurs by a process assessing the relative degree of similarity between the degraded primary memory trace and those eligible secondary memory traces forming the secondary memory search set. Similarity is determined by feature matching between the degraded trace and the secondary memory traces and is calculated from, and related to, the distances in psychological space between the degraded trace and the potential candidates. Selection of a secondary memory trace for retrieval uses a similarity-based choice rule (Nofosky, 1985, 1986) that determines the probability of selection by comparing the relative similarity of each candidate against the cumulative similarities of all possible choices. The search set typically contains the representations of the most recently presented list items (Neath & Nairne, 1995) however under some conditions this may be extended to include other items. For example, those contexts that present items in a common category (for example digits) are likely to include all categorical members in the search set. Furthermore, experimental conditions that involve the use of open stimulus sets are likely to encourage a wider search set.

After a secondary memory trace has been chosen, a final recovery stage determines whether it will be outputted. This property of the Feature model is included to account for the presumed impact of output interference, argued to be the agent responsible for the strong primacy effects observed in serial recall (Neath & Nairne, 1995). Specifically, as all members of the secondary memory search set remain available for comparison with degraded primary memory vectors throughout the item retrieval process, the probability of recovery of a sampled item is dependent on whether it has already been recalled; previous recall of an item makes its repeated recall less likely. Consequently, recall performance decreases across serial positions, not because the model incorrectly samples an item, but because it is less likely to be recovered if it has been previously selected and recalled (Neath & Nairne, 1995).

The final list item is assumed to benefit from the preservation of modality-dependent features as there is no externally generated trace that follows it. However, this advantage is moderated by the decreased likelihood that it will be successfully recovered, because it was incorrectly outputted in a former list position. The recency effect will manifest so long as the modality-dependent features are sufficiently distinctive to maintain good correspondence with the correct secondary memory representation (Neath & Nairne, 1995).

A further issue with the distinctiveness of traces relates to predictions regarding the relative levels of order memory. This was not a major focus for Nairne (1990) in his initial work, however it is typically found that frequency, and LTM effects in general, have little or no impact on order memory in serial recall tasks (Poirier & Saint-Aubin, 1996; Hulme et al., 2003). If greater item distinctiveness corresponds to the more accurate selection of items in secondary memory, it might be expected that HF words in pure lists facilitate greater order memory than LF words in pure lists. Nairne (1990) recognised this weakness and proposed the inclusion of a time-based perturbation process operating at feature and trace levels (Lee & Estes, 1977; 1981) as a means of establishing levels of order memory reflecting greater influence from an item-independent mechanism. Additionally, expanding the secondary memory search set to include other items would offer the potential for some order errors to be recategorised as intrusions (see Neath, 2000).

Similarity within the stimuli, either in terms of modality-dependent or modality-independent features, impairs recall. Attributes of items that affect recall performance, for example concreteness or frequency, are modeled in the feature model by varying the range of values that features can take on, rather than by changes in the number of features in an item representation. Therefore distinctiveness of feature values is what preserves item identification according to manipulation of long-term linguistic attributes. Neath (1994, as cited in Neath & Nairne, 1995) successfully modelled the concreteness effect by assuming that concreteness-based features could take on non-zero integer values within the ranges [-1, 1] and [-2, 2] for abstract and concrete words respectively.

Therefore in pure lists, based on the assessments of Neath and Nairne (1995), variations in LTM variables are better modeled in terms of the range of values that features can take on. The question is whether variations in feature value ranges, in combination with directionally sensitive retroactive interference and a similarity-based item retrieval process can account for the frequency effect in mixed lists.

8.3.1.2.1 A qualitative assessment of the Feature model

In order to examine more closely the type of qualitative recall pattern the model would produce for pure and alternating lists, a version of the basic model (Nairne, 1990) was created in Excel Visual Basic (Microsoft, 2003). The code is presented in Appendix A.

The model was constructed based on the assumptions outlined by Nairne (1990), and then developed to account for frequency effects according to the prescription of Neath and Nairne (1995). Initially, memory traces were assumed to contain 20 modality-dependent and 20 modality-independent features that were set to –1 or 1 before trace overwriting. An equal number of features of the same value were randomly positioned within the modality-dependent and modality independent portions of each item trace.

Overwriting occurred between adjacent features when they were the same value with probability F being set to 1. The last item experienced trace overwriting of its modality-independent features by the creation of an internal trace through the random selection of a previous list item.

The similarity of each degraded primary memory vector to all intact secondary memory vectors was calculated according to the relation

$$s(i, j) = e^{-d_{ij}}$$

where d_{ij}, the distance between vectors in psychological space, was formed from the ratio of mismatched to compared features,

$$d_{ij} = \frac{a\sum b_k M_k}{N}.$$

The M_k reflect whether feature mismatches have occurred (M_k = 1 for mismatches with k^{th} features, and M_k = 0 when features correspond). The scaling parameter a was set to 7.0. The values b_k, defined as attention weighting parameters were set to 1.

Using the similarity measures $s(i, j)$, the sampling probabilities for each item in each serial position could be calculated using a similarity-based choice rule (Nofosky, 1986),

$$P_s(SM(j) \mid PM(i)) = \frac{w_j s(i, j)}{\sum_k w_k s(i,k)} .$$

The weights w_j and w_k representing response bias were set to 1. For each serial position the sampling probabilities were sequenced to form a cumulative probability distribution. A random number between 0 and 100 was generated and the item corresponding to its location on the cumulative probability distribution was selected for potential recovery. Successful recovery was determined according to the equation

$$P_r = e^{-cr},$$

where the probability of successful recovery P_r is a function that decreases with previous recall of an item ($r > 0$). The scale constant c was set to 2.0.

Initial modelling replicated the conditions outlined in Nairne (1990), namely eight-item lists simulated for 1000 trials. Variations in code were amended by comparison with existing code (Neath, personal communication) and the basic properties of the model were checked against those previously reported (see Figure 8.1). The frequencies of list items were modelled as a greater range of modality-independent feature values for HF than LF words, where each feature value occurred an equal number of times. The list length was shortened to reflect the typical length used in frequency-based serial recall experiments (6 items) and a value of F was chosen to tune model results to the levels more typically experienced in empirical studies (0.6). Simulations involved running the program for 1000 trials and determining the proportion correct serial recall that resulted. These are presented in Figure 8.2. It is clear from this simulation that an attenuated sawtooth pattern emerges with mixed lists and the level of recall for mixed lists is similar to that for pure HF lists. In addition, as would be predicted from the formulation of the Feature model (Nairne, 1990; Neath & Nairne, 1995), variations in order memory occur between HF and LF items in pure lists and are mostly responsible for the observed difference in correct serial recall, which stands in contrast to the empirical data.

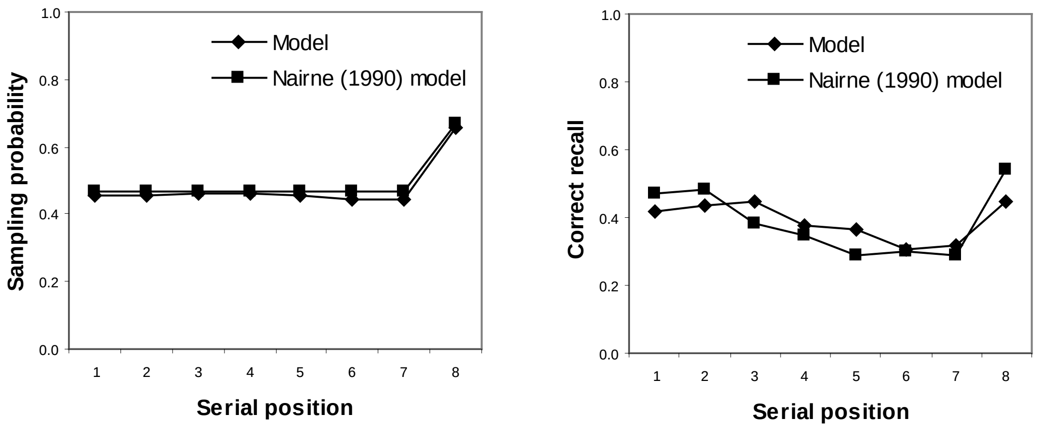


Figure 8.1. The sampling probabilities and correct recall produced from the Feature model (1000 trials).

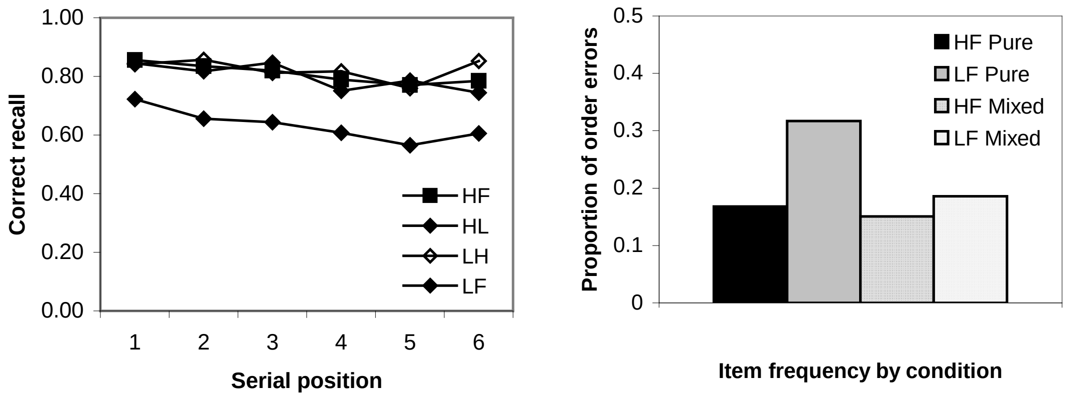


Figure 8.2. Correct recall for alternating frequency lists (6 items, $F = .6$) as predicted by the Feature model. All HF modality-independent features were non-zero integer values in the range [-2, 2].

8.3.2 Association-based models

Poirier and Saint-Aubin (1996) noted that TODAM (Lewandowsky & Murdock, 1989) had been used to model the word frequency effect, specifically the results of O.C. Watkins and Watkins (1997) and M.J. Watkins (1977). In this model the strength of order encoding is assumed to be directly proportional to its frequency, resulting in stronger inter-item associations between HF items. In contrast, LF words are assumed to have higher item information than HF words, as supported by the better recognition memory for LF words. This distinction however creates difficulties for the model to replicate a frequency advantage for item memory with no difference in order memory. More widely, inter-item associations have been found to be unsuitable as a generalised order mechanism in STM (e.g. Henson et al., 1996).

Hulme et al. (2003) examined models from the free recall literature to determine whether any would be suitable candidates on which to further develop an account of frequency in immediate serial recall (e.g. Howard & Kahana, 2002a, 2002b). In such accounts the inter-item associations and pre-existing associations between list items form a critical contribution to the recall of items, and influence the likelihood of ordered recall. Hulme et al. (2003) thought that such a mechanism might also be present in the retrieval of items in serial recall. However, as presented in section 5.4.7.2, the TCM has been found to contain a number of weaknesses from a computational perspective (see Farrell & Lewandowsky, 2008), although at a conceptual level, the possibility of pre-existing inter-item associations affecting serial recall performance at a late-stage remains a valid one (Hulme et al., 2003; Stuart & Hulme, 2000).

8.3.3 SIMPLE

Saint-Aubin and LeBlanc (2005) referred to SIMPLE (Scale Independent Memory, Perception, and Learning, as presented in Hulme et al., 2004) as a model potentially capable of describing the frequency effects observed in their isolate experiments (see Section 6.3.1), motivated by SIMPLE’s success in replicating the qualitative patterns found in the recall of alternating word length lists (Hulme et al., 2004). This model has been more fully presented in subsequent work (G.D.A. Brown, Neath, & Chater, 2007; Neath & Brown, 2006), and an updated version of the model addressing word length effects was presented in Hulme et al. (2006).

SIMPLE is based on a number of assumptions (G.D.A. Brown et al., 2007). Firstly, the location of episodic memory traces occurs along a time-based continuum, taking its point of reference as the present moment, and this information contributes to the positioning of memory events in sequence. Furthermore, memory traces recede and become logarithmically compressed as time passes. As a consequence, recent events are usually more discriminable from each other than are more distant events.

More generally, traces can be located within a multidimensional psychological space that contains the episodic continuum as one dimension. Memory retrieval is thought to be a problem of discrimination; those traces containing information setting them apart from close neighbours in psychological space will be advantaged. Therefore discriminability of traces along other dimensions may also contribute to their memorability.

In its unidimensional form, SIMPLE models the confusability of memory traces on the temporal continuum as a function of the ratio of the distance in time between these events and the point of retrieval. This assumption reflects the observed sensitivity of recency of encoding to the time of test, and the growing insensitivity of memory traces to recency as they recede in time. Multi-dimensional variants of the model allow for items to be encoded on additional dimensions (e.g. phonological similarity). Variation on a second dimension allows the traces of distinctive items to be more readily discriminated from each other, as these are positioned in psychological space more sparsely, even when temporal encoding has been compressed. In contrast, less distinct items will be densely positioned on this second dimension, and as traces are compressed with elapsed time, selection for item retrieval becomes increasingly difficult. In cases where multiple dimensions are used, a weighting parameter sets the degree to which the temporal dimension is used in the selection of traces for retrieval.

A final assumption is that the likelihood of trace retrieval is related to the confusability of a trace. Specifically, the probability of successful retrieval is the inverse of its summed confusability with other traces.

With respect to word length, two versions of SIMPLE have been proposed. The original account outlined in Hulme et al. (2004) assumed that events were encoded along temporal, item and class dimensions. The item dimension was argued to reflect the distinctiveness between short and long words; short words, due to their less complex phonological make-up, were seen as more distinctive than long words. The class dimension was included to reflect the perceived length differences between items when they are presented in mixed rather than pure lists. Weighting parameters adjusted the extent to which differences between item representations on each dimension would contribute to the calculated similarity between items. These calculations underpinned the probability of selection of an item at retrieval, given a cue. The second form of SIMPLE used to explain word length effects was a simplified version of the first (Hulme et al., 2006; Neath & Brown, 2006). The class dimension was excluded in this case.

Saint-Aubin and LeBlanc (2005) argued that the frequency effect might also be explained in terms of item distinctiveness. However, they found that in addition to distinctiveness, constraints placed on STM by the processing requirements of the list conditions would be required for their isolate results to be replicated. Furthermore, if the frequency effect in alternating lists was modeled as a familiarity-based analogue to word length (Neath & Brown, 2006), that is, if HF words were assumed to be more distinctive than LF words, the distinctiveness of adjacent items in alternating lists would be similar to the distinctiveness that occurs in pure HF lists, leading to equivalent recall across these conditions. The intermediacy of the alternating lists cannot be reproduced.

Therefore, SIMPLE does not appear capable, without extension, of providing a distinctiveness-only account of frequency-based results. G.D.A. Brown et al. (2007) acknowledged that SIMPLE was, as such, incomplete as a model for frequency effects in serial recall; they stated that “Bayesian redintegration processes along with richer multidimensional semantic representations would need to be combined with the temporal dimension” (p. 569) to explain lexicality and frequency effects in STM in terms of local distinctiveness.

SIMPLE has been criticised more widely regarding the assumption that recall is driven by retrieval of items along a temporal continuum. Several recent investigations in the serial recall literature have shown that recall is not sensitive to unpredictable temporal isolation effects and temporal variability in events at encoding (Lewandowsky, Brown, Wright, & Nimmo, 2006; Nimmo & Lewandowsky, 2005; 2006) or to differences in recall rate (Lewandowsky, Duncan, & Brown, 2004), although it does appear that temporal information is used in circumstances where output order is unconstrained (Lewandowsky, Nimmo, & Brown, 2008) and positional information is of little use (e.g. running memory span, Geiger & Lewandowsky, 2008). It is also apparent that temporal information is automatically encoded even it is not always relied upon (Lewandowsky et al., 2008). SIMPLE can accommodate these apparent incompatibilities by including an additional positional dimension that encodes serial order when modeling memory for temporally variable material (G.D.A. Brown et al., 2007; Geiger & Lewandowsky, 2008; Lewandowsky et al., 2006; Lewandowsky et al., 2008).

8.4 Conclusion

While a number of conceptual arguments have been developed to account for the frequency effect in mixed lists (Hulme et al., 2003; Saint Aubin & LeBlanc, 2005; Morin et al., 2006; Stuart & Hulme, 2000), no viable computational model supporting these results has been produced. Most existing models of STM have treated the frequency effect as essentially an item-specific one, leaving them incapable of replicating the smooth serial position curves found with alternating lists. Two other candidates identified in the literature as models more likely to accomodate list-based variations in the frequency effect, Nairne's (1990; Neath & Nairne, 1995) feature model and SIMPLE (G.D.A. Brown, et al., 2007; Neath & Brown, 2006) do not describe the empirical data. A computational model of episodic memory used in the free recall literature (TCM, Howard & Kahana, 2002a, 2002b) suggested by Hulme et al. (2003) to be a starting point for future STM model development, specifically with respect to redintegrative processes, has been shown to contain a number of flaws (Farrell & Lewandowsky, 2008), although it may yet provide some conceptual direction. Importantly, the general inability across models to adequately describe the frequency effect in its various guises suggests that mechanisms underpinning the formation of the effect are currently absent from STM models.

Chapter 9

The Interaction of Frequency and Word Concreteness: A Review of the Concreteness Effect in Serial Recall

9.1 Introduction

Given that the frequency effect is argued to be dependent on the strength of association between list items due to co-occurrence in language (Hulme et al., 2003; Stuart & Hulme, 2000, 2009), and this reflects some aspect of semantic organisation, it is also possible that the frequency effect is sensitive to variations in other semantic properties of words. Tse and Altarriba (2007) have shown that the frequency effect is smaller for high than low LSA words. Another variable that could similarly impact the size of the frequency effect is word concreteness. The initial studies presented in this thesis examine whether the level of word concreteness influences the magnitude of the frequency effect (Chapters 10 and 11 respectively). Before presenting these experiments however, it is appropriate to provide some background on the concreteness effect in serial recall, and the theoretical interpretations of this effect. [1]

9.2 The concreteness effect in serial recall

9.2.1 Definitions and related terms

The property of word concreteness indexes the extent to which a word refers to a material object in contrast to an abstract quality or action (Walker & Hulme, 1999). For example, words such as *broom* and *glass* are highly concrete, while the words *harm* and *truth* possess low levels of concreteness and are more commonly referred to as abstract words. Imageability, a measure of how easily a mental image of a word can be produced, is highly correlated with concreteness. Studies reporting on concreteness or imageability tend to use the terms interchangeably.

---

[1] Sections of Chapters 9, 10 and 11 have been published in "The interaction of word frequency and concreteness in immediate serial recall", by L.M. Miller and S. Roodenrys, 2009, *Memory & Cognition*, *37*,pp. 850-865. Copyright 2009 Psychonomic Society Inc. Reprinted with permission.

9.2.2 The word concreteness effect

The word concreteness effect is the better serial recall of concrete than abstract words (Allen & Hulme, 2006; Romani et al., 2008; Walker & Hulme, 1999, although see Tse & Altarriba, 2009, who claim the effect is dependent on the emotional valence of words). Those studies that have manipulated stimuli on the basis of imageability have reported equivalent results; high imageability words produce better recall than low imageability words (Bourassa & Besner, 1994; Caza & Belleville, 1999; Majerus & Van der Linden, 2003; Tse & Altarriba, 2007). The difference in performance with word sets varying in concreteness is presumed to reflect differential levels of semantic processing in LTM; concrete words are argued to have richer semantic representations than abstract words (Jones, 1985; R.C. Martin & Lesch, 1996; Neath, 1997). A second related explanation is that concrete words are advantaged by the representation of visual features in addition to verbal ones (Paivio, 1986, 1991).

9.3 Imageability effects in word span and the serial recall of supraspan lists

Bourassa and Besner (1994) determined that an apparent superiority in memory span for content words (nouns, verbs, adjectives, and some adverbs) over function words (prepositions, conjunctions, pronouns, articles, auxiliary verbs, and some adverbs) was the result of a confound with imageability. When function and content words were matched for imageability, there was no difference in performance, suggesting that imageability was the influential property affecting recall rather than the semantic/syntactic differences between content and function words. However, Caza and Belleville (1999) later showed that an effect of word class independent of imageability was possible.

Majerus and Van der Linden (2003) explored the serial recall of high and low imageability words across four age groups from young children to young adults. They found a small imageability effect in serial recall in all age groups except adolescents, however when the recall of all items regardless of order was analysed all age groups showed an effect. The authors therefore discounted the possibility that developmental influences altered the presence of the imageability effect, and argued that this effect, in combination with other lexico-semantic effects, was a consistent feature in the recall of children and adults, and provided evidence of LTM influence in the retention of verbal material over the short-term.

Tse and Altarriba (2007) examined the relationship between imageability and a semantic association-based contribution to serial recall, as measured by LSA-cosine. They noted that that while manipulations of word frequency were reported to be consistent with LSA-cosine measures (Hulme et al., 2003, see also section 5.4.8), that is HF word sets had greater LSA-cosine values than LF words sets, similar analysis of the manipulations of concreteness and imageability sets used in the literature revealed that abstract or low imageability sets possessed greater LSA-cosines than concrete or high imageability word sets. They argued that if imageability, unlike frequency, was independent of inter-item association the effect of LSA-cosine would be the same regardless of the level of imageability.

The authors performed an orthogonal manipulation of semantic association (using LSA-cosine) and imageability. Effects of both variables were found, indicating that they each made a contribution to recall performance. However there was no interaction – the LSA cosine contribution was the same (recall was greater by approximately 10% for high LSA cosine conditions) regardless of the level of imageability. Tse and Altarriba (2007) argued that the absence of an interaction demonstrated that inter-item associativity was not a mechanism involved in the formation of the imageability effect in serial recall. They proposed that the imageability effect could be explained in terms of Paivio's (1986) dual coding hypothesis, where highly imageable items benefit from both verbal and visual encoding, or as a consequence of item-specific redintegration (Hulme et al., 1997) where the richer semantic representations of high imageability words promotes the greater likelihood of successful item reconstruction.

9.4 Concreteness effects in the serial recall of supraspan lists

9.4.1 Walker and Hulme (1999)

Walker and Hulme (1999) investigated the serial recall of concrete and abstract words, where recall was either spoken or written. They determined that output modality was not an influential factor in the creation of the concreteness effect, except in altering shape of the serial position curves. Word length was also manipulated in these experiments, and was found not to interact with concreteness.

Furthermore, Walker and Hulme (1999) compared the differences between frequency and concreteness effects with particular reference to how they manifest across serial position. In their experiments, the concreteness effect was found to display a consistent difference in recall for medial positions of the list sequence. The authors compared this pattern with those previously observed for word frequency (e.g. Hulme et al., 1997), where there was a typical increase in the effect across serial positions, except for the last. The absence of a change in effect of word concreteness across serial positions was seen by Walker and Hulme (1999) as evidence that concreteness acts on the degree and type of semantic encoding, and is more resistant to interference from output processes than the phonological traces presumed to underpin the frequency effect.

This difference between serial position interactions for frequency and concreteness motivated Walker and Hulme (1999) to extend Hulme et al.'s (1997) description of redintegration to allow for multiple STM codes and LTM representations operating in parallel, with the frequency and concreteness effects the outcome of separate redintegrative processes. The authors however, did not elaborate on the specific role short-term semantic encoding would play in the restoration of items, or how this contribution would link with redintegrated output from the phonological system (Romani et al., 2008).

Walker and Hulme (1999) also found that the concreteness effect was maintained in backwards recall while the word frequency effect had been eliminated (Hulme et al., 1997). Backwards recall is viewed to be a complex procedure that may involve other retrieval strategies (e.g. visuo-spatial, Li & Lewandowksy, 1995), however Hulme et al. (1997) interpreted the failure to observe a frequency effect under these conditions as reflecting retrieval that was based on semantic information. The preservation of the concreteness effect in backwards recall was seen by Walker and Hulme (1999) as support for this idea.

As reviewed in Chapter 7, Walker and Hulme (1999) used a matching span task on lists made of either concrete or abstract words. This recognition-based procedure delivered a null result. As a consequence, the authors argued that the concreteness effect was a product of the output processes associated with STM recall.

9.4.2 The concreteness effect in serial recall and speech perception and speech production tasks: Allen and Hulme (2006)

Allen and Hulme (2006) examined the recall of concrete and abstract words along with a number of speech perception (auditory lexical decision and word identification in noise) and speech production (definition naming, delayed repetition, and maximal rate of articulation) tasks. They found a concreteness effect that was significant for all serial positions except the last, and a small order effect reflecting more order errors for abstract than concrete words. Item errors were also more likely to occur for abstract than concrete words, in particular omissions and phonological approximations to actual items. Performance in speed and accuracy of lexical decision, word identification in noise accuracy, and delayed repetition speed were not affected by concreteness. However, in definition naming concrete words were better named and named more quickly than abstract words. Concrete words were also articulated faster than abstract words, although the difference for articulation rate was small.

Allen and Hulme (2006) conducted within-subjects regressions (Lorch & Myers, 1990) where for individual participants, the speech perception and speech production measures for each item were regressed onto serial recall performance of that item. The coefficients of regression were then analysed to assess their significance. This process found that definition naming accuracy was the only speech related task that had a significant relationship to serial recall performance, either in bivariate or multivariate regression analyses. Furthermore, the authors demonstrated that the effects of concreteness were explained by differences in definition naming accuracy, suggesting common processes for these tasks, and that the recall of words might depend on how successfully semantic representations can facilitate the appropriate speech output representations during recall.

9.4.3 Concreteness effects in serial recall and other STM tasks: Romani et al. (2008)

Romani et al. (2008) examined the influence of concreteness across a range of STM tasks that varied the emphasis on item and order information. Their results motivated a language-based model that was developed from the model of R.C. Martin et al. (1999) (see Figure 9.1). They proposed the removal of the lexical-semantic buffer and the inclusion of a placeholder operating at the phonological level (N. Martin & Saffran, 1997) to reflect the observed relationship between the maintenance of serial order and the integrity of phonological representations. Order would be maintained through the buffered phonological representations that exist for speech perception or speech production. The role of the input buffer is in the conversion of acoustic information into phonological representations, while the output buffer is responsible for the transformation of phonological representations into an articulatory code. The "virtual" lexical-semantic buffer in this model would be formed from the activated lexical-semantic representations in the LTM knowledge structure. The encoding order of lexical-semantic representations is directed from the buffered phonological representations in the placeholder.

Please see print copy for images

Figure 9.1. A language-based model of short-term memory. The model includes a placeholder at the phonological level to maintain the serial order of items. From "Concreteness effects in different tasks: Implications for models of short-term memory", by C. Romani, S. McAlpine, and R.C. Martin, 2008, *Quarterly Journal of Experimental Psychology,* 61, p. 315. Copyright 2008 by Taylor & Francis Group, http://www.informaworld.com. Reprinted with permission.

These authors replicated the concreteness effect and produced a concreteness by serial position interaction similar to that found by Walker and Hulme (1999). More specifically, the interaction was driven by the presence of a concreteness effect that is relatively constant across serial positions except for the last two, where it is absent. Romani et al. (2008) noted this pattern as similar to other reports of lexical-semantic effects that were more prominent in the primacy than recency portions of the curve (Kintsch & Buschke, 1969; O.C. Watkins & Watkins, 1977). In particular, O.C. Watkins and Watkins (1977) found that the frequency effect diminished for recency positions (see section 4.2). Romani et al. (2008) disagreed with Hulme et al.'s (1997) interpretation of the frequency by position interaction, citing the assumption of a widening effect due to disruption of phonological traces from output sources as *ad hoc*. They considered another possibility put forward by Hulme et al. (2003), namely that the frequency and concreteness effects are due to different processes; while concreteness depends on item-based redintegration, frequency relies on a different form of co-occurrence mechanism. However, their favored position was to interpret the pattern observed for lexical-semantic effects, that is, a narrowing of effect in the last two positions in the list, as a product of masking by a strong phonological record.

As highlighted in section 7.3.2, Romani et al. (2008) demonstrated that the failure to produce a concreteness effect in matching span is a function of task design. They compared matching span performance between open and closed sets of items under control and articulatory suppression conditions. In control conditions the closed sets replicated the outcome observed by Walker and Hulme (1999), however the open sets produced a facilitative effect of concreteness. The authors declared a reduction in the difference in item information between concrete and abstract words, due to repeated presentation, was responsible for the lack of effect with the closed sets. Furthermore, the addition of articulatory suppression to the task yielded results that were negatively offset to those in the control condition; there was no effect of concreteness with open sets and a reversed concreteness effect with closed sets. These results were argued to indicate that under suppression, additional phonological input serves to confuse the phonological record, disrupting the order of items and undermining any item identity advantage concreteness can provide. In the case of open sets, the concreteness advantage was nullified. When closed sets were used the order information was so weak that item and order information could become decoupled, an outcome thought to harm concrete words more than abstract ones. It was argued that concrete words would be more likely than abstract words to produce item information capable of overriding order information, as the latter experience less strong lexical-semantic activation, that in turn, encourages participants to use available phonological information, however weak. It should be noted that the performance data of the suppression conditions in this study included sizeable proportions of participants with near chance performance, although all results were replicated in a second sample of participants.

Finally, Romani et al. (2008) compared performance in order reconstruction and free recall under control and articulatory suppression conditions to test the size of the concreteness effect in tasks with differing emphases on order and item information, and more generally to test the proposition that order information relies on the integrity of phonological representations while item identity is influenced by the strength of lexical-semantic codes. The effect of concreteness was greater in free recall than order reconstruction, reflecting the expected differences in task demands. Concreteness effects were found in the order reconstruction tasks in both control and suppression conditions, implying that lexical-semantic effects operate from the time of presentation onwards. Free recall under suppression was less influenced by order information than in the control condition as indicated by a measure of seriation, supporting the view that serial order is associated with the integrity of phonological representations.

9.4.4 Word concreteness and word valence in serial recall: Tse and Altarriba (2009)

Tse and Altarriba (2009) reported that the concreteness effect was contingent on the emotional valence of items. Specifically, they found that positively valenced emotion words and neutral words produced a concreteness effect, while negatively valenced emotion words did not. They also reported an order effect in one experiment such that concrete positive words showed better order memory than abstract positive words, while there was no difference in order memory for negatively valenced concrete and abstract words. The failure to find a concreteness effect with negatively valenced words was replicated in a second experiment.

The authors argued that the results were consistent with item-specific redintegration (Hulme et al., 1997) and that this process was dependent on both the concreteness and valence of items. While for positive and neutral words the richer semantic representations associated with concrete words facilitates their superior redintegration, the encoding of negatively valenced words is marred by attentional avoidance (Mackintosh & Mathews, 2003)[2]. This mechanism serves to divert attention away from mildly negatively valanced material when it is visually presented, negating any advantage to concrete words in terms of reconstruction at recall. Alternatively, Tse and Altarriba (2009) considered that the effects of concreteness and word valence might be derived from two separate redintegrative systems, much in the same manner that Walker and Hulme (1999) proposed that concreteness and word frequency were products of distinct systems.

9.5 Conclusion

The general finding across these experiments is that concrete words are better recalled than abstract words. Redintegration accounts (Tse & Altarriba, 2007, 2009; Walker & Hulme, 1999) argue that the effect derives from the richer semantic representations of concrete items that better facilitate the reconstruction of items at recall. Furthermore, Walker and Hulme (1999) based on differences in the serial recall performances between frequency and concreteness, claimed that separate phonological and semantic redintegration systems supported STM recall, and that the temporary traces of items would therefore need to include both phonological and semantic features in order to underpin these processes. Alternatively, language-based models (N. Martin & Saffran, 1997; R.C. Martin et al., 1999; Romani et al., 2008) view concreteness, by virtue of its influence on the semantic representations of words, as a property that necessarily interacts with lexical-phonological representations in a multi-representational and interactive activation system.

Finally, while Tse and Altarriba (2007) found that LSA-cosine measures interacted with frequency, such that an LSA-based contribution was only evident for LF words, an analogue experiment manipulating imageability with LSA-cosine found that the effect of LSA was the same for high and low imageability words and therefore independent of imageability. Accordingly, experiments investigating the

[2] MacKintosh and Mathews (2003) also found evidence of attentional avoidance for mildly positively valanced stimuli, although how this finding integrates with the results of Tse and Altarriba (2009) was not addressed.

possible interaction of frequency and concreteness in serial recall would provide complementary information regarding the relationship between semantically-based variables and the lexical properties of words, and provide a basis for a more thorough interpretation of the findings of Tse and Altarriba (2007).

Chapter 10

Experiment 1: The Interaction of Frequency and Word Concreteness with Visual Presentation and Spoken Recall

10.1 Introduction

The aim of Experiments 1 and 2 was to test the nature of the LTM contributions of word frequency and word concreteness to STM recall performance when these variables are manipulated within the same experiment. Furthermore, these experiments provided the opportunity to examine the compatibility of the approaches used to explain concreteness and frequency effects, namely the dual redintegration approach of Walker and Hulme (1999) and language-based models (for example Romani et al., 2008) with the observed outcomes.

Both these frameworks can accommodate the possibility of a frequency by concreteness interaction although the arguments for its existence are quite different. In language-based models word frequency is assumed to affect the activation levels of the lexical nodes (R.C. Martin et al., 1999); HF items have stronger lexical activations than LF words, and these serve to deliver stronger feedback to the connected representations at the phonological level. Word concreteness is reflected in the richness of semantic representation, and this determines the amount of feedback to lexical-phonological representations, that subsequently provide feedback to buffered output phonological representations. Therefore this class of model explicitly incorporates the interaction of frequency and concreteness via the structure of long-term knowledge and continual access to long-term representations throughout the encoding, retention and retrieval of items (N. Martin, 2009). In contrast, within the dual redintegration model of Walker and Hulme (1999), an interaction might be generated from the increasing numbers of items that can be successfully redintegrated from both semantic- and phonologically-based processes when frequency and concreteness are greater. Recall of an item presumably requires successful redintegration from a single system, and HF words that are also highly concrete would produce redundant information when both systems yield the same output. Should no interaction arise this would be harder to explain within the language-based view, given the specific links between these types of representations. The dual redintegration approach might have greater scope to accommodate such a finding, in terms of a ‘one-shot’ process, where an item could be redintegrated by one of the phonological or semantic mechanisms, but not both.

From a qualitative perspective, these approaches might anticipate different forms of serial position interactions for each of the variables involved. The Walker and Hulme (1999) model relies on a qualitative difference between the interactions with serial position for frequency and word concreteness reflecting the different processes that produce them, while Romani et al. (2008) suggest that interactions of lexical-semantic variables with serial positions should be of the same form, that is, marked by a reduction in effect for the recency positions of the curve. The first experiment was run to investigate the variations in recall performance of lists of words in a factorial manipulation of word frequency and word concreteness. Items were visually presented with the specific aim of examining the behaviour of these lexical-semantic variables in a context where arguably phonological information is weaker, relative to auditory presentation (Tolan & Tehan, 1999), and possibly less likely to mask lexical-semantic effects (Romani et al., 2008).

10.2 Method

10.2.1 Participants

Forty University of Wollongong undergraduate and postgraduate students (6 males and 34 females) either participated for course credit or volunteered participation. The mean age of the sample was 21.7 years (*SD* = 6.5 years) and all participants had English as a first language.

10.2.2 Materials

Four stimulus sets of twelve words each were selected such that a factorial manipulation of word frequency and concreteness was achieved (namely, word sets contained items that were either low frequency and low concreteness – LFLC, low frequency and high word concreteness – LFHC, high frequency and low concreteness – HFLC, or high frequency and high concreteness - HFHC). Word frequency ratings were derived from the Celex database (Baayen et al., 1993) and were the composite of database entries for the same orthography and across those entries with the same word identification number, so that for example, the frequency counts for *bird* and *birds* were combined. This was done to make the frequency counts more representative of frequency-based effects in serial recall when list items are not directly related. Concreteness ratings were sourced from the MRC Psycholinguistic Database (Coltheart, 1981). The sets were matched on phonological neighbourhood using values calculated from the Celex database, the number of phonemes, and number of letters of set items. The phonological similarity of items within each word set were examined using an Excel Visual Basic program to determine dissimilarity ratings based on the methodology outlined by Mueller, Seymour, Kieras and Meyer (2003, PSYMETRICA). This is a pairwise analysis, and involves decomposing words according to syllable structure and phoneme clusters. As the words in the current experiment were monosyllabic, effects of syllable structure are not considered here. Within each item, the vowel nucleus is identified, and then phoneme clusters either side of the vowel become the onset (phonemes preceding the vowel) and coda (phonemes following the vowel). Alignment of clusters (onset, nucleus and coda) occurs between word pairs, and the phonological features of these elements are compared. This process results in three dissimilarity measures for each unique pairwise combinations of items in the set. A dissimilarity profile for the entire set is defined to be the average of these dimensions across all pairwise comparisons. Dissimilarity values lie on a scale of zero to one, where zero indicates identical phonology, and numbers closer to one indicate greater dissimilarity.

A MANOVA was performed on the word sets, Wilks Lambda, $\Lambda = .005$, $p < .001$, the analysis demonstrating that (i) HF sets differed significantly from LF sets – $F(3, 44) = 37.69$, $p < .001$, $MSE = 1877.083$, Tukey's HSD homogeneous subset analysis identified HF and LF conditions as significantly different; and (ii) high concreteness sets differed significantly from low concreteness sets – $F(3, 44) = 565.98$, $p < .001$, $MSE = 689.233$, Tukey's HSD homogeneous subset analysis identified high and low concreteness conditions as significantly different. No sets differed significantly from the others with respect to phonological neighbourhood – $F(3, 44) = 0.76$, $p = .522$, $MSE = 207.087$, number of phonemes - $F(3, 44) = 0.21$, $p = .887$, $MSE = .261$, number of letters – $F(3,44) = 1.66$, $p = .189$, $MSE = .402$, or any of the phonological similarity measures – onset, $F(3,44) = 1.66$, $p = .189$, $MSE = .002$, nucleus $F(3,44) = 1.70$, $p = .181$, $MSE = .002$, and coda $F(3,44) = 0.79$, $p = .504$, $MSE = .002$. The word sets and word attributes are presented in Appendix B.

Pseudo-random six-word lists were generated for each condition and each participant such that no word appeared twice in the same list, and each word appeared once in each serial position across the set of lists. The order of conditions (four sets of twelve lists) was factorially counterbalanced across participants to minimise additional influences from learning and practice effects.

10.2.3 Procedure

All participants were tested individually. The total time to complete the experiment was approximately half an hour. Testing was controlled via the DMDX program (Forster & Forster, 1999) run on an IBM compatible computer and commenced after two practice trials. Initiation of each trial occurred when the participant pressed the spacebar of the keyboard. The program would then present each word in the trial at a rate of one word per second. Words were presented in the centre of a black screen, and were in white, 40 point Times New Roman font. After the sixth word, a recall prompt appeared ("?????") indicating that participants should commence recall. Spoken recall was according to strict serial recall criteria, that is, (i) words were recalled in order of presentation; (ii) if a word could not be recalled the participant would indicate by saying 'pass'; and (iii) previous items were not to be recalled after moving on to successive items in the list.

10.3 Results

For each participant and each condition, recall was scored and collapsed across trials to provide the number of correct items by serial position. Correct recall was scored using a strict criterion, namely responses were considered correct if a word was recalled in the position corresponding to its serial order in presentation. The mean proportion of correctly recalled items by serial position and condition is shown in Figure 10.1 (means and standard deviations of correct recall for each serial position in each experiment are reported in Appendix C). Performance collapsed across serial positions was the greatest for the HFHC condition ($M = .670$, $SD = .149$), then the

HFLC condition (M = .644, SD = .148), the LFHC condition (M = .602, SD = .151), and finally the LFLC condition (M = .514, SD = .130).

10.3.1 Serial recall

An alpha level of .05 was applied to the following statistical tests. A 2 x 2 x 6 (frequency x concreteness x serial position) repeated measures ANOVA was conducted on the serial recall data. All main effects were significant: frequency, $F(1,39) = 62.80$, $p < .001$, $MSE = .037$; concreteness, $F(1,39) = 27.80$, $p < .001$, $MSE = .028$; and serial position, $F(5, 195) = 160.67$, $p < .001$, $MSE = .110$. Thus HF words were better recalled than LF words, concrete words were better recalled than abstract words, and the performance across serial position revealed a typical pattern for visually presented material; a decline in performance from the first item through to the fifth item, followed a modest increase at the last position. The frequency by concreteness interaction was found to be significant, $F(1,39) = 10.77$, $p = .002$, $MSE = .022$.

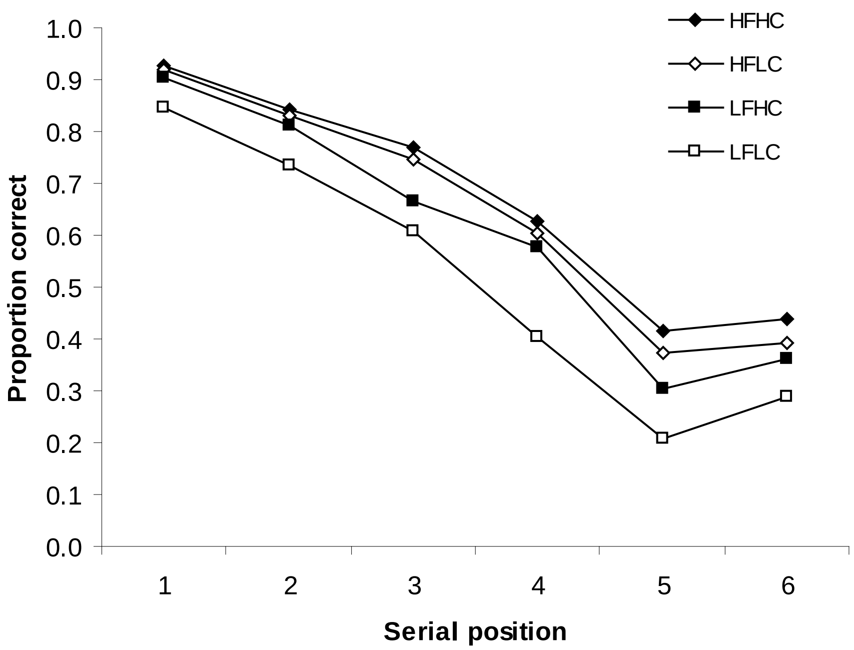


Figure 10.1. Serial recall of words as a function of frequency and concreteness with visual presentation and spoken recall. HFHC - High frequency, high concreteness, HFLC – High frequency, low concreteness, LFHC - Low frequency, high concreteness, LFLC – Low frequency, low concreteness.

The frequency effect was smaller for concrete words than for abstract words, or alternatively, the concreteness effect was smaller for HF than LF words. The frequency by serial position interaction was also significant, $F(5, 195) = 3.81$, $p = .006$, $MSE = .018$, as was the concreteness by serial position interaction, $F(5,195) = 2.66$, $p = .035$, $MSE = .011$. These results appear to be driven by a widening of the respective effects for the latter serial positions together with the possibility of a ceiling effect operating on the first position. The three-way interaction was non-significant – $F(5, 195) = 1.76$, $p = .122$, $MSE = .013$.

10.3.2 Serial position interactions

The influence of the ceiling effect on the interactions with serial position was investigated by reanalysing the data for the last five serial positions. While the frequency by position interaction remained significant, $F(4, 156) = 2.50$, $p = .045$, $MSE = .015$, the concreteness by position interaction was marginal, $F(4, 156) = 2.12$, $p = .081$, $MSE = .010$, suggesting performance in the first position artificially contributed to the results in the full analysis. With this in mind, the differences between effects for the primacy and recency portions of the curves were compared. The average difference in the frequency effect for the primacy positions ($M = .079$, $SD = .076$) was, after Bonferroni correction for multiple comparisons ($\alpha = .025$), not significantly different than that for the recency positions ($M = .119$, $SD = .120$), $t(39) = -2.03$, $p = .049$. In the case of word concreteness, the average difference in the primacy positions ($M = .039$, $SD = .077$) was smaller than the average difference in effect for recency positions, ($M = .075$, $SD = .090$), $t(39) = -2.35$, $p = .024$. It appears unlikely in this instance, despite limitations to the magnitude of differences in the primacy portion of the curve that effects for either variable would be larger in this region than in the recency portion of the curve.

10.3.3 Item analysis

Error analysis was performed on the data to further examine the impact that word frequency and concreteness may have on the recall of items. The responses for each trial were classified as either correct (the correct item in the correct serial position), an order error (the item recalled corresponded to a word that was presented elsewhere in the trial) or an item error (the response did not match any item presented in the trial or was a repetition of a previously recalled list item). Non-repetition item errors were further broken down according to the classification suggested by Allen and Hulme (2006), each error being either an omission (when participants said ‘pass’ to an item, or indicated that they could not recall an item in that position), an intra-set intrusion (when an item from within the current experimental set but outside the list presented was recalled - ISI), an intra-experiment intrusion (when the response was from one of the other experimental sets - IEI), an extra-experiment intrusion (when the response did not correspond to any item within the any of the experimental sets - EEI) or a phonological approximation (when a response approximated a list item by at least 50% of the presented item’s phonemes - PA). The proportion of errors of each type, collapsed across serial position and participants is given in Table 10.1, together with the proportion of items correct.

Table 10.1

Mean proportion of items correctly recalled and mean proportions of different error categories by condition in Experiment 1

	Correct	Order errors	Item Errors
HFHC	.670 (.149)	.090 (.062)	.240 (.123)
HFLC	.645 (.148)	.080 (.054)	.275 (.128)
LFHC	.603 (.151)	.086 (.059)	.311 (.129)
LFLC	.515 (.130)	.080 (.054)	.405 (.114)

	Item errors by category					
	Repetitions	Omissions	EEI	ISI	IEI	PA
HFHC	.003 (.006)	.213 (.125)	.000 (.002)	.023 (.026)	.000 (.002)	.001 (.004)
HFLC	.004 (.006)	.244 (.125)	.002 (.005)	.017 (.020)	.004 (.010)	.005 (.009)
LFHC	.005 (.009)	.279 (.126)	.002 (.007)	.019 (.021)	.002 (.006)	.005 (.009)
LFLC	.002 (.005)	.342 (.121)	.003 (.007)	.019 (.021)	.026 (.028)	.012 (.020)

Note. HFHC - High frequency, high concreteness, HFLC – High frequency, low concreteness, LFHC - Low frequency, high concreteness, LFLC – Low frequency, low concreteness, EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Rep – Repetitions, PA – Phonological approximations. Standard deviations are given in brackets.

The proportions of order errors were conditionalised by dividing the total number of order errors by the number of items correctly recalled regardless of order. This procedure avoids confounding different levels of order memory with differing levels of item memory (Murdock, 1976; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999). The resultant conditionalised rates were .125 for HFHC, .116 for HFLC, .134 for LFHC and .142 for LFLC respectively. A 2 x 2 repeated measures ANOVA identified that neither frequency, $F(1,39) = 2.14$, $p = .152$, $MSE = .006$, nor concreteness, $F(1,39) = 0.00$, $p = .982$, $MSE = .005$, nor the frequency by concreteness interaction $F(1,39) = 0.96$, $p = .333$, $MSE = .003$, affected memory for order.

A 2 x 2 repeated measures ANOVA was performed on the total item errors by condition. This analysis revealed significant main effects of frequency, $F(1,39) = 118.02$, $p < .001$, $MSE = .122$, and concreteness, $F(1,39) = 58.02$, $p < .001$, $MSE = .104$, and a significant frequency by concreteness interaction, $F(1,39) = 12.27$, $p = .001$, $MSE = .103$. Higher frequency and more concrete words yielded better memory for items, and the change in effects of frequency and concreteness was such that the effect of one variable was smaller at higher levels of the second variable.

Of all item error categories omissions formed the largest contribution and these were analysed separately to ascertain the sensitivity of items as a function of condition to the likelihood that recall would fail altogether. This analysis yielded significant main effects, namely frequency, $F(1,39) = 79.84$, $p < .001$, $MSE = .123$, and concreteness $F(1,39) = 38.50$, $p < .001$, $MSE = .084$, while the interaction reached borderline significance $F(1,39) = 4.10$, $p = .050$, $MSE = .092$. The remaining error categories were not analysed by category as the data was considered too sparse, and totals too small to be meaningful. Repetitions, EEI and ISI do not vary much across conditions while the LFLC rates for both IEI and PA categories is many times those of any other condition. This is likely to be due to the substitution of the LFLC item *truce* with *truth* (an item from the HFLC set) on some occasions.

10.4 Discussion

Several results of Experiment 1 are consistent with previous research. The main effects were in line with prior findings; HF words were better recalled than LF words (Allen & Hulme, 2006; Hulme et al., 1997; Hulme et al., 2003; Majerus & Van der

Linden, 2003; Poirier & Saint-Aubin, 1996; Roodenrys & Quinlan, 2000; Stuart & Hulme, 2000; Tehan & Humphries, 1988; Watkins & Watkins, 1977), concrete words were better recalled than abstract ones (Allen & Hulme, 2006; Romani et al., 2008; Walker & Hulme, 1999), and the effect of serial position was broadly consistent with experiments using visual presentation of words (Poirier & Saint-Aubin, 1996; Roodenrys & Quinlan, 2000; O.C. Watkins & Watkins, 1977).

Importantly however, this experiment also identified a new characteristic of LTM influence on STM recall, the frequency by concreteness interaction. The size of effect of word frequency will be dependent upon the concreteness of the items used, and vice versa. Support for the direction of this interaction is given by R.C. Martin et al. (1999), who reported the results for a group of controls on a list repetition task involving sets of words that varied in terms of imageability and frequency, when investigating the STM performance of an anomic patient MS. They did not report results for the controls by each condition, collapsing performance across imageability and frequency, but did report a marginally significant interaction, consistent with the direction observed in the current experiment.

The presence of an interaction between lexical-semantic variables is seen to be a natural outcome of the architecture of language-based models (N. Martin, 2009), although it suggests that feedback activation to phonological representations is functionally limited, in that activation from items that are both highly frequent and highly concrete is not substantially greater than the lexical-semantic activation from items either highly frequent or highly concrete. This result can be likewise accommodated within the dual redintegration framework of Walker and Hulme (1999) if both redintegrative mechanisms operate on item retrieval. In this case, the interaction occurs because of the greater proportion of items successfully redintegrated by both systems when frequency and concreteness is high.

The finding of an interaction in this experiment rests on the finding that the frequency effect is smaller for high than low concreteness words, or equivalently, the concreteness effect being smaller for HF than LF words. Tse and Altarriba (2009) found that the concreteness effect did not occur for negatively valenced words, but was present (and indifferent) for neutral or positively valenced words. Accordingly, if there were differences in the numbers of negatively valenced items across word sets, the interaction observed in Experiment 1 could have resulted from the failure to control word valency. Post hoc examination of the stimuli determined that many stimuli did not have available ratings (Altarriba, Bauer, & Benvenuto, 1999; Bradley & Lang, 1999)[1]. However, from these ratings the HF word sets (both high and low concreteness) were found to each contain some positively valenced or neutral items. The indication, though sketchy, is that the size of the concreteness effect for the HF sets was not influenced by word valency. Moreover, given the sizeable concreteness effect that occurred for the LF sets, it can be inferred based on the results of Tse and Altarriba (2009) that the emotional valence of these items in these sets was, on average, at least neutral. Accordingly, it is unlikely that the pattern of results observed in Experiment 1 was unduly influenced by differences in the emotional valence of the word sets.

Analysis of the current experiment identified that neither variable affected memory for order. For word frequency this is the general finding when recall is immediate (Allen & Hulme, 2006; Hulme et al., 2003; Morin et al., 2006; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000), although two experiments have been reported where a small difference in memory for order was detected (Hulme et al., 2003, Experiment 2; Tse & Altarriba, 2007). Word concreteness has also been reported not to be associated with memory for order (Walker & Hulme, 1999), however Allen and Hulme (2006, Experiment 1) identified a small advantage for concrete words. It would appear that order effects, when present, are slight and may arise from demand characteristics of individual experiments or variations in participant pools. Despite such anomalies, it is clear that the results for order errors, in combination with those for item recall as reflected in total item errors, identify the impact of lexical-semantic variables as one affecting item memory (Hulme et al., 1997; Hulme et al., 2003; Poirier & Saint-Aubin, 1996; Romani et al., 2008; Walker & Hulme, 1999).

The findings that both variables and their interaction influence the total number of item errors is repeated in the pattern for the dominant item error category of omissions. Both main effects have been reported previously (Allen & Hulme, 2006; Hulme et al., 1997; Hulme et al., 2003; Stuart & Hulme, 2000). In terms of redintegration, these results reflect the relative advantage higher levels of LTM variables can provide in the retrieval of items at recall; a greater proportion of items degraded beyond a recoverable state for phonological or semantic redintegration when

[1] Tse and Altarriba (2009) reported that many of their neutral stimuli also did not have ratings.

word frequency is low and items are abstract, intermediate levels of omission when items are either high frequency, highly concrete, but not both, and the least proportion of omissions when word frequency is high and items are concrete. This last circumstance corresponds to the event where a degraded STM trace is most likely to be recovered from either system. Language-based models would posit that higher levels of concreteness and word frequency assist the maintenance of phonological representations at encoding and during retention. Such items are less likely to be unidentifiable at the point of recall because their representations provide greater lexical-semantic feedback activation to phonological traces across the life of the trial.

Both variables were found to interact with serial position, however there are a number of qualifications to be considered before any conclusions regarding these effects can be drawn. Firstly, the reanalysis of the data for the last five serial positions found that the frequency by serial position interaction retained significance, while the concreteness by serial position interaction failed, implying that the latter had been more influenced by a ceiling effect in the full analysis. The combination of presentation and recall modalities used in the current experiment is comparatively rare, however Schweickert et al. (1999) reported a reanalysis of data from a similar investigation by Roodenrys and Hulme testing the frequency effect in five-item lists where they found a greater effect for the middle serial positions. The frequency by serial position interaction generated in the current instance could be viewed in a similar way. However, other complications in this case include the anomalies present in the third and fourth serial positions for the LFHC condition and their influences on the frequency and concreteness effects across positions, and the possible extension of a ceiling effect operating on the second serial position. Furthermore, the smaller concreteness effect might be constrained by power issues with respect to the detection of patterns across the recall curve. It would appear that in the present case when both effects have been observed within the same experiment, the argument that frequency and concreteness serial position interactions are different results from concessions in data interpretation rather than the presence of distinctly different patterns. While it is possible that contextual factors, such as the choice of combination of presentation and recall modalities, may have contributed to a lack of clear difference between serial position interactions in this experiment, as it stands, the evidence for distinct redintegrative systems, as Walker and Hulme (1999) would predict, is not overly convincing.

However, the examination of serial position in terms of effects in primacy and recency portions of the curve failed to support predictions from the language-based viewpoint, as in each case the effects in the primacy positions of serial recall were not greater than those in the recency positions (Romani et al., 2008). These comparisons are also constrained to some degree by the presence of ceiling effects in the early part of the curve and the anomalies associated with recall in the LFHC condition, however it does appear unlikely that greater effects would have otherwise resulted in the primacy region. It is also possible that design features in this task, namely presentation and recall modalities, contributed to the extent to which lexical-semantic effects were displayed across the curve by limiting the masking from the phonological record for the latter serial positions (Romani et al., 2008). As previous opportunities to observe performance in this context were few, and language-based models have grown from an auditory-verbal perspective (e.g. N. Martin, Saffran, & Dell, 1996), the impacts of such task constraints may not have been widely considered.

10.5 Conclusion

Experiment 1 demonstrated that frequency and concreteness do interact in serial recall. However, the results were inconclusive regarding the determination of differences or otherwise between concreteness and frequency effects across serial positions, and therefore did not provide a robust test of the models under consideration. Accordingly, a second experiment was performed in order to better differentiate the explanatory capacities of the dual redintegration and language-based accounts of serial recall.

Chapter 11

Experiment 2: The Interaction of Frequency and Word Concreteness with Auditory Presentation and Written Recall

11.1 Introduction

Experiment 2 was a replication of Experiment 1 with a different choice of presentation and recall modalities. The motivation for this experiment was twofold, firstly relating to the possibility of a rival explanation for the existence of the frequency by concreteness interaction in Experiment 1, and secondly, a desire to reduce the ambiguity surrounding the serial position interactions observed in the experiment.

While Experiment 1 had used visual presentation with a view to placing greater emphasis on lexical-semantic representations, short-term traces from visually presented material are known to rapidly degrade, and with them their capacity to support recall (Tolan & Tehan, 1999). Furthermore, one argument regarding the differences between concrete and abstract words involves additional sensory as well as semantic encoding available for concrete but not abstract items (Paivio, 1986, 1991). It is possible therefore that the presentation modality may have encouraged the use of different recall strategies across conditions, and this would provide an alternative explanation for the observed interaction between frequency and concreteness. Specifically, when items are highly concrete, a visual imagery strategy might have attenuated the difference between HF and LF words, as reliance on processes reflecting the contribution of lexical-semantic effects would be reduced. A visual imagery strategy would be much less effective with abstract words however, leading to a greater dependence on whatever phonological and semantic features of items were retained. Accordingly, differences in the frequency effect between these conditions would be observed in these circumstances.

Hence it is possible that visual presentation emphasized a non-systematic contribution that is less likely to exist in experiments using auditory presentation. In these cases phonological encoding is direct and has greater duration than encoding from visually presented stimuli (Penney, 1989; Tolan & Tehan, 1999). Experiment 2 was designed to determine whether the interaction between frequency and concreteness would be replicated under conditions where phonological encoding, and arguably phonologically based recall strategies, had greater influence.

Given the level of performance observed in Experiment 1, the change in presentation modality raised the possibility that ceiling effects for early serial positions would eventuate. Walker and Hulme (1999) had shown that for concrete stimuli with auditory presentation, the first few items in a list are recalled less well when written, instead of spoken, recall is employed. Thus, in order to minimize the likelihood of ceiling effects operating in Experiment 2, written recall was chosen to capture output.

This selection of presentation and output modalities contained the additional benefit of testing performance for frequency and concreteness under conditions that have been reported elsewhere, for both frequency (O.C.Watkins & Watkins, 1977) and concreteness (Romani et al., 2008; Walker & Hulme, 1999). This facilitated a more direct comparison of the behaviour of effects across serial position in tasks with similar design features, and tested the generalisability of the current task manipulations with reference to these.

11.2 Method

11.2.1 Participants

Approximately 120 University of Wollongong undergraduate students participated in the experiment as a class exercise. From this initial pool, the data set was reduced for several reasons. Firstly, some data was lost due to answer sheet formatting problems. From the remaining participants, all those who had English as second language were removed. Any participants who had inadvertently skipped a trial (see procedure below) were omitted. Because this was a class-based exercise, experimental list files were used multiple times (5 unique files of each of 4 counterbalanced orders of conditions). Furthermore, the number of times each experimental list file was used varied according to class size (up to 20 students). In order to determine data for the final analysis, the number of unique files and number of participants who were tested with these files was matched against counterbalancing constraints within the remaining participant pool. List files that had less than 3 eligible participants were eliminated - this determined the greatest possible number of list files per counterbalancing condition. In cases where there were more eligible participants for a particular list file, or there were more list files for a particular counterbalancing condition, random selection of participants or file was used to determine inclusion. This process resulted in a final participant pool of 48 students (9 male, 39 female) with mean age of 22.0 years (SD = 3.2 years).

11.2.2 Materials

The stimulus sets were twenty of the list sets used in Experiment 1. However, as auditory presentation of list items was being used in this experiment, some additional considerations were required. Some items were homophones, for example *sun* and *steak*, and accordingly, it was important to establish that any changes arising from the presence of homophones in the stimuli were unlikely to change the factorial manipulation of the independent variables. This was done by examining the frequency counts of any homophones of list items, and altering their values by summing homophone frequencies. Additionally, where possible, and in the majority of instances this was the case, the concreteness ratings of the homophones were extracted, and weighted average concreteness ratings, using the individual frequency counts as the basis for the weighted contributions, were derived.

A MANOVA was performed on the amended word sets, Wilks Lambda, Λ = .023, $p < .001$, the analysis demonstrating once again that (i) HF sets differed significantly from LF sets – $F(3, 44) = 27.925$, $p < .001$, $MSE = 2784.360$, Tukey's HSD homogeneous subset analysis identified HF and LF conditions as significantly different; and (ii) high concreteness sets differed significantly from low concreteness sets – $F(3, 44) = 95.29$, $p < .001$, $MSE = 3537.866$, Tukey's HSD homogeneous subset analysis identified high and low concreteness conditions as significantly different.

The stimuli were digitally recorded in a native Australian female voice, and using the ProTools LE software on a G4 Macintosh computer converted to sound files. A response sheet that asked for demographic information (age, sex, first language) and provided spaces for written recall of list items was used for data collection.

11.2.3 Procedure

Participants were tested in groups of up to twenty and testing took approximately half an hour to complete. Each group was given oral instructions, supported by overhead material, as to how to correctly participate in the experiment. The experiment was conducted on individual IBM compatible computers connected to headphones, controlled by purpose written software. To reinforce adherence to correct participation, an instructions screen appeared prior to the participants commencing the experiment. No practice trials were given. Initiation of each trial occurred when the participant pressed the left mouse button. The program would then present each word in the trial at a rate of one word per second. After the sixth word, an auditory prompt (a *beep*) was played to indicate the commencement of the recall phase. Written recall was according to the strict serial recall criteria outlined in Experiment 1 and was self-paced. Response sheets were collected at the conclusion of each experimental session.

11.3 Results

Recall was calculated in the same manner as for Experiment 1. The mean number of correctly recalled items by serial position and condition is shown in Figure 11.1. The ranking of condition by recall performance collapsed across serial positions replicated that for Experiment 1, namely, recall being greatest for the HFHC condition (M = .760, SD = .139), followed by the HFLC condition (M = .711, SD = .149), the LFHC condition (M = .690, SD = .127), and lastly the LFLC condition (M = .595, SD = .139).

11.3.1 Serial recall

An alpha level of .05 was again the criterion for significance. A 2 x 2 x 6 (frequency x concreteness x serial position) repeated measures ANOVA was conducted on the serial recall data. Once again all main effects were significant: frequency, $F(1,47) = 61.47$, $p < .001$, $MSE = .041$; concreteness, $F(1,47) = 46.67$, $p < .001$, $MSE = .032$; and serial position, $F(5,235) = 130.85$, $p < .001$, $MSE = .089$. The serial position curves for this experiment are consistent with other experiments that have auditory presentation of material and written recall in that, when compared to those for Experiment 1, there is a marked recency effect for each condition. The frequency by concreteness interaction was found to be significant, $F(1,47) = 5.50$, $p = .023$, $MSE = .028$, and this interaction manifested in the same way as for Experiment 1. The frequency by serial position interaction was significant once more, $F(5,235) = 7.14$, $p < .001$, $MSE = .011$. Comparing recall for HF and LF conditions, the frequency effect increased for the first three positions and stayed constant until the last position where it closed again. The concreteness by serial position interaction was significant, $F(5,235) = 4.71$, $p = .001$, $MSE = .016$. The difference in recall between high and low concreteness words increased to the fourth serial position and then decreased for the final two positions. The three-way interaction was non-significant, $F(5, 235) = 1.69$, $p =.156$, $MSE = .013$.

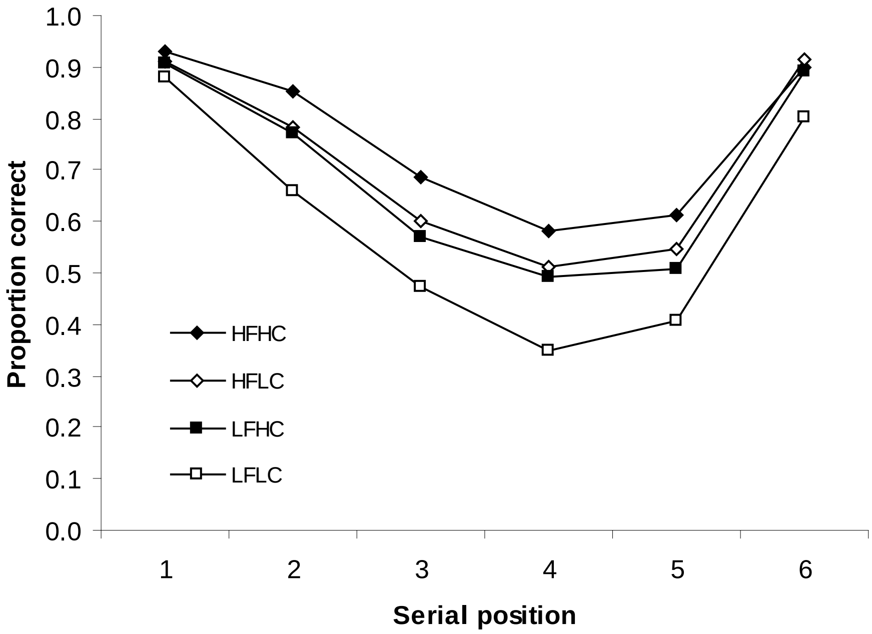


Figure 11.1. Serial recall of words as a function of frequency and concreteness with auditory presentation and written recall. HFHC - High frequency, high concreteness, HFLC – High frequency, low concreteness, LFHC - Low frequency, high concreteness, LFLC – Low frequency, low concreteness.

11.3.2 Serial position interactions

The level of performance in the HFHC condition for the first serial position indicated that there could be a ceiling effect operating on the distribution of the data. As an additional check to discount the interaction of each variable with serial position as a consequence of range-restricted data, a further repeated measures ANOVA on the last five serial positions was run. Interactions of each variable with serial position persisted despite the removal of the first position data; the frequency by serial position interaction was significant, F (4, 188) = 3.16, p = .017, MSE = .011, as was the concreteness by serial position interaction, F (4, 188) = 2.74, p < .038, MSE = .015, and the three-way interaction remained non-significant, F (4, 188) = 1.49, p =.216, MSE = .013. All other results were consistent with those determined from the full data. Despite the persistence of these interactions with serial position, a strong case can be made that these are a direct reflection of the changes in effects in the last serial position. It is apparent that both frequency and concreteness effects throughout the medial positions changed very little.

The effects observed for primacy and recency portions of the curve were compared for both frequency and concreteness. The average frequency effect for the primacy positions (M = .084, SD = .104) was not different to the average effect for recency positions (M = .103, SD = .088), t(47) = -1.39, p = .172. The comparison for the concreteness effect also yielded a non-significant result, t(47) = -0.34, p = .734, as the average effect for the primacy (M = .070, SD = .014) and recency (M = .075, SD = .084) portions of the curve was the same.

11.3.3 Item analysis

The same classification system as outlined for Experiment 1 data was employed in this experiment. A summary of the categories for item errors is presented in Table 11.1.

Conditionalising order errors yielded rates of .117 for HFHC, .150 for HFLC, .129 for LFHC and .173 for LFLC respectively. A 2 x 2 repeated measures ANOVA found significant main effects for frequency, F(1,47) = 4.60, p = .037, MSE = .003, and concreteness, F(1,47) = 20.54, p < .001, MSE = .003. Thus significantly more order errors occurred in the recall of lists with LF words and abstract words, respectively. The frequency by concreteness interaction however was non-significant, $F(1,47) = 0.47$, $p = .495$, $MSE = .003$.

The total item errors were analysed using a 2 x 2 repeated measures ANOVA. This analysis indicated that the main effects were significant; frequency, $F(1,47) = 78.71$, $p < .001$, $MSE = .188$; and concreteness, $F(1,47) = 43.58$, $p < .001$, $MSE = .095$. A significant frequency by concreteness interaction, $F(1,47) = 8.10$, $p = .007$, $MSE = .007$ was also observed. These results replicated those for Experiment 1, namely that LF and abstract words generated more item errors than HF and concrete words respectively, and that the difference in errors between LF and HF words was smaller for concrete words than for abstract words.

Table 11.1

Mean proportion of items correctly recalled and mean proportions of different error categories by condition in Experiment 2

	Correct	Order errors	Item Errors
HFHC	.760 (.131)	.097 (.061)	.143 (.100)
HFLC	.711 (.149)	.120 (.070)	.169 (.109)
LFHC	.690 (.127)	.098 (.050)	.212 (.102)
LFLC	.595 (.139)	.119 (.064)	.286 (.116)

	Item errors by category					
	Repetitions	Omissions	EEI	ISI	IEI	PA
HFHC	.010 (.013)	.092 (.090)	.004 (.013)	.025 (.023)	.001 (.004)	.011 (.020)
HFLC	.010 (.013)	.121 (.103)	.003 (.009)	.023 (.020)	.002 (.008)	.013 (.016)
LFHC	.016 (.017)	.144 (.104)	.011 (.019)	.025 (.029)	.003 (.011)	.010 (.020)
LFLC	.011 (.014)	.187 (.116)	.008 (.016)	.035 (.035)	.029 (.031)	.011 (.018)

Note. HFHC - High frequency, high concreteness, HFLC – High frequency, low concreteness, LFHC - Low frequency, high concreteness, LFLC – Low frequency, low concreteness, EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Rep – Repetitions, PA – Phonological approximations. Standard deviations are given in brackets.

As in Experiment 1, omissions made up the bulk of item errors and were analysed separately. Patterns for omissions exhibited significant main effects for both frequency, $F(1,47) = 40.68$, $p < .001$, $MSE = .148$, and concreteness, $F(1,47) = 28.64$, $p < .001$, $MSE = .078$. More omissions occurred with LF than HF items, and more omissions were evident for abstract items than concrete items. The interaction however was non-significant $F(1,47) = 0.92$, $p = .342$, $MSE = .083$. The failure to find an interaction in this case appears due to a floor effect on word sets with at least one high-end attribute.

The effects of other item errors were small in comparison to omissions, and were not subjected to inferential tests. However, it should be noted that more EEI errors were committed in LF conditions, and an order magnitude difference was present in IEI errors, driven predominantly by the substitution error of *truth* for *truce*.

11.4 Discussion

This experiment confirmed the expected main effects of frequency and concreteness and yielded serial position curves typical for a task employing auditory presentation and written recall (Romani et al., 2008; Walker & Hulme, 1999; O.C. Watkins & Watkins, 1977). Additionally, a frequency by concreteness interaction was found in Experiment 2 and followed the pattern observed in Experiment 1. The effect size for each variable was modulated by the level of the second, and effects diminished with frequently used or concrete items. Replication across two tasks varying in input and output requirements suggests that this interaction is a stable feature of STM recall. Additionally, its presence in a context where a recall strategy reliant on the visual features of words would be less productive, namely when phonological traces are stronger and the additional burden of written recall would encourage efficient output, reinforces the likelihood that the interaction between these lexical-semantic variables is an outcome of orthodox STM activity.

In this experiment, and unlike Experiment 1, variations in memory for order across conditions were observed. Memory for order was influenced independently by both variables however the magnitudes of these variations are small; between 1-2% for word frequency and 3-4% for word concreteness. Small order effects for frequency and concreteness have been reported in other experiments (Allen & Hulme, 2006; Roodenrys et al., 2002; Tse & Altarriba, 2007; Walker & Hulme, 1999). More generally order effects have been interpreted as additional evidence that the locus of effect for lexical-semantic variables extends well beyond the scope outlined by late-stage redintegration theories (Monnier & Syssau, 2008; Saint-Aubin, Ouellette, & Poirier, 2005).

The overall results for item memory replicated those for Experiment 1. Total item errors indicated that the frequency by concreteness interaction was predominantly a product of better item memory for HF and concrete words that was limited in effect when items were both concrete and frequent in language use. The effects of frequency and concreteness were evident with the pattern of omissions, presumably reflecting the relative advantages higher levels of each variable provide in either restoring or retaining STM traces. The failure to find an interaction in this error category appears due to the low levels of omissions for HF and concrete items, and is testament to the durability of short-term traces that arise from auditory presentation.

In this experiment the frequency by position and concreteness by position interactions were both significant, and are different to those in Experiment 1; the size of either effect varied little in the medial positions. It is possible that this difference may be due to the tendency to omit early medial items so that the latter items in the lists can be outputted before they are lost entirely. The increase in recall performance in the final items is similar for most conditions, except perhaps for the HFHC condition, where performance at recency may have suffered, relative to other conditions, because pre-recency items were sufficiently intact to be outputted, thus slowing the output process and increasing the degree of output interference encountered at the recency position (Hulme et al., 1999). Alternatively, given the high level of performance at the recency position, the HFHC condition may have been restricted by a ceiling effect.

Therefore in the context of this experiment, both frequency and concreteness behaved similarly when performance across serial positions is considered. Each was also found to exhibit the same size of effect in primacy and recency portions of the curve. The frequency by serial position interaction can be compared with that found by O.C. Watkins and Watkins (1977) with auditory presentation. These authors determined that the frequency effect was greater in the primacy positions than the recency positions, and specifically was absent for the last two positions. This contrasts with the present findings where the effects of frequency can be seen for all positions but the first. With respect to the concreteness by serial position interaction, two additional experiments are relevant; Walker and Hulme's (1999) Experiment 2 and Romani et al.'s (2008) Experiment 1B. The current results are closer to those of Walker and Hulme (1999) in that the effect is present in the penultimate position but not the last, where this is reversed in Romani et al.'s (2008) experiment. Furthermore, Romani et al.'s (2008) finding that the concreteness effect was greater in the primacy than recency region was not replicated in Experiment 2. It would appear that the claim regarding the reduction of lexical-semantic effects prior to the last serial position is not as generalised as Romani et al. (2008) imply, although the results in this instance do support their contention that serial position interactions for frequency and concreteness should be similar.

11.4.1 A statistical analysis of the generalisability of the effects found in Experiment 2

A final analysis was performed to ascertain whether the discrepancy between the frequency by serial position and concreteness by serial position interactions found in the present studies and those of Walker and Hulme (1999) was a result of a lack of statistical power in their experiments, where sample sizes of sixteen participants were used, and in one experiment the concreteness by position interaction was found to be non-significant. To test this idea, a series of analyses was run on the data from Experiment 2 to simulate the pattern of results that could be generated from samples of this size ($N = 16$). The analyses involved a random selection of 16 participants' data, without replacement, and the calculation of the F-statistics for all effects. This was repeated 1000 times. The distributions of each set of results were examined to gauge the likelihood that a significant interaction between concreteness and serial position could be obtained with a single sample of 16 participants. For Experiment 2 this was possible for 42% of the cases. Therefore with a sample size the same as Walker and Hulme (1999) there was less than a 1 in 2 chance of obtaining a significant interaction. On the basis of this demonstration it appears probable that smaller effects, such as those associated with concreteness, require greater sampling in order to be detected and for the concreteness by serial position interaction to stabilise.

11.5 Discussion of Experiments 1 and 2

11.5.1 Explanatory frameworks

Experiments 1 and 2 were in agreement regarding the pattern of results for correct recall; there is evidence that the effects of word frequency and word concreteness combine in the short-term recall of verbal material. Specifically, the nature of this combination is such that the more concrete the stimuli, the smaller the effect of word frequency. The structure of the language-based models of STM (R.C. Martin et al., 1999; Romani et al., 2008), and the inherent properties that derive from them, are compatible with the finding of an interaction in the data if it is assumed that a limit exists in the level of activation a lexical node can realise.

Using the structure of these models as a framework, it is proposed that word frequency affects the strength of activation at the lexical layer in the long-term knowledge store (R.C. Martin et al., 1999). With respect to pure lists, an advantage to HF items will occur because they activate the lexical nodes for these items more strongly than for LF words, thus providing stronger feedback activation to the linked nodes at the phonological level. Concreteness, reflecting the richness of semantic representation, is a marker of the quantity of semantic nodes that connect with a lexical node (R.C. Martin & Lesch, 1996). When a semantic feature is activated it provides feedback activation to lexical entries that are connected to that semantic feature. While feedback will return to the lexical node that activated it in the first place, it will also activate semantic competitors at the lexical level. Therefore, those lexical items with more semantic features will be better reinforced by interactive activation, and this in turn will assist the preservation of activated phonological representations in LTM. As concrete words provide greater feedback activation from the semantic level to the initiating lexical nodes than abstract words, concreteness effects should be evident between conditions that vary on this attribute. Furthermore, if the level of activation of nodes in the lexical layer has some upper bound, then a frequency by concreteness interaction of the type observed could result. That is, the activation levels experienced by lexical nodes of words that are both highly frequent and concrete would be less than the sum of activation levels achieved by items that are highly frequent or concrete alone.

The existence of a frequency by concreteness interaction with pure lists can also be accommodated within a redintegration account that assumes separate redintegrative capacities for phonological and semantic information (Walker & Hulme, 1999). Word frequency affects the redintegration of short-term phonological traces through the accessibility and availability of phonological long-term representations, either in an item-specific (Hulme et al., 1997; Saint-Aubin &

LeBlanc, 2005; Saint-Aubin & Poirier, 2005) or associative (Hulme et al., 2003; Stuart & Hulme, 2000) manner. Word concreteness influences the strength of semantic representation and the uniqueness of semantic features (Walker & Hulme, 1999), and therefore impacts on the likelihood that short-term semantic codes are correctly redintegrated. It is assumed that as items increase in either frequency or concreteness the likelihood of successful redintegration from phonological or semantic mechanisms, respectively, increases. Therefore, items that are both frequent in use and highly concrete will benefit less in relative terms, because an increasing proportion of these will be reconstructed from both processes.

11.5.2 Differences between the processing of frequency and concreteness in serial recall: Are they real?

11.5.2.1 The nature of the frequency by serial position interaction

Despite the plausibility of a frequency by concreteness interaction within a system that contains separate redintegrative capacities, Walker and Hulme's (1999) assertion that separate redintegrative mechanisms are evidenced by the distinct signatures of frequency and concreteness effects across serial position was not supported in these experiments. Although Experiment 1 was inconclusive with regard to genuine differences between serial position interactions for frequency and concreteness, Experiment 2 produced interactions with serial position that were similar for both variables and invariant across the medial positions of the recall curve, as proponents of language-based models have suggested they should be (Romani et al., 2008).

It is therefore important to address the apparent conflict between observations made here and those made by Walker and Hulme (1999), and offer an explanation as to their origin. A survey of the serial recall literature involving lists of pure word frequency reveals that earlier experiments (e.g. Hulme et al., 1997) observing a definite increase in effect across serial positions until the last item, used very small stimulus sets, and high numbers of trials. For example, the serial recall experiments of Hulme et al. (1997) employed sets of eight items per condition, and each condition constituted 25 seven-item trials. Thus, each item was presented throughout the course of the experiment on average 22 times. Typically, the number of presentations for each item in the majority of other reported experiments is less than 10 and in the case of open sets of words this reduces to one or two presentations. Furthermore, examination of the serial position interactions in experiments where the number of item presentations is not large does not reveal a consistent pattern of performance. Ceiling effects in the first and sometimes second serial position are often evident, leading to a possible masking of actual effect size in the primacy positions and the artificial creation of statistically significant interactions. The regularity of the increase in effect across serial positions varies, and in some instances diminishes further into the list. Therefore apparent qualitative differences in frequency by serial position interactions exist in reported experiments, and these differences loosely correlate with how often items are used within an experiment.

This is perhaps most clearly demonstrated in a comparison between the frequency by serial position interactions for Hulme et al.'s (2003) Experiment 1 and Experiment 2. The first of these experiments used eight-item closed sets from Hulme et al. (1997) in a comparison of recall performance between six-item lists of pure and alternating frequency. The recall curves for pure lists replicate the pattern identified by Hulme et al. (1997) and display a monotonically increasing effect from positions one through to five. In contrast, the second experiment, a replication of the first using an open set of stimuli where items were presented twice in the life of the task, reveals curves for the recall of pure lists that do not possess this feature. This data is arguably affected by a ceiling effect on the first serial position masking the true level of performance for HF lists and the effect sizes are constant for the medial serial positions[1]. Furthermore, the experiments reported by Hulme et al. (2003) did not block conditions, and so presentation effects of set size on serial position interactions would appear to survive the intermittent presentation of items belonging to specific conditions.

This difference between serial position interactions resulting from the use of closed and open item pools is suggested to result from an improvement in recall in the early serial positions for LF words. Research involving pre-exposure to pairs of items or items individually (Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000), has shown that recall for familiarised LF items is markedly greater than recall of unfamilarised LF items. It seems likely that intra-experimental familiarity effects

[1] It is notable that the pure list serial recall curves presented in Experiment 3, Experiment 5, and Experiment 6 of this thesis conform to this pattern and were products of open stimulus sets.

occur in circumstances that incorporate the repeated presentation of items. The improved recall for early serial positions might well reflect the incompleteness of familiarisation in these circumstances, and an inability to sustain processing benefits throughout the list. Thus, the oft-cited serial position interaction for word frequency (e.g. Hulme et al., 1997) may in reality be an artefact of experimental method.

The stimuli in the serial recall experiments of Walker and Hulme (1999) were presented 7 times within their respective conditions. The current experiments used each stimulus item 6 times. Therefore, if a corresponding presentation effect were to be present in sets manipulating levels of concreteness, then these experiments are adequately matched. While Romani et al. (2008) used open sets of words for their serial recall experiment the pattern observed is similar. Furthermore, a difference in this case relates to the lack of concreteness effect at the penultimate position with open sets and is therefore one unlikely to be related to the multiple presentations of items.

11.5.2.2 Backwards recall

One final distinction made by Walker and Hulme (1999) regarding differences in short-term processing between items varying in frequency and concreteness was in relation to backwards recall. Hulme et al. (1997) had observed no effect of frequency in this task, while Walker and Hulme (1999) found a facilitative effect of concreteness. These results were presumed to reflect the use of a semantic retrieval strategy in task execution. While no direct disconfirmation of this position is available at present, given recent research identifying conditions under which the effects of lexical-semantic variables can go undetected, and variations in the size of stimulus sets used by Hulme et al. (1997) and Walker and Hulme (1999) respectively, this contrast too should perhaps be treated with suspicion. Hulme et al. (1997) had again used closed sets of 8 items for blocks of 25 trials per condition in a design that tested the forward and backwards recall in lists of differing frequency. Therefore, depending upon the counterbalancing of conditions, some participants would have been presented with each item 44 times by the end of the backward recall trials. In contrast, each stimulus in the backwards recall experiment of Walker and Hulme (1999) was presented only 7 times across the condition to which it belonged. Here the absence of effect for word frequency may reflect the influence of increased activation of LF words due to repeated presentation rather than the adoption of a semantic retrieval strategy per se. Furthermore, if a frequency effect was shown to be present under conditions similar to those for Walker and Hulme’s (1999) word concreteness task, a more general lexical-semantic influence over backwards recall performance might be responsible, and relate to the degree to which these variables assist in maintaining the intactness of phonological traces (Romani et al., 2008).

11.5.3 Order effects

Experiments 1 and 2 differed with respect to the results for memory for order. Experiment 1 indicated no involvement of lexical-semantic variables in the sequencing of list items, where Experiment 2 showed small effects for frequency and concreteness with the reordering of items more evident for LF and abstract words, respectively. In the context of redintegration theory, the presence of an order effect associated with semantic similarity has been previously explained by the admission of semantic features to the short-term trace (Saint-Aubin et al., 2005) and Walker and Hulme’s (1999) extension of redintegration includes short-term semantic encoding. While semantic similarity is claimed to increase the confusability of items and compromise memory for order, concreteness is argued to determine the uniqueness of the semantic representation (Walker & Hulme, 1999). If this is the case, then degraded short-term traces of abstract items might be more similar than degraded short-term traces of concrete words, thereby facilitating better order memory for concrete words.

The language-based models presented here admit the influence of lexical-semantic variables from the point of item presentation onwards, but their role is primarily the maintenance of item identity through interactive activation in LTM. In the model described by Romani et al. (2008) the management of serial order occurs through the maintenance of buffered phonological representations in the placeholder that interact with LTM at the phonological level. These in turn are supported by connected lexical-semantic activation. Small order effects might reflect perturbations arising from the interaction of the contents of the placeholder with the phonological representations in LTM as a result of supportive lexical-semantic activation, however this explanation would benefit from greater specification of the serial order mechanism overseeing the preservation of order memory (Majerus, 2009).

11.5.4 The loci of lexical-semantic effects

While the current experiments have yielded results that imply lexical-semantic effects exist at output, they do not directly address specific processes that are responsible for the encoding and maintenance of representations. A number of ERP studies have produced indications that lexical-semantic activation may operate in the serial recall process from the point of encoding onwards (Cameron et al., 2005; Ruchkin et al., 1999; Ruchkin et al., 2003), as language processing models of STM assume (N. Martin & Saffran, 1997; N. Martin, 2009; R.C. Martin et al., 1999; Romani et al., 2008). However, other researchers (Thorn, et al., 2009; Thorn, et al., 2005) have suggested, based on close examination of item recall errors, that variables such as language familiarity, word frequency and phonotactic frequency have an influence on the retention of items as well as their production in serial recall, while the effect of lexicality, that is differential performance for words versus nonwords, is present in serial recall, but not earlier as indexed by the numbers of completely incorrect recall attempts. In consequence, Thorn et al. (2009; Thorn et al., 2005) are less convinced that lexical-semantic variables play an integrated and on ongoing role in the retention of items before recall, and propose that short-term memory activity is punctuated by multiple mechanisms at differing stages of the task, involving different LTM variables. This account therefore has similarities with the redintegration model of Walker and Hulme (1999) in that it identifies specific roles and processes for LTM variables, although it relies on alternative data to argue for the separate treatment of these, and so far has been confined to variables with a phonological base.

11.5.5 Interactions between lexical-semantic variables

It is worth considering how the frequency by concreteness interaction sits in relation to the findings of Tse and Altarriba (2007). These researchers found that frequency interacted with LSA cosine such that an LSA-cosine effect was present for LF but not HF lists, while an LSA cosine effect was invariant for high and low imageability lists. The LSA measure had been used as an index of the strength of inter-item associativity in these experiments, however as discussed in Section 5.4.3.2 this parameter might be better defined as a semantic relatedness variable than a measure of the co-occurrence of items in language use. Should this be the case, the interaction found between LSA-cosine and word frequency bears some resemblance to the pattern observed between frequency and concreteness in Experiments 1 and 2. That is, a much smaller concreteness effect for HF in comparison to LF words is consistent with the finding of an effect of LSA-cosine for LF but not for HF lists. Given this similarity and the correspondence between concreteness and imageability, it is possible that concreteness and LSA-cosine reflect different aspects of semantic processing that interact with the lexical representations of items. In contrast, as LSA-cosine and imageability were found to be independent this suggests that these semantic properties are represented and processed separately.

11.5.6 Practical considerations

Lastly, on a practical note, the interaction between frequency and concreteness suggests that the domain over which stimuli are controlled for in one variable will determine the size of the manipulated variable's effect (Miller, 2004). This moderation may explain, in part, variations in effect sizes observed between experiments. Furthermore, experimenters should be aware that effects can be smaller than anticipated if their control of the other variable is at the higher end of the respective measurement scale.

11.6 Conclusion

In summary, Experiments 1 and 2 confirmed the presence of a frequency by concreteness interaction in serial recall. The results of these experiments were examined in relation to two explanatory frameworks that considered frequency and concreteness separately; the dual redintegration framework of Walker and Hulme (1999) and the language-based model of Romani et al. (2008). Each position is capable of explaining the presence of an interaction between these lexical-semantic variables, however the redintegration account has greater difficulty accommodating the observations of similar interactions with serial position for both variables. Indeed the distinction between serial position interactions is a corner stone of the dual redintegration theory (Walker & Hulme, 1999). Closer inspection of the previous experimentation on which this distinction was based indicates that confounds with set size may be responsible for differences in observed patterns, supporting the findings of the current work and the predictions drawn from models with a language-based architecture (Romani et al., 2008), namely that serial position interactions of lexical-semantic variables should take on a similar form. However, while effects for medial positions were found to be constant, the suggestion that lexical-semantic effects are eliminated in the final two positions of serial recall, and effects are greater in the primacy than recency portions of the curve, did not generalise to the current experiments. It is likely that additional experiment-specific factors influence the extent to which lexical-semantic effects are reduced in the recency portion of the curve.

Chapter 12

Experiment 3 - A Replication of Hulme et al. (2003) Experiment 2

12.1 Introduction

The second facet of this thesis concerns how the effect of frequency of each list item affects recall across the serial positions. The frequency effect in mixed lists possesses different manifestations, dependent upon variations in list composition and item arrangement (Hulme et al., 2003; Miller, 2004; Morin et al., 2006; Roodenrys et al., 2002; Roodenrys, unpublished; Saint-Aubin & LeBlanc, 2005). In circumstances where HF and LF items are alternated, the frequency effect is eliminated (Hulme et al., 2003; Morin et al., 2006). In these lists LF words are recalled as well as HF words, particularly when item information is considered. In isolate experiments, individual HF items embedded in an LF list produce a frequency effect, as their recall exceeds that of LF items in a pure list (Roodenrys, unpublished; Saint-Aubin & LeBlanc, 2005). However, the recall of HF isolates is not as great as the recall of HF items when they are recalled with other HF items in a pure list. An LF isolate within a HF list fails to produce a frequency effect between list items as the isolate is recalled as well as HF items in pure lists, although a small effect may be detectable with greater power (see section 6.3, Saint-Aubin & LeBlanc, 2005). When compared to the level of recall produced by LF items in pure lists however, this difference is marked.

When the order (and number) of HF and LF items within a list is randomised the frequency effect can vary in magnitude (Miller, 2004; Roodenrys et al., 2002). Drawing on the mixed list results obtained in the literature thus far it might be argued that modulation of the effect might be due to variation in the distributions of the stimuli used, and how these identify items that are more likely to be 'isolates'. The mean frequency of items used by Roodenrys et al. (2002) was reported to be 41.4 (*SD* = 104.6), while those used by Miller (2004) had a mean frequency of 44.4 (*SD* = 70.72). The stimuli of Roodenrys et al. (2002) did contain more LF items (items with counts less than 10) than the Miller (2004) set and this may have created an LF background against which HF items might have been better recalled. However, the lists used here did not necessarily conform to isolate conditions *per se*, so other variables might have contributed to the observed difference. The frequency effect may well be a composite effect resulting from the interplay of several factors (Tse & Altarriba, 2007).

Given the complexity associated with randomly constructed lists, where list composition (in terms of the number of HF and LF items) and item arrangement are varied together, the number of testable list forms is large, and variations in outcomes might also arise from associated random differences in list types between experiments, it was decided to adopt a focused approach in the current mixed list studies by pursuing the line of research commenced by Hulme et al. (2003). In addition, this line of investigation would facilitate integration with previous literature. Within this approach list composition is fixed and balanced, and item arrangement is varied to examine differences in recall. This direction therefore would further inform the research base regarding the contributions arising from the relationships between list items in mixed lists, and further the theoretical development of the nature of inter-item associations in serial recall. Pursuit of this aspect of study was thought the most profitable, as manipulations in list composition had already been performed, albeit across studies (e.g. Hulme et al., 2003; Saint-Aubin & LeBlanc, 2005), and information on item arrangement would be of a complementary nature.

According to Hulme et al. (2003; Stuart & Hulme, 2000), if the pre-existing inter-item associations between list items form a mutually supporting redintegrative network in a non-directional and nonspecific way, then all lists containing equal numbers of HF and LF items should produce recall at a level intermediate to that for pure HF and pure LF lists, and HF and LF items should be recalled at equal levels across serial positions. In contrast, if the inter-item associations are sensitive to the order of items in the list, as a TCM (Howard & Kahana, 2002a, 2002b) approach to redintegration would argue, then recall patterns should be different for lists that vary item arrangement.

The difference in the size of the frequency effect when the number of HF and LF items is systematically varied (isolate, alternating and pure lists) has been explained in terms of item distinctiveness combined with broader processing effects that are a function of the list composition (Saint-Aubin & LeBlanc, 2005). No directional effects with this explanation are assumed. Accordingly, a distinctiveness plus processing account of lists containing equal numbers of HF and LF items would predict the same recall performance across item arrangements as neither the distinctiveness of items, nor the processing requirements should change.

In the absence of additional indications that the frequency of an item influences how well that item is recalled in position, namely order errors, language-based approaches to serial recall (e.g. R.C. Martin et al., 1999; Romani et al., 2008) would not offer any differentiating predictions. The nature of long-term representational support (a mutually supportive versus directionally supportive activation of presented items) could be accommodated by links between lexical entries that varying according to their co-occurrence in language.

12.2 Experiment 3 – Aims and objectives

In order to test the effect of item arrangement in mixed lists, it was necessary to construct stimulus sets and demonstrate that they were capable of reproducing the pattern of results with pure and alternating lists as reported by Hulme et al. (2003, Experiments 1 and 2). Therefore Experiment 3 was a replication of Hulme et al.'s (2003) Experiment 2 where open sets of words were used. This experiment extended the generality of the frequency effect with the use of CVC stimuli (Hulme et al., 2003 had used 2-3 syllable words).

12.3 Method

12.3.1 Participants

Forty-one undergraduate University of Wollongong students (34 females, 7 males) participated in the experiment for course credit. The data from 5 participants were excluded from analysis because they were either not native Australian-English speakers (4) or reported hearing difficulties (1). The mean age of the remaining 36 participants was 21.9 years (*SD* = 7.4 years).

12.3.2 Materials

Two stimulus sets of 96 CVC words were created based on the manipulation of word frequency. Frequency ratings were taken from the Celex database (Baayen et al., 1993) and calculated in the same manner as the stimuli for Experiment 1. Additionally, care was taken to ensure that the frequency counts of homophones were included in the manipulation of frequency. The LF set had a mean log_{10} frequency rating of 0.70 per million, and a standard deviation of 0.36, while the HF set had a mean log_{10} frequency rating of 2.21, with a standard deviation of 0.27. The word sets were matched on concreteness, using values obtained from the MRC database (Coltheart, 1981), and in cases where homophones existed a weighted average concreteness value was calculated using the relative frequencies as weights. Sets were matched on the number of phonological neighbours, as calculated from an in-house Celex-based program, and the number of phonological neighbours present within each word set was also determined. In addition, the stimulus sets were matched on phonological similarity using the PSIMETRICA methodology (Mueller et al., 2003). Differences between sets as a function of vowel quality (short and long vowels, and diphthongs) were identified (DeCara & Goswami, 2002).

A series of nonparametric tests were run on the word set attributes to identify where differences existed. Word sets were shown to be significantly different with respect to frequency ratings, *Mann-Whitney* $U = 0.00$, $p = .000$. There was no difference between words sets for phonological neighborhood – $U = 4560.50$, $p = .902$, concreteness - $U = 4495.50$, $p = .770$, or any of the phonological similarity measures – onset, $U = 4117.00$, $p = .202$, nucleus, $U = 4398.00$, $p = .585$, and coda, $U = 4365.00$, $p = .527$. A comparison between the number of phonological neighbors of each item within their respective word sets revealed no difference between LF and HF words – $U = 4555.50$, $p = .890$.

The number of words with short, long or diphthong vowels for LF and HF sets was analysed using a χ^2 test for independence. The result indicated that there was a difference in the proportion of vowel type by frequency group, χ^2 (3) = 7.04, $p = .030$; specifically there were more words with short vowels, and less words with long vowels in the LF set. While several studies dispute a direct link between speech rate and recall (see G.D.A. Brown & Hulme, 1995; Hulme, Newton, Cowan, Stuart, &

Brown, 1999; Tehan & Humphreys, 1988; but see Woodward et al., 2008 for an alternate view on speech and memory processes), any marked duration-based advantage for recall in this case should operate in opposition to the standard frequency effect for pure lists, and thus be detectable in the form of a smaller than usual effect. Furthermore, as HF items are typically articulated faster than LF items of the same length, the difference in proportions of items with short and long vowels would go partway to controlling differences in speech rate. The word sets, word attributes and summary statistics are presented in Appendix D.

The stimuli were digitally recorded in a native Australian female voice, and using the ProTools LE software on a G4 Macintosh computer converted to sound files. Experimental sessions were run on an IBM compatible PC using prepared script files that were loaded into SuperLab v. 2.0.4. Amplification of the sound files was via an external speaker attached to the computer.

Each script file contained 64, six word trials testing four conditions (pure HF lists, pure LF lists, alternating lists beginning with a HF item (HL), and alternating lists beginning with an LF item (LH)). In these trials, each item of the LF and HF word sets was presented twice; once in a pure list and once in alternating list formats. Allocation of items to serial positions was random within the constraints of each list format, and the presentation of individual trials by condition was randomised.

12.3.3 Procedure

All participants were tested individually. The total time to complete the experiments was approximately 45 minutes. Initiation of each trial occurred when the participant pressed the spacebar of the keyboard. The program would then play the sound files of each word in the trial at a rate of one word per second. After the presentation of the sixth word, a recall prompt appeared (“?????”) on the screen indicating that participants should start the recall phase. Spoken recall was according to strict serial recall criteria, that is, (i) words were recalled in order of presentation; (ii) if a word could not be recalled the participant would indicate by saying ‘blank’; and (iii) previous items were not to be recalled after moving on to successive items in the list. The experimental program presented four practice trials that were used to familiarise participants with the task requirements before the commencement of the experimental phase.

At the conclusion of the experiment, participants were given a list of 12 pairs of HF words and 12 pairs of LF words, randomly selected from the experimental pools such that all items were sampled every four participants. The HF and LF word pairs were randomly arranged within the list. The participants were asked to repeat each word pair 10 times as quickly as possible. The length of time required to complete each set of repetitions was taken with a stopwatch and recorded in milliseconds. This data was converted to measures reflecting the speech rate of items by frequency type, in terms of words per second.

12.4 Results

For each participant and each condition, recall was scored and collapsed across trials to provide the number of correct items by serial position. Correct recall was scored using a strict criterion, namely responses were considered correct if a word was recalled in the position corresponding to its serial order in presentation. The mean number of correctly recalled items by serial position and condition is shown in Figure 12.1.

12.4.1 Serial recall

An alpha level of .05 was applied to the following statistical tests. A 4 x 6 (list type x serial position) repeated measures ANOVA was conducted on the serial recall data. Both main effects were significant: list type, $F(3,105) = 43.87$, $p < .001$, $MSE = .022$, and serial position, $F(5,175) = 109.63$, $p < .001$, $MSE = .089$. The effect of list type was driven by differences in recall between pure HF, mixed, and pure LF lists, while the effect of serial position arose from the general pattern observed across conditions, where recall decreased from the first to the fifth item but showed a marked recovery at the recency position. In addition, a significant interaction between list type and serial position was found, $F(15,525) = 1.74$, $p = .040$, $MSE = .011$, reflecting differences in the frequency effect between pure and mixed lists across positions; pure lists produced a somewhat constant differential across the recall curve, while the observed difference between mixed lists reversed in sign at the last position.

Following Hulme et al. (2003), the interaction was explored further by running separate ANOVAs on the recall data for pure and mixed lists. The analysis for pure lists revealed a significant frequency effect, $F(1, 35) = 110.56$, $p < .001$, $MSE = .026$, a significant effect of position, $F(5, 175) = 99.08$, $p < .001$, $MSE = .043$, but a non-

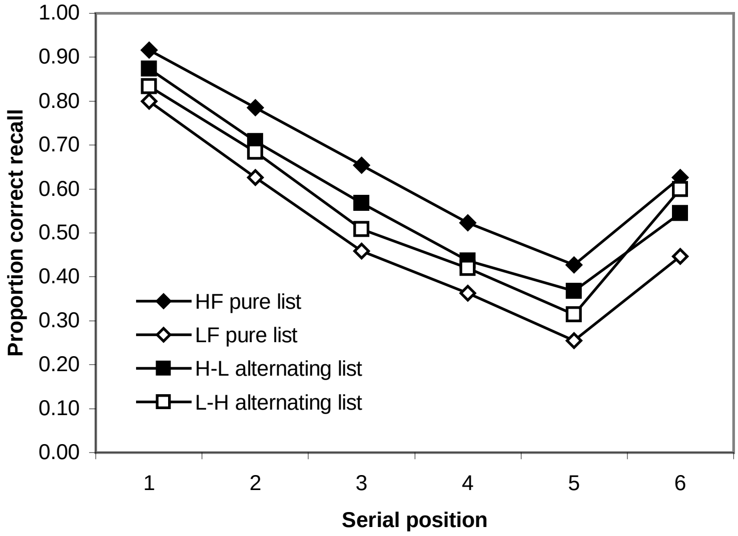


Figure 12.1. Serial recall of words as a function of frequency and list composition in Experiment 3.

significant list type by position interaction, $F(5, 175) = 1.09$, $p = .368$, $MSE = .012$. In contrast, the recall performances on mixed lists was not affected by list type, $F(1, 35) = 2.70$, $p = .109$, $MSE = .022$, but a significant effect of position, $F(5, 175) = 79.58$, $p < .001$, $MSE = .052$, and a significant list type by position interaction, F (5, 175) = 3.08, $p = .011$, $MSE = .012$, were found. The absence of main effect of list type coupled with the significant interaction was driven by performance of the list types at the last position however, where the recall curves were seen to cross over. Re-analysis on the first 5 serial positions found a significant effect of list type, $F(1, 35) = 5.89$, $p = .021$, $MSE = .023$, a significant effect of position, $F(4, 140) = 106.98$, $p < .001$, $MSE = .046$, and a non-significant interaction, $F(4, 140) = 0.66$, $p = .700$, $MSE = .011$. That is, recall for lists beginning with HF items was reliably higher than for lists beginning with LF items across the first five positions in the list.

Despite the significant HF advantage across these serial positions, it is obviously small; the superior recall of HF words is much affected by their co-presentation with LF words. It is also apparent that the mixed lists exhibit a weak sawtooth pattern across serial positions, possibly reflecting an item-specific component to recall. To test this possibility, a post-hoc contrast was run comparing the differences in recall between alternating list types for odd and even serial positions. This contrast would therefore compare the serial positions where alternating lists diverge and converge, respectively. This contrast was performed using the statistical analysis software PSY (Bird, Hadzi-Pavlovic, & Isaac, 2000). The differences between odd and even serial positions across all serial positions was found to be significant, $F(1,35) = 10.32$, $p = .003$, $MSE = .016$, supporting the likelihood that a small item-specific component to recall was present in this case. However, performances on the final serial position determined the significance of this finding; a corresponding contrast that examined differences for the first five serial positions by adjusting contrast coefficients revealed for these serial positions the effect was not reliable, $F(1,35) = 2.13$, $p = .153$, $MSE = .019$.

Speech rate scores identified a small but reliable difference between HF (mean 3.00 words per second) and LF items (mean 2.88 words per second; $t(35) = 3.87$, $p < .001$). The speech rate data was included as a covariate in a re-analysis of the above data, to determine whether variations in the articulation of items, by item type, altered the pattern of results produced. However, before presenting these analyses, it is appropriate to address concerns raised more recently in the literature regarding the salience of articulation duration measures, their possible source, and the implications for interpreting serial recall results (Woodward et al., 2008).

12.4.2 A coarticulation explanation of the frequency effect

Woodward et al. (2008) proposed that the frequency effect in serial recall could be explained in terms of differences in the familiarity of articulation of items at the word boundaries. They argued that HF words might appear in more articulatory contexts than LF words resulting in the better negotiation of the articulatory transitions between words in the lists used in STM tasks. They gave participants a serial recall task using the short HF and LF stimuli of Hulme et al. (1997) that was followed by the measurement of articulatory duration of single items, pairs of items and sequences of 6 items using a reading aloud task. Measures of single and pairwise item duration involved the length of time taken to perform 10 repetitions. The serial recall task produced a frequency effect with HF items better recalled than LF items.

The articulation duration data showed that HF words were produced reliably better than LF items, and all forms of duration measure revealed frequency effects. However, the effect for single item duration was less than for the pair or six-item sequences.

Based on a series of previous experiments where Woodward et al. (2008) had manipulated item familiarity and produced articulation duration measures that shadowed serial recall performance, they argued that insufficient consideration of articulation duration had been given to previous studies of word frequency. They questioned whether the frequency effect was adequately explained by a redintegrative account assuming differences in LTM representations, as differences in articulatory fluency correlated with serial recall performance, and fluency was likely to impact the efficiency of rehearsal in serial recall. Rehearsal was viewed as the repeated engagement of a list level speech output plan to be used during the output phase, and more difficult articulatory transitions between items would degrade speech plan quality for this purpose; more complex assemblies between items would result in more time taken to produce them and lead to poorer recall performance. Therefore, serial recall performance might better be viewed as an activity that depends on "general perceptual and motor planning processes, rather than reflecting the operation of bespoke stores and processes" (Woodward et al., 2008, p.63).

These authors do raise the possibility that the efficiency of speech in the production of a sequence of items may play a larger role in serial recall than previously acknowledged. However, there are a number of considerations to make regarding their strong claim that serial recall does not rely on mnemonic processes as such. A basis for the experiments performed by Woodward et al. (2008) was the assertion by Murray and Jones (2002) that the ease of coarticulation between items facilitated serial recall, as determined by performance in a serial order reconstruction task using stimuli matched on duration and frequency. However, examination of their high- and low- articulatory complexity stimuli revealed that a confound of phonological neighbourhood size existed (low complexity – $M = 31.38$, high complexity – $M = 17.75$, $t(14) = 3.02$, $p = .009$). A replication of this experiment with different stimuli contained a similar confound with phonological neighbourhood size (low complexity – $M = 36.63$, high complexity – $M = 22.38$, $t(14) = 5.24$, $p < .001$). Phonological neighbourhood size has been found to influence serial order reconstruction performance (L. Clarkson & S. Roodenrys, personal communication, September 9, 2009). Therefore, it is difficult to unambiguously attribute the better performance observed with low-complexity items to better articulatory fluency.

Secondly, Woodward et al.'s (2008) position rests on the assumption that the tasks used to obtain articulatory duration measures replicate serial recall conditions without memory load, and nothing else. If processes other than those used in the serial recall task are involved, and these are also sensitive to familiarisation, differences in articulatory duration due to these factors might be misattributed to articulatory fluency. For example, fixation times in reading differ between HF and LF words (Raney & Rayner, 1995; Monsell, 1991), and lexical decision and naming tasks performance suggest that identification processes (e.g. access to orthographic representations, mapping from orthographic to phonological representations) are slower for LF items (Monsell, 1991; Monsell et al., 1989; Sears, Siakaluk, Chow, & Buchanan, 2008). Cumulative effects are likely to result for longer sequences of items, and cannot be teased apart from speech production in the same manner that speech onset latencies can be removed in single word production. Furthermore, other experiments performed by Woodward et al. (2008) involved nonword stimuli where a serial recall task, used to manipulate familiarity, was sandwiched between reading aloud tasks used to obtain articulation measures, and this sequence of activity might have facilitated cross-modal priming, leading to differences in reading efficiencies that went unacknowledged. In tasks where restricted item sets were familiarised separately the transfer of set size effects to recognition processes might also have contributed to differences in articulatory durations for pure and mixed sequences post-familiarisation.

These authors also examined how training on a specific nonword set using speeded spoken production affects the articulatory duration of the same item set, a different item set with the same articulatory transitions across word boundaries, and a set with no common word boundaries. Woodward et al. (2008) obtained transfer of performance to the item set that shared common item boundaries, although duration improvement was not as great as for the familiarised set. A second procedure where serial recall was conducted either side of the speeded spoken production task showed that familiarisation with a nonword set containing common item boundaries could also facilitate recall. While these results are consistent with the possibility that better coarticulation of items leads to better serial recall, items with common word boundaries will be necessarily related in terms of neighbourhood membership, and consequently a mnemonic contribution to performance cannot be ruled out.

Lastly, Woodward et al. (2008) claimed that articulatory fluency acts on the process of rehearsal in serial recall, and differences in rehearsal result in differential recall performance. Yet it has been shown on more than one occasion that a frequency effect in serial recall persists under articulatory suppression (Gregg et al., 1989; Tehan & Humphreys, 1988). While the use of suppression throughout presentation and recall has resulted in a smaller frequency effect than the effect under quiet conditions (Gregg et al., 1989), and this conceivably might correspond to word fluency differences, the frequency effect has not been eliminated as a strong interpretation of the coarticulation hypothesis would predict.

Accordingly, while Woodward et al. (2008) have reliably produced familiarity effects in both articulatory duration and serial recall measures, it would appear premature to recommend that firstly, these effects are necessarily the products of the same mechanisms, and that consequently the influences of memory processes are discounted altogether in the serial recall task. However, given that the differences in word boundaries between item sets might contribute to performance differences, post-hoc examinations of the item sets used in Experiments 3-6 were conducted to determine whether it was likely that this factor played a role in the current tasks.

12.4.2.1 Tests of word boundary differences for the HF and LF word sets used in Experiments 3 – 6

12.4.2.1.1 Phonetic analysis

Murray and Jones (2002) conducted a broad analysis of articulatory complexity of the word boundaries for Welsh and English digits. This involved phonetic analysis for all coda-onset combinations for the digits 1-9 (72 boundaries in total) in each language. They suggested that the better performance for English digit span could be attributed to differences between Welsh and English digits in terms of place of articulation, particularly with respect to differences in the numbers of boundaries with the same place of articulation and the numbers of major place changes when negotiating word boundaries. On the basis of this analysis they argued that the articulation of sequences of Welsh digits may involve more changes than the articulation of sequences of English digits.

The possibility that articulatory complexity contributed to differences in the serial recall of HF and LF words in Experiments 3 – 6 was investigated by conducting a corresponding phonetic analysis on the stimulus sets used. As each set contained 96 items, 9120 word boundaries were assessed in terms of voicing, manner of articulation and place of articulation. As all stimuli adhered to CVC structure, this analysis did not involve CV, VC or VV boundaries. The results of this examination are presented in Table 12.1, where each figure represents the number of transitions of each type. The manner of articulation refers to movement between the lower lip and the tongue as active articulators, and the upper surfaces of the oral tract when a speech sound is produced (Davenport & Hannahs, 1998). The place of articulation refers to the location in the oral tract responsible for the production of the sound. Voicing indicates whether the production of a speech sound has involved the vocal chords; for example the ‘v’ in *van* is voiced, while the ‘f’ in *fan* is not.

Table 12.1
Phonetic characteristics of word boundaries for the HF and LF word sets used in Experiments 3-6.

Phonetic characteristic	HF	LF
Manner of articulation		
Combinations with the same manner of articulation	2411 (.264)	2541 (.279)
CC combinations requiring a change in manner of articulation	6709 (.736)	6579 (.721)
Place of articulation		
CC combinations with the same place of articulation	2965 (.325)	2906 (.319)
CC combinations with a different place but assimilation likely	1198 (.131)	815 (.089)
CC combinations requiring a change of place of articulation		
Minor place change (e.g. between 2 anterior Cs)	2134 (.234)	2822 (.309)
Major place change (e.g. between anterior/posterior Cs)	2823 (.310)	2603 (.285)
Voicing		
CC combinations involving no change in voicing	4641 (.509)	4703 (.516)
CC combinations involving a change in voicing	4479 (.491)	4417 (.484)

Note. C = consonant. Proportions are in brackets.

Table 12.1 demonstrates that in this case, phonetic analysis does not readily indicate a difference in articulatory complexity between stimulus sets that corresponds to performance in serial recall. While more instances of assimilation, that is, readily coarticulated coda-onset consonants (e.g. *pho*<u>*n*</u>*e* → <u>*b*</u>*all*), appear likely for HF than LF words, this is offset by a greater proportion of word boundaries requiring major place changes in articulation in the HF set.

12.4.2.1.2 Transitional probabilities

As a further check on the differences in coda-onset coarticulation between words, an analysis using the transitional probabilities of biphone frequency in language was conducted. Transitional probabilities are measures of the likelihood that two phonemes occur in order, as indexed by their frequency in language. Transitional probabilities of phonemes spanning word boundaries are typically much lower than those for consecutive phonemes within words, and it is thought that identification of word boundaries in speech makes use of these statistical differences (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Toro, Nespor, Mehler, & Bonatti, 2008). If frequency is inversely related to difficulty, transitional probabilities could be used as a metric to reflect expected differences in coarticulation fluency. As Woodward et al. (2008) argue that serial recall performance is a function of the articulatory fluency between words, it would be expected that differences in the transitional probabilities between HF and LF words should exist.

A biphone frequency database (Frankish, unpublished) derived from Celex information was used to determine the transitional probabilities of all CC coda-onset combinations in language. The coda-onset transitions for all possible combinations of the HF and LF words were then derived (9120 transitions for each set) and the associated transitional probabilities were identified (range 0 - .227). As these distributions were positively skewed the data was subject to a square root transformation. The resultant difference in mean values between HF and LF sets (M = .002) was not significant, $t(18238) = 1.62$, $p = .105$. Accordingly, if the legitimacy of the transitional probability as a proxy for expected word boundary fluency is accepted, it would appear that variations in coarticulation between HF and LF words in the current experiments were minimal. Therefore, as indicated by both phonetic analysis and transitional probability, coarticulation is unlikely to have contributed substantially to the differences observed in serial recall across the current experiments.

12.4.3 Serial recall with speech rate as a covariate

A within-subjects ANCOVA was run with all 4 list type conditions included using the speech rate values, determined from the timed word pair repetitions (see section 12.3.3), as covariates. This analysis replicated the results determined with the uncontrolled analysis, namely significant effects of list type and serial position, $F(3, 104) = 26.07$, $p < .001$, $MSE = .02$, and $F(5, 174) = 109.63$, $p < .001$, $MSE = .04$, respectively, and a significant interaction $F(15, 524) = 1.74$, $p = .040$, $MSE = .01$.

ANCOVA sub-analyses of the pure and alternating lists were then performed. In the case of pure lists, the nature of results was identical to those found when speech rate was not controlled for: list type and position were both significant, $F(1, 34) = 63.56$, $p < .001$, $MSE = .03$, and $F(5, 175) = 99.08$, $p < .001$, $MSE = .02$, while the interaction between these variables was non-significant, $F(5, 175) = 1.09$, $p = .368$, $MSE = .01$. The results from the alternating lists' data also were not altered by the addition of the covariate, with list type non-significant, $F(1, 34) = 2.70$, $p = .109$, $MSE = .02$, serial position significant, $F(5, 174) = 79.58$, $p < .001$, $MSE = .03$, but a significant interaction - $F(5, 174) = 3.09$, $p = .011$, $MSE = .01$. The list type by serial position interaction was explored further by running the contrast on differences between covariate-adjusted data of the alternating list types for odd and even serial positions. This analysis yielded a result similar to those found for the unadjusted data, $F(1,35) = 10.95$, $p = .002$, $MSE = .044$.

Limiting the ANCOVA analysis to the first 5 positions resulted in the effect of list type becoming marginally significant, $F(1, 35) = 3.68$, $p = .063$, $MSE = .02$, while the significance of the result for serial position was unaltered, $F(4, 139) = 104.98$, $p < .001$, $MSE = .03$. Similar to the uncontrolled analyses, the interaction between list type and position considering only the first 5 positions was non-significant, $F(4, 139) = 1.03$, $p = .394$, $MSE = .01$. Finally, the contrast comparing the differences in recall between odd and even positions for all positions other than recency produced a non-significant result, $F(1,35) = 1.85$, $p = .182$, $MSE = .046$, underscoring the importance of the final position to the presence of this effect. Thus, any adjustments to data according to speech rate differences were immaterial to the pattern of results observed for this experiment with the correct serial recall data.

12.4.4 Item analysis

Error analysis was performed on the data to examine more closely the influences of word frequency and list composition on serial recall. The responses for each trial were classified as either correct (the correct item in the correct serial position), an order error (the item recalled corresponded to a word that was presented elsewhere in the trial) or an item error (the response did not match any item presented in the trial, or was a repetition of a previously recalled list item). Order errors were broken down into either transpositions (where items from the same frequency type switched positions) or intra-list intrusions (where an item of one frequency type intruded onto the position of the other type - ILI). This latter measure was collected to see whether there existed any tendency for one frequency type to replace the other within a list. Non-repetition item errors were classified using the system of Allen and Hulme (2006), where errors are identified as either; an omission (when participants said ‘pass’ to an item, or indicated that they could not recall an item in that position); an intra-set intrusion (when an item from within the current experimental set but outside the list presented was recalled - ISI); an intra-experiment intrusion (when the response was from one of the other experimental sets - IEI), or an extra-experiment intrusion (when the response did not correspond to any item within the any of the experimental sets - EEI) . An additional classification ‘Other’ captured errors that were repetitions or, on rare occasions, nonwords). The proportion of errors of each classification, for each item type in each list type, collapsed across serial position and participants is given in Table 12.2, together with the proportion of items correct.

Transposition curves were generated for each of the conditions and are shown in Figure 12.2. While Hulme et al. (2003) reported that the transposition curves for their Experiment 2 revealed a weak tendency for items of the same frequency type to swap positions in the alternating conditions this pattern is not evident for Experiment 3. In addition, the breakdown of order errors (transpositions and ILI) for HF and LF items presented in alternating lists suggests that the movement of items within these lists was similar. Total order errors were conditionalised by dividing the total number of order errors by the number of items correctly recalled regardless of order. This was done to prevent the confounding of order errors with the level of item memory observed for each stimulus-condition type (Murdock, 1976; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999). Conditionalised order error rates were .123 for HF in pure lists, .131 for HF in mixed lists, .142 for LF in pure lists and .134 for LF in mixed lists, respectively. A 2 x 2 (frequency x list format) repeated measures ANOVA identified that neither frequency, $F(1,35) = 2.54$, $p = .120$, $MSE = .002$, nor list format, $F(1,35) = 0.00$, $p = .999$, $MSE = .003$, nor the frequency by list format interaction $F(1,35) = 1.15$, $p = .291$, $MSE = .002$, affected memory for order.

Table 12.2

Mean proportion of items correctly recalled and mean proportions of different error categories as a function of item type and list type in Experiment 3

	Correct	Errors			
		Order			Item
		Total	ILI	Trans	
HF in pure lists	.655 (.137)	.087 (.058)	-	.087 (.058)	.257 (.112)
LF in pure lists	.492 (.155)	.073 (.041)	-	.073 (.041)	.436 (.133)
HF in alt. lists	.583 (.142)	.082 (.043)	.068 (.039)	.015 (.014)	.332 (.128)
LF in alt. lists	.560 (.146)	.082 (.043)	.065 (.034)	.017 (.017)	.360 (.120)

	Item errors by category				
	Omissions	EEI	ISI	IEI	Other
HF in pure lists	.174 (.103)	.043 (.026)	.023 (.018)	.011 (.010)	.006 (.009)
LF in pure lists	.284 (.128)	.085 (.051)	.029 (.013)	.027 (.017)	.011 (.015)
HF in alt. lists	.218 (.121)	.059 (.032)	.029 (.022)	.016 (.015)	.008 (.009)
LF in alt. lists	.229 (.116)	.074 (.032)	.023 (.014)	.027 (.018)	.010 (.011)

Note. ILI – Intra-list intrusions, Trans – transposition to a same frequency position in the list, EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Other – repetitions and nonwords. Standard deviations are given in brackets.

A 2 x 2 repeated measures ANOVA was performed on the total item errors. This analysis revealed a significant main effect of frequency, $F(1,35) = 166.42$, $p < .001$, $MSE = .002$, but the effect of list format was non-significant, $F(1,35) = 0.00$, $p = .986$, $MSE = .002$, while the interaction was significant, $F(1,35) = 61.60$, $p < .001$, $MSE = .003$. High frequency words produced less item errors than LF words however this difference was attenuated in mixed lists where error levels were intermediate with respect to those for pure lists.

Separate categories of item error large enough to produce meaningful results were then explored to determine whether they revealed differing sensitivities to item type or context of presentation. The pattern of results for omissions reflected that for total items - a significant main effect of frequency, $F(1,35) = 71.91$, $p < .001$, $MSE = .079$, a non-significant effect of list format, $F(1,35) = 0.70$, $p = .408$, $MSE = .059$, and a significant interaction $F(1,35) = 29.80$, $p < .001$, $MSE = .082$. Again, these effects were driven by a sizeable difference in the number of omissions between HF and LF words in pure lists, with HF lists producing less omissions than LF lists, but this difference was abolished, and the error level observed was intermediate, when words appeared in alternating lists.

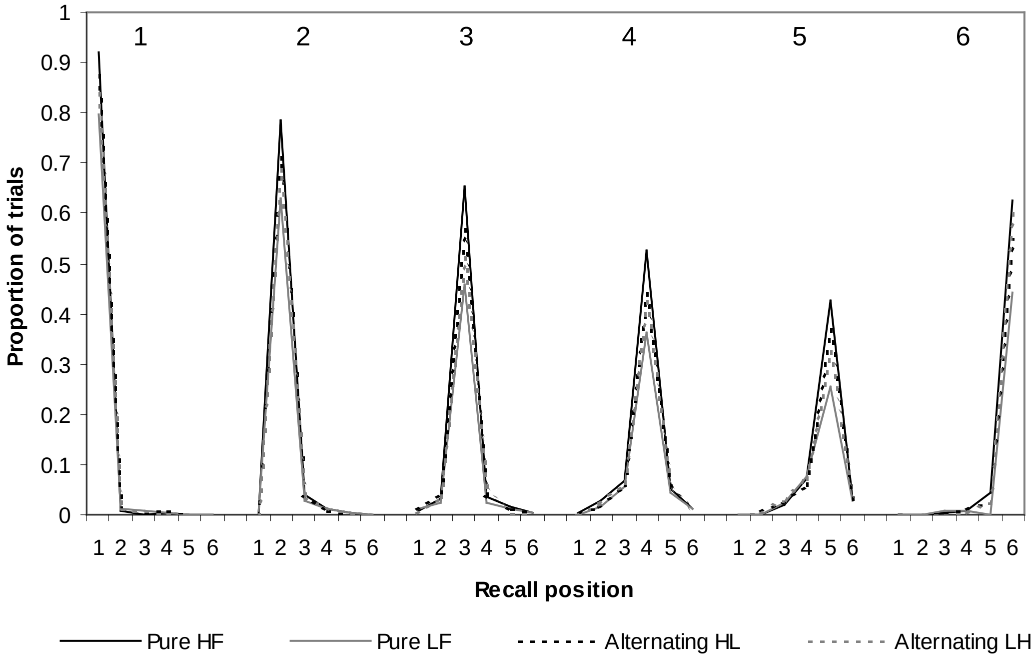


Figure 12.2. Transposition gradients for serial recall as a function of word frequency and list type in Experiment 3.

The results for the number of intrusions from outside all experimental sets (EEI) followed the pattern observed for omissions but to a lesser degree. The number of EEI was influenced by frequency, $F(1,35) = 22.51$, $p < .001$, $MSE = .035$, but not list format, $F(1,35) = 0.24$, $p = .627$, $MSE = .033$, while the interaction was significant, $F(1,35) = 17.21$, $p < .001$, $MSE = .024$. More intrusions occurred with LF words in either list context, and list type served to moderate error levels by decreasing EEI in alternating lists for LF words, but increasing the level in mixed lists for HF words.

The remaining error types together contributed between 4-6% of the total errors for each frequency by list type category. Low frequency words appeared more likely to be replaced by non-list HF set members than vice versa. There was no obvious systematic variation in the ISI errors with the rates for LF items in pure lists and HF items in alternating lists slightly elevated with respect to the levels of the other item categories.

12.5 Discussion

Experiment 3 was performed as a replication of Hulme et al.’s (2003) Experiment 2 to confirm that new sets of open stimuli could reproduce the mixed list paradox for word frequency in serial recall. Therefore it was anticipated that the frequency effect for pure lists would be abolished for alternating lists. Importantly, speech rate differences were not found to alter the salient features of the results for Experiment 3.

Experiment 3 found a frequency effect for pure lists that was similar to the effect reported on several occasions; HF words were consistently better recalled than LF words across serial positions (Allen & Hulme, 2006; Hulme et al., 1997; Hulme et al., 2003; Poirier & Saint-Aubin, 1996; Roodenrys & Quinlan, 2000). However, in contrast with the majority of experiments (e.g. Hulme et al., 1997; Hulme et al., 2003), a significant frequency by serial position interaction was not reported. As discussed in section 11.5.2.1, Hulme et al. (1997) identified that the size of effect became larger across serial positions, except for the final position, and proposed the increase in effect was consistent with a redintegrative process that was more successful for HF than LF words as short-term traces became more degraded. Nonetheless, a closer inspection of the recall pattern for pure lists in Hulme et al.’s (2003) Experiment 2 identified that the interaction in this case was most likely to be driven by a ceiling effect operating on the first serial position; the size of effect in positions 2-5 did not vary much, and in most respects appeared similar to the recall patterns for pure lists reported here. It is likely therefore, that the frequency by serial position interaction, thought to be a standard feature of the frequency effect with pure lists, is less well generalised than otherwise assumed.

The frequency effect for alternating lists in Experiment 3 produced a number of similarities with the results of Hulme et al.'s (2003) Experiment 2. The current experiment found that when all serial positions were taken into account, there was no frequency effect for alternating lists, however after investigation of the accompanying interaction and removal of the last position from analysis, a small but marginal frequency effect in the form of better recall for alternating lists beginning with HF words was apparent. This result is consistent with the slight advantage for these lists found by Hulme et al. (2003) when open sets were used.

The current experiment identified a possible item-specific contribution to recall in the form of a sawtooth component superimposed on the recall curves for mixed lists. This effect required the inclusion of the final serial position in order to be reliable nonetheless it is consistent with other research that suggests that both item-specific and associative influences are present in the frequency effect (Tse & Altarriba, 2007).

Experiment 3 did not find evidence of any frequency related order effects. This observation is in line with the majority of the serial recall literature that has reported null effects of frequency on order accuracy and reinforces the position that frequency impacts the level of item memory in the recall of a list of words (Allen & Hulme, 2006; Hulme et al., 2003; Morin et al., 2006; Poirier & Saint-Aubin, 1996).

The analysis of item errors revealed a pattern consistent with the change in frequency effect found in the serial recall of pure and alternating lists. As per Hulme et al.'s (2003) report, LF items produced substantially more errors than HF words in pure lists, while errors for LF items in alternating lists occurred at a similar rate to errors for HF items. Furthermore, the error rates for mixed lists were intermediate with respect to the levels found for HF and LF words in pure lists. The largest category of item errors, omissions, followed this pattern. The number of EEI loosely conformed to this outline also, where the proportion of intrusions for LF items was less for alternating than pure lists, while these errors increased for HF items in alternating lists when compared to pure lists.

Therefore, the results of this experiment were generally consistent with the explanation provided by Hulme et al. (2003; Stuart & Hulme, 2000) positing inter-item associative mechanisms at recall. Pure HF lists would be most advantaged in this respect, as HF words are most likely to co-occur in language with other HF words, and accordingly their inter-item association would be high. Pure LF lists would produce the least supportive circumstance due to the relative rarity of co-occurrence between LF words. Finally, if due to the high usage of HF words, associations between HF and LF words are, on average, stronger than associations between LF words but weaker than associations between HF words, items in alternating lists of HF and LF words would be expected to be recalled at an intermediate level to pure HF and pure LF lists. Furthermore, alternating lists commencing with either HF or LF words should be recalled at equivalent levels.

12.3 Conclusion

Experiment 3 demonstrated that the recall of HF and LF words in pure and alternating lists of CVC words is similar to the pattern reported by Hulme et al. (2003, Experiment 2). In consequence, the generalisability of the frequency effect in the recall of pure and alternating lists with open sets of stimuli was demonstrated, and a point of continuity with the existing literature was established.

Given that Experiment 3 reproduced the important features of the preceding research, a test directly comparing the recall of alternating lists with lists where each half is composed of items of one frequency type (i.e. HHHLLL or LLLHHH) could be considered. The results from such a study will provide further information on the nature of associative effects between list items. This experiment is the subject of Chapter 13.

Chapter 13

Experiment 4: A Comparison of the Serial Recall of Alternating and Half Lists

13.1 Introduction

Experiment 4 compared the recall patterns for alternating and half lists, the latter constructed so that items of the same frequency type occurred consecutively in the first or the second half of the list. Therefore in this experiment all lists had the same composition (3 HF items and 3 LF items), but item arrangement was varied to determine whether serial recall is a function of the strength of association between consecutive list items.

If the nature of effect of inter-item associations during redintegration was non-directional and nonspecific as Hulme et al. (2003; Stuart & Hulme, 2000) proposed, then there should be no reliable differences in the frequency effect for alternating and half lists in any list position. If however this redintegrative contribution was directionally sensitive (Howard & Kahana, 2002b), then different patterns of recall would be anticipated for alternating and half lists. More specifically, according to Howard and Kahana (2002b), in free recall strongly associated items are better recalled in sequence than weakly associated items. Hulme et al. (2003) suggested that if a similar mechanism was to operate within late-stage redintegration then similar directional sensitivities would be observed; sequences of HF words would achieve better recall performance than sequences of LF words in the same serial positions.

In addition, based on the observations of alternating list performance in Experiment 3, it was predicted that the recall of HF words in alternating lists would be recalled at a similar or slightly greater level than LF words in those lists. Furthermore, a subtle see-sawing pattern with list type might be present in alternating lists, suggesting the superposition of an item-specific contribution onto the list-level component of word frequency (Tse & Altarriba, 2007).

13.2 Method

13.2.1 Participants

The participants for Experiment 4 were 39 undergraduate University of Wollongong students (29 females, 10 males) who participated for course credit. The data from 3 participants were excluded from analysis because they were either not native Australian-English speakers (1) or their performance showed floor effects. The mean age of the 36 participants whose data was retained for analysis was 23.9 years (*SD* = 8.4 years).

13.2.2 Materials

The stimuli used in Experiment 3 were used in Experiment 4. This experiment arranged items according to alternating and half list formats. Alternating lists were as defined for Experiment 3, while half lists comprised either three HF items followed by three LF items (HHHLLL), or vice versa (LLLHHH). The procedure followed to generate list sets for experimental sessions was identical to that for Experiment 3.

13.2.3 Procedure

The procedure for administering the experiment was the same as for Experiment 3.

13.3 Results

The scoring methods used in Experiment 3 were applied to this data. The mean number of correctly recalled items by serial position and condition is shown in Figure 13.1.

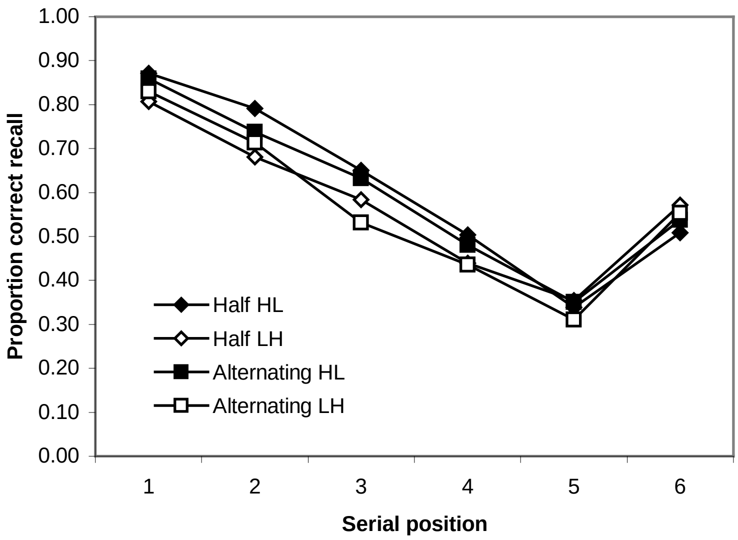


Figure 13.1. Serial recall of words as a function of frequency and list composition in Experiment 4.

13.3.1 Serial recall

As before, an alpha level of .05 was applied to all statistical tests. An analogous set of analyses to those performed in Experiment 3 was run on the data. The 4 x 6 (list type x serial position) repeated measures ANOVA identified that list type, $F(3,105) = 5.85$, $p = .001$, $MSE = .019$, serial position, $F(5,175) = 116.81$, $p < .001$, $MSE = .064$, and the list type by position interaction, $F(15,525) = 3.37$, $p < .001$, $MSE = .017$, were significant. Half lists were better recalled than alternating lists. The interaction is likely to have been driven by a combination of crossover effects between list pairs, and possibly an irregularity in recall at the third serial position for the LH alternating list type. Repeated contrasts identified that lists beginning with HF items were better recalled than lists beginning with LF items; the difference between HL half and HL alternating lists was non-significant, $F(1,35) = 0.73$, $p = .399$, $MSE = .026$, the difference between the alternating lists was significant, $F(1,35) = 8.80$, $p = .005$, $MSE = .301$,while the difference between LH alternating and LH half lists was non-significant, $F(1,35) = 0.53$, $p = .471$, $MSE = .022$.

Separating the list type by format (half versus alternating) explored the interaction further and permitted a closer examination of list type differences. In the case of half lists all results were significant, namely frequency, $F(1,35) = 9.65$, $p = .004$, $MSE = .016$, serial position, $F(5,175) = 91.49$, $p < .001$, $MSE = .037$, and the list type by position interaction, $F(5,175) = 6.89$, $p < .001$, $MSE = .011$. These lists crossed over at approximately position five, one position later than the change in item frequency.

The alternating lists also revealed significant results for list type, $F(1, 35) = 8.80$, $p = .005$, $MSE = .017$, serial position, $F(5, 175) = 93.81$, $p < .001$, $MSE = .035$, and a significant interaction, $F(5, 175) = 2.40$, $p = .039$, $MSE = .011$. Unlike Experiment 3, when all serial positions including the recency position were considered, lists starting with an HF item were better recalled than those that began with an LF item, despite performances converging at the recency position. As in the case of Experiment 3, the significance of the interaction rested on the final position; when the first 5 serial positions were re-analysed, frequency, $F(1, 35) = 16.40$, $p < .001$, $MSE = .013$, and serial position, $F(4, 140) = 127.14$, $p < .001$, $MSE = .032$, remained significant effects, but the interaction became non-significant, $F(4, 140) = 1.46$, $p = .218$, $MSE = .012$.

Experiment 3 found a significant item-specific effect, evidenced by a small see-saw in the serial recall of alternating lists. Figure 13.1 indicates that such a see-saw pattern is not readily apparent in the current data. Nevertheless, the differences in alternating serial recall positions for these lists in Experiment 4 were subjected to the same contrast. However, in this instance, the contrast was found to be non-significant, $F(1,35) = 2.68$, $p = .111$, $MSE = .029$. The equivalently weighted contrast across the first five serial positions was also non-significant, $F(1,35) = 0.67$, $p = .419$, $MSE = .029$.

Once more a small but reliable difference between the speech rates between HF (mean 3.10 words per second) and LF items was recorded (mean 2.94 words per second; $t(35) = 5.25$, $p < .001$). A parallel series of ANCOVAs were performed to assess the level to which differences in speech rate according to frequency type altered the nature of results. An ANCOVA including all 4 list types replicated the pattern of the unadjusted data; significant effects of list type, $F(3, 105) = 5.85$, $p = .001$, $MSE = .02$, and position, $F(5, 175) = 116.80$, $p < .001$, $MSE = .04$, and a significant interaction, $F(15, 524) = 2.48$, $p = .002$, $MSE = .01$ were found.

The significances of half list results did not change when speech rate data was included in the analysis; list type, $F(1, 35) = 9.65$, $p = .004$, $MSE = .02$, serial position, $F(5, 175) = 91.49$, $p < .001$, $MSE = .03$, and the interaction, $F(5, 174) = 4.66$, $p = .001$, $MSE = .01$, all remained significant. Furthermore, a simple effects analysis adopting a familywise error rate of .05 and using adjusted data found that differences in recall between list types were significant for positions one, two and four. It is noteworthy that recall at the point of change in list composition remained significantly better for lists beginning with HF words; for this position LF words were consistently better recalled than HF words. Beyond this point there were no differences in performance between list types.

While half lists did not exhibit any changes when the covariate was considered, alternating lists were altered by speech rate adjustments. The effects of list type, $F(1, 35) = 8.80$, $p = .005$, $MSE = .02$, and position, $F(5, 175) = 93.81$, $p < .001$, $MSE = .03$, remained significant, however the interaction $F(5, 174) = 1.77$, $p = .121$, $MSE = .01$, was non-significant. The frequency effect between list types was reinforced by a single significant difference at position three, as identified by simple effects on the adjusted data. Unsurprisingly, covariate analysis of the first 5 serial positions did not change the status of results established with the corresponding unadjusted data; this revealed significant effects of list type, $F(1, 34) = 10.82$, $p = .002$, $MSE = .01$, and position, $F(4, 140) = 127.14$, $p < .001$, $MSE = .02$, and a non-significant interaction, $F(4, 139) = 1.24$, $p = .300$, $MSE = .01$. Taken together, these last analyses reveal the possibility that the convergence of alternating lists in the final position is a result of item-specific influences that operate there, removed to some extent by controlling for speech rate differences. Additionally, this explanation could be extended to the crossover in the levels of recall at the recency position observed between alternating lists in Experiment 3.

13.3.2 Item analysis

Table 13.1 presents the item and order error rates for this experiment according to the same classification system as used in Experiment 3, and the proportion of items correctly recalled. The latter can be seen to be similar for HF and LF words respectively, regardless of list context. The total order error rates for each frequency type do not appear to be affected by list format either, however the breakdown by order error type does differ between half and alternating lists, and reflects the tendency to recall items in positions adjacent to the position of presentation (see Figure 13.2).

Conditionalised order error rates were .110 for HF in half lists, .117 for HF in alternating lists, .097 for LF in half lists and .093 for LF in alternating lists, respectively. A 2 x 2 (frequency x list format) repeated measures ANOVA identified that the effect of frequency was significant, $F(1,35) = 11.06$, $p = .002$, $MSE = .001$, but list format, $F(1,35) = 0.02$, $p = .888$, $MSE = .002$, and the interaction $F(1,35) = 0.68$, $p = .415$, $MSE = .001$, were both non-significant. On the basis of the total number of items recalled, LF items were the subject of fewer order errors than HF words. The propensity to recall items in positions adjacent to the presented position was also reinforced by the breakdown of ILI and transposition errors between formats; a greater proportion of ILI was produced in the recall of alternating lists.

Table 13.1

Mean proportion of items correctly recalled and mean proportions of different error categories as a function of item type and list type in Experiment 4

	Correct	Errors			
		Order			Item
		Total	ILI	Trans	
HF in half lists	.611 (.118)	.071 (.044)	.026 (.022)	.045 (.028)	.319 (.095)
LF in half lists	.573 (.127)	.057 (.036)	.022 (.018)	.034 (.026)	.368 (.100)
HF in alt. lists	.600 (.137)	.070 (.044)	.053 (.033)	.017 (.018)	.347 (.124)
LF in alt. lists	.563 (.139)	.054 (.030)	.044 (.025)	.010 (.011)	.365 (.105)
	Item errors by category				
	Omissions	EEI	ISI	IEI	Other
HF in half lists	.213 (.093)	.056 (.036)	.029 (.017)	.014 (.013)	.008 (.011)
LF in half lists	.218 (.101)	.085 (.046)	.025 (.020)	.031 (.024)	.009 (.010)
HF in alt. lists	.238 (.113)	.050 (.039)	.029 (.022)	.022 (.018)	.008 (.007)
LF in alt. lists	.231 (.109)	.072 (.035)	.027 (.018)	.027 (.021)	.009 (.011)

Note. ILI – Intra-list intrusions, Trans – transposition to a same frequency position in the list, EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Other – repetitions and nonwords. Standard deviations are given in brackets.

The total item errors were examined using a 2 x 2 (frequency by list format) repeated measures ANOVA. This analysis revealed a significant main effect of frequency, $F(1,35) = 15.79$, $p < .001$, $MSE = .003$, but the effect of list format was non-significant, $F(1,35) = 2.02$, $p = .164$, $MSE = .003$. However, the interaction was marginally significant, $F(1,35) = 3.66$, $p = .064$, $MSE = .002$. High frequency words produced less item errors than LF words, and this pattern tended to be greater for half lists.

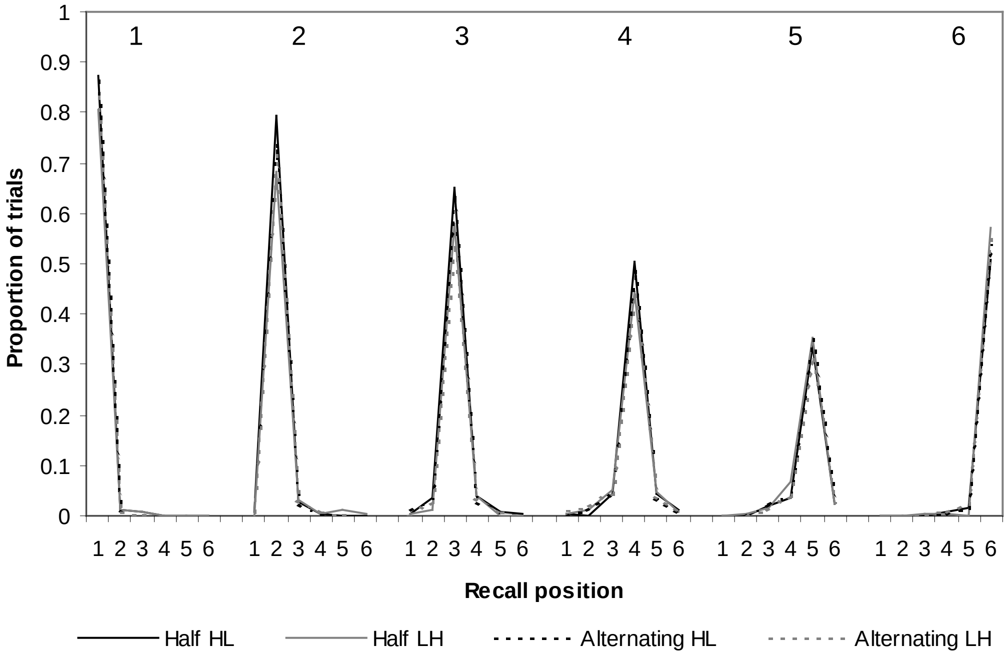


Figure 13.2. Transposition gradients for serial recall as a function of word frequency and list type in Experiment 4.

Differences in omissions were influenced by the arrangement of items in the list rather than their frequency, with analysis identifying a non-significant effect of frequency, $F(1,35) = 0.01$, $p = .921$, $MSE = .002$, but a significant effect of list format, $F(1,35) = 5.55$, $p = .024$, $MSE = .002$, and a non-significant interaction $F(1,35) = 1.14$, $p = .293$, $MSE = .001$. Specifically, half lists suffered less omission errors than alternating lists. Furthermore the finding of no effect of frequency concurs with the absence of a frequency effect for alternating lists reported in the analysis of the Experiment 3 data. It would appear that the effect of frequency for omissions is a result of list composition (pure versus mixed) and when list composition is matched on item type, variation in this error classification is influenced by the arrangement of items in the list, i.e. half lists show an effect of frequency while alternating lists do not.

The number of intrusions from outside all experimental sets, EEI, was influenced by both frequency, $F(1,35) = 38.52$, $p < .001$, $MSE = .001$, and list format, $F(1,35) = 4.92$, $p = .033$, $MSE = .001$, but the interaction was non-significant, $F(1,35) = 0.65$, $p = .426$, $MSE = .001$. As with Experiment 3, more intrusions of this kind occurred with LF words, but for these items more intrusions occurred with half lists than alternating ones.

The remaining error classifications totalled between 5-7% of the frequency by list type categories and were not analysed further. The ISI rates appear somewhat constant across categories, while the tendency for LF items to be replaced by HF items from outside the list (IEI) was again evident.

13.4 Discussion

Like Experiment 3, the results of the analyses of the raw serial recall data and the data adjusted to account for speech rate differences for the current experiment were highly comparable. A directional contribution was evident in the pattern of serial recall for half lists, as the frequency effect for the first two serial positions was significant, and a crossover of the serial recall curves was present. However this occurred after the change in list composition, at approximately the fifth serial position. The difference between half lists was found to be significant at the fourth serial position, but recall of half-lists was not significantly different for the last two positions. Furthermore, in the fourth serial position the recall of LF words was greater than the recall of HF words, continuing the better recall of half lists beginning with HF items than lists beginning with LF words to this point. Therefore, an inter-item associative effect operating on consecutive items was found to lag behind transitions between HF and LF words. On reflection, this would seem a reasonable outcome, if the recall of an item is in some way dependent on the recall of the previous item.

The pattern of results for the serial recall of half lists appears to lend support to the findings of M.J. Watkins (1977) who found better memory span for mixed lists beginning with HF than LF list halves. However, strategic factors encouraged by the task demands of the present experiment, and not present in span tasks, may also have contributed to the serial recall curves obtained in Experiment 4. Accordingly, support for M.J. Watkins' (1977) finding cannot be viewed as unqualified.

More generally, lists beginning with HF rather than LF items were better recalled, and the recall of half and alternating lists beginning with items of the same frequency type did not differ. The first of these findings is consistent with the marginal effect found between alternating lists in Experiment 3 for serial positions 1-5. That this superiority occurs with alternating lists, although small, suggests there is an advantage to commencing list recall with a highly used item that, additionally, is likely to be better associated with the next item in the sequence. It is possible that start-of-list relies more on the item-specific properties of an item, and that this advantage can be perpetuated with later list items through associative means.

However, the serial recall curves for alternating lists in Experiment 4 were not as well-formed as in Experiment 3, with an anomaly in serial recall performance inverting the anticipated levels of recall for lists beginning with LF words in the third serial position. Performance at this serial position no doubt contributed to the reliable difference between alternating lists across all serial positions. Concerns regarding the form of these curves aside, the non-significance of the interaction when speech rate was extracted from the recall data does suggest that recall in the recency position might contain a greater item-specific component than the recall for pre-recency serial positions.

Order errors were found to be sensitive to item type, but not the arrangement of items within the list. Furthermore HF items were observed to be slightly, but reliably more likely than LF words to be recalled in the wrong position. The history regarding order errors for mixed lists is not extensive, but Roodenrys et al. (2002) also found that HF items were more likely to be recalled in the wrong list position when mixed with LF items, and Hulme et al. (2003) reported weak effects for LF words in their Experiment 2. The current item analysis cannot provide any detail as to specific list positions that might be more susceptible to item reordering.

In Experiment 4, as in Experiment 3, the vast bulk of errors were item errors and these were found to be more prevalent for LF words. In this experiment, item errors varied marginally with the arrangement of list items suggesting that the frequency effect reported here possibly contained an item-specific component. However omissions did not contribute to this difference as the proportions of omissions were not different for HF and LF words. In contrast, more omissions were found for alternating lists than half lists demonstrating that the largest error category was influenced by item arrangement. This suggests that the structure of half lists produces a small advantage in the retention of item information that is available for HF and LF items. The numbers of EEI varied by both frequency and list format, indicating that at least some element of this error was due to item arrangement. More EEI were produced for LF than HF items, and more EEI occurred for half lists than alternating lists. It is possible that the better item retention for half lists, as evidenced by the lower rate of omissions, also translated to better retention of partial item information that in turn, created more opportunities for intrusions into the list (e.g. Roodenrys et al., 2002).

Despite the apparent presence of small item-specific contributions in the item error data, the item-specific effect witnessed in the serial recall of alternating lists in Experiment 3 was not replicated in this experiment. The effect was not present when either all serial positions or the first 5 serial positions were considered. Therefore, it is likely that that this element of recall is transitory in nature and not a reliable feature of alternating lists. However, the irregularities in the serial recall curves from this experiment suggest also that this data might not be the best from which to generalise.

13.5 Conclusion

Experiment 4 was designed to examine whether two forms of mixed lists, half lists and alternating lists, produced patterns of recall consistent with non-directional long-term associative support of all list items (Hulme et al., 2003; Stuart & Hulme, 2000) or directionally sensitive associativity between adjacent list items reflecting the strength of semantic association. Some evidence was found regarding a directional component to the inter-item associations of list members; differences between half lists in early serial positions and the cross-over of the serial recall curves of these lists imply some effect of relationship between consecutive list members. However, based on the full data from this experiment, the story for the frequency effect might not be a simple one, as effects associated with item type, but unrelated to list format, suggest that item-specific properties could also be influential in the production of item errors. Despite the concordance of the item error data for alternating lists between Experiment 3 and Experiment 4, the serial recall patterns observed in this experiment contain some anomalies that may have influenced the pattern of errors obtained in this case.

Having examined the recall of half lists against the recall of alternating lists, it was decided to compare the recall of half lists with pure lists. As an equivalent comparison had already been performed with alternating and pure lists (Experiment 3, Hulme et al., 2003), an experiment producing the corresponding information with half lists would offer complementary information to that found in Experiment 4, and possibly offer a clearer picture of the serial recall behaviour of half lists.

Chapter 14

Experiment 5: A Comparison of the Serial Recall of Pure and Half Lists

14.1 Introduction

The previous experiment had produced serial recall curves for half lists suggesting that at least some component of the frequency effect responds to the degree of associativity between adjacent list items. However, the data from this experiment was not as decisive as anticipated, as these were indications that the frequency effect, at least in the context of Experiment 4, could be a product of a number of factors. While differences between half lists existed for early serial positions, they did not for the latter lists positions, and the changeover in performance between half lists beginning with HF and LF words lagged behind the change in list composition. This could be explained by the conditional nature of directionally sensitive associativity (that is, successful recall of an item at serial position i is in part a function of successful recall at position i-1), however the lack of distinction between half and alternating lists in late serial positions suggests this proposal is, on the basis of the evidence so far, speculative.

Experiment 4 did not produce information that would indicate how closely recall performance on portions of half lists would mimic recall on pure lists across serial positions. The aim of Experiment 5 was to provide this contrast. Furthermore, this data would allow distinctions between the performances for half lists and alternating lists in relation to the recall of pure lists to be drawn.

14.2 Method

14.2.1 Participants

Forty undergraduate and postgraduate University of Wollongong students (27 females, 13 males) participated in this experiment for course credit. The data from 4 participants were excluded from analysis because they were not native Australian-

English speakers (1), had hearing difficulties (1), failed to follow the recall procedure (1), and had a speech impediment making responses difficult to perceive and evaluate. The mean age of the remaining 36 participants whose data was retained for analysis was 23.1 years (*SD* = 7.6 years).

14.2.2 Materials

The same stimuli that were used in Experiment 3 were employed in Experiment 5. This experiment arranged items according to pure and half list formats. Pure lists were as defined for Experiment 3, while half lists were as defined for Experiment 4 (HHHLLL or LLLHHH). The procedure generating list sets for experimental sessions was identical to that for Experiment 3.

14.2.3 Procedure

The procedure for administering the experiment was the same as for Experiment 3.

14.3 Results

The scoring methods used in Experiment 3 were applied to this data. The mean number of correctly recalled items by serial position and condition are shown in Figure 14.1.

14.3.1 Serial recall

The criterion for significance and statistical tests were the same as for Experiment 3. The 4 x 6 (list type x serial position) repeated measures ANOVA identified that list type, $F(3,105) = 62.13$, $p < .001$, $MSE = .021$, serial position, $F(5,175) = 120.84$, $p < .001$, $MSE = .086$, and the list type by position interaction, $F(5,525) = 6.20$, $p < .001$, $MSE = .016$, were significant. As in Experiment 4, lists starting with HF items generated better performance than those beginning with LF items. In a strict sense, the serial recall curves for half lists were observed to crossover at approximately the fifth serial position, while the pure lists showed the typical difference between HF and LF items across all serial positions. However in contrast to Experiment 4, the frequency effect at the fourth serial position was clearly influenced by the transition from HF to LF words, and vice versa, on recall.

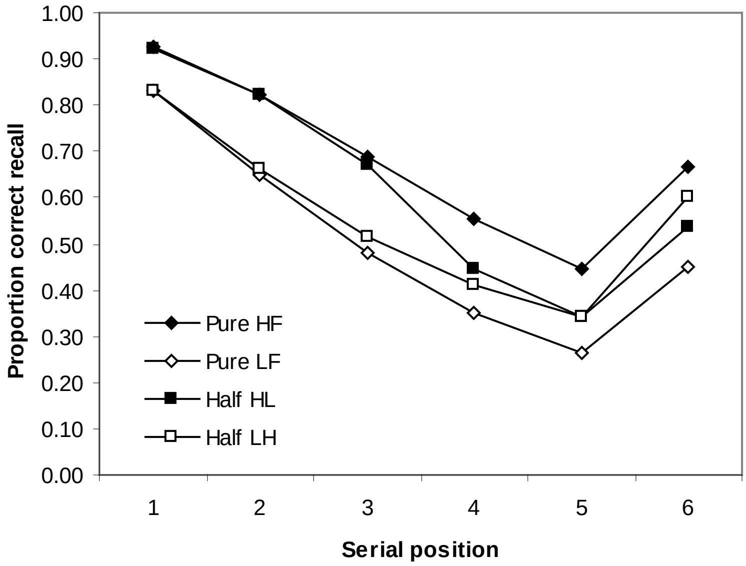


Figure 14.1. Serial recall of words as a function of frequency and list composition in Experiment 5.

The interaction was explored further by conducting individual analyses for each list format (pure or half). In the case of pure lists all results were significant, namely frequency, $F(1,35) = 123.76$, $p < .001$, $MSE = .028$, serial position, $F(5,175) = 86.08$, $p < .001$, $MSE = .056$, and the list type by position interaction, $F(5,175) = 3.56$, $p = .004$, $MSE = .010$. In this instance, the interaction was driven by a ceiling effect operating for HF items in the first serial position 1. A 2 x 5 ANOVA on the last five serial positions identified that while the main effects of list type, $F(1,35) = 127.08$, $p < .001$, $MSE = .028$, and serial position, $F(4,140) = 50.29$, $p < .001$, $MSE = .048$, remained significant, the interaction disappeared, $F(4, 140) = 0.53$, $p = .713$, $MSE = .012$.

The half lists also produced significant results for list type, $F(1, 35) = 29.85$, $p < .001$, $MSE = .014$, serial position, $F(5, 175) = 118.70$, $p < .001$, $MSE = .037$, and a significant interaction, $F(5, 175) = 13.33$, $p < .001$, $MSE = .017$. Lists beginning with HF words had better recall overall and as previously mentioned, performance on these lists crossed over at approximately the fifth serial position, one item after the change in list composition, though recall was clearly affected prior to this point.

Once more a higher speech rate for HF (mean 2.97 words per second) than LF items (mean 2.85 words per second; $t(35) = 3.56$, $p = .001$) was observed from the speech rate data. Accordingly, a parallel series of ANCOVAs were performed to assess the level to which differences in the speech rate of HF and LF items altered the nature of results. An ANCOVA including all 4 list types replicated the pattern of results seen in the unadjusted data; significant effects of list type, $F(3, 105) = 39.83$, $p < .001$, $MSE = .02$, and position, $F(5, 175) = 120.84$, $p < .001$, $MSE = .04$, and a significant interaction, $F(15, 524) = 4.19$, $p < .001$, $MSE = .01$ were found.

Examining pure lists with speech rate as a covariate revealed significant results for list type, $F(1, 34) = 75.10$, $p < .001$, $MSE = .03$, serial position, $F(5, 175) = 86.08$, $p < .001$, $MSE = .03$, and the interaction, $F(5, 175) = 3.56$, $p = .004$, $MSE = .01$. Consistent with the analysis on the unadjusted data, the list type by serial position interaction disappeared when the first serial position data was omitted from the ANCOVA, $F(4,140) = 0.53$, $p = .714$, $MSE = .01$.

The ANCOVA examining half lists produced significant effects of list type, $F(1, 35) = 29.85$, $p < .001$, $MSE = .01$, and position, $F(5, 175) = 118.70$, $p < .001$, $MSE = .02$, and the interaction $F(5, 174) = 7.73$, $p < .001$, $MSE = .01$. To explore the nature of the interaction further, a simple effects analysis on the adjusted means was performed with a familywise error rate of .05. This procedure identified that recall between list types in the first three serial positions only was significantly different.

14.3.2 Item analysis

In line with the error analyses presented for Experiments 3 and 4, the rates of the classifications of items errors, together with the proportion of correct recall, were determined for each frequency by list format combination, and are given in Table 14.1. In terms of correct recall a frequency effect was evident in both list formats, although it was attenuated in the case of half lists, due to the non-significant difference between lists for positions four, five and six.

Figure 14.2 gives the transposition curves for this experiment. All conditions conformed to the expected pattern of transposition errors. Conditionalised order errors identified that .100 of the HF items in pure lists were recalled in the wrong position, while LF items in pure lists were recalled out of position on .102 of occasions. The equivalent error values for HF and LF items in half lists respectively were .105 and .111. When submitted to a 2 x 2 (frequency x list format) repeated measures ANOVA, neither frequency, $F(1,35) = 0.32$, $p = .575$, $MSE = .002$, list type, $F(1,35) = 0.86$, $p = .360$, $MSE = .002$, nor the interaction $F(1,35) = 0.06$, $p = .808$, $MSE = .002$, were significant.

Table 14.1

Mean proportion of items correctly recalled and mean proportions of different error categories as a function of item type and list type in Experiment 5

	Correct	Errors			
		Order			Item
		Total	ILI	Trans	
HF in pure lists	.684 (.143)	.071 (.046)	-	.071 (.046)	.244 (.117)
LF in pure lists	.504 (.134)	.052 (.033)	-	.052 (.035)	.444 (.119)
HF in half lists	.623 (.126)	.071 (.043)	.024 (.019)	.047 (.032)	.303 (.110)
LF in half lists	.560 (.135)	.065 (.045)	.022 (.016)	.043 (.034)	.384 (.116)
	Item errors by category				
	Omissions	EEI	ISI	IEI	Other
HF in pure lists	.170 (.107)	.037 (.028)	.022 (.022)	.009 (.011)	.007 (.010)
LF in pure lists	.275 (.129)	.097 (.040)	.026 (.017)	.035 (.028)	.011 (.012)
HF in half lists	.212 (.114)	.048 (.032)	.026 (.018)	.010 (.011)	.008 (.012)
LF in half lists	.230 (.114)	.083 (.034)	.026 (.017)	.028 (.025)	.017 (.017)

Note. ILI – Intra-list intrusions, Trans – transposition to a same frequency position in the list, EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Other – repetitions and nonwords. Standard deviations are given in brackets.

The corresponding analysis for total item errors identified that there was a significant main effect of frequency, $F(1,35) = 276.54$, $p < .001$, $MSE = .003$, no effect of list format, $F(1,35) = 0.01$, $p = .931$, $MSE = .004$, but a significant interaction, $F(1,35) = 49.61$, $p < .001$, $MSE = .003$. High frequency words produced less item errors than LF words, but this difference was smaller for half lists.

Differences in omissions replicated those for the total item errors, namely a significant effect of frequency, $F(1,35) = 77.26$, $p < .001$, $MSE = .002$, a non-significant effect of list format, $F(1,35) = 0.07$, $p = .793$, $MSE = .001$, *ns*, and a significant interaction $F(1,35) = 29.91$, $p < .001$, $MSE = .002$. Low frequency words produced more omissions than HF words in pure lists, however this difference was negligible for half lists.

The number of intrusions from outside all experimental sets, EEI, followed the same pattern with a significant effect of frequency, $F(1,35) = 113.28$, $p < .001$, $MSE = .001$, but not list format, $F(1,35) = 0.11$, $p = .742$, $MSE = .001$, while the interaction was significant, $F(1,35) = 11.13$, $p = .002$, $MSE = .001$. Low frequency items were often replaced by words from outside both stimulus sets and this effect was greater in pure lists.

The residual classifications amounted to between 4-7% of all items for each category examined. From Table 14.1 it can be noted that the level of ISI across categories was invariant, while LF items were apparently more likely to suffer intrusion from a HF word not presented in the list (IEI).

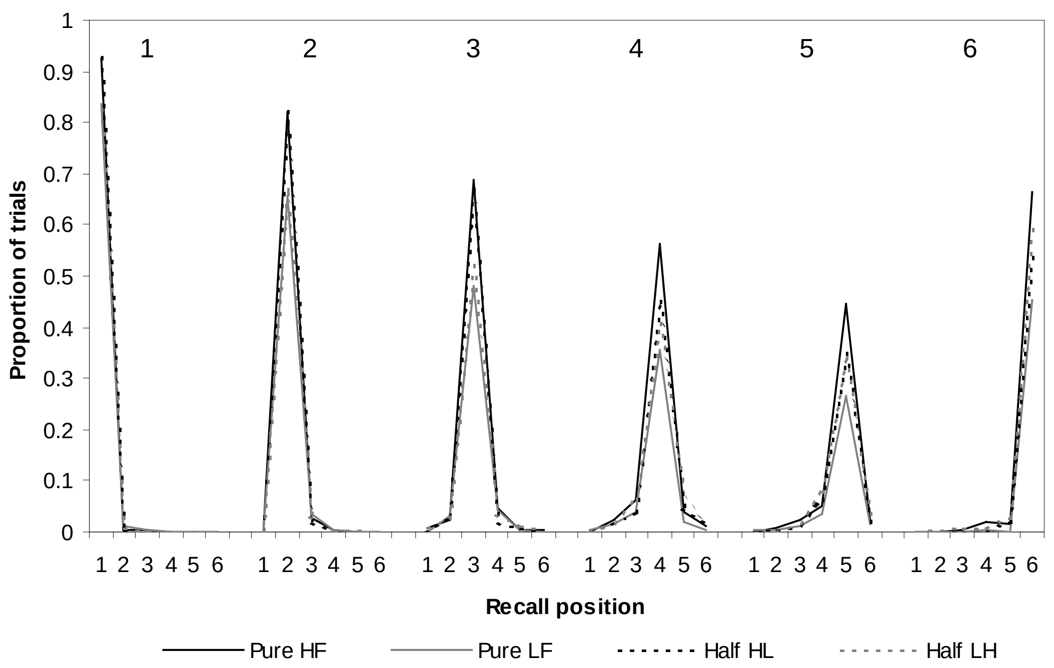


Figure 14.2. Transposition gradients for serial recall as a function of word frequency and list type in Experiment 5.

14.4 Discussion

This experiment was designed to contrast the recall of pure lists with half lists, and to provide a comparison between the recall of alternating and half lists in the context of experiments involving the recall of pure lists. Once more, statistical adjustment of the data to accommodate differences in speech rate between HF and LF words did not alter the pattern of results obtained with the unadjusted scores.

When items were presented in pure lists an HF advantage in recall across serial positions was found. Although a frequency by serial position interaction was present, this was attributable to the presence of a ceiling effect operating on the first serial position. Therefore, the results for pure lists are comparable to those found in both Experiment 3 and Hulme et al. (2003, Experiment 2), when open sets of items were used.

The half lists were found to follow the envelope of the first three positions of the pure lists; the differences between half lists beginning with HF and LF words were significant across these positions. In contrast, at and beyond the change in item type in half lists, recall was intermediate with respect to the pure serial recall curves. Furthermore, in the second half of the lists, recall for half lists was not different in any serial position, although the non-significant trend identified that the curves crossed over at the fifth position, as they did in Experiment 4, and separated again at the recency position. In line with Experiments 3 and 4, the correct recall of mixed lists across serial positions was greater for lists beginning with HF than LF words.

As highlighted in Experiment 4, the lag of the crossover in half list recall behind the change in list composition is not necessarily inconsistent with the notion of inter-item association. While changes in inter-item association would be expected to alter the rate of change in correct recall between items, the absolute difference between serial position curves prior to a change in list composition might dictate how far beyond a change in list composition the crossover actually occurs. Furthermore, when absolute recall levels are similar, as occurs at crossover points, the capacity of item-to-item associativity to generate significant effects between lists in subsequent positions might be limited. Alternatively, a non-significant difference between half lists observed at the recency position might be due to an item-specific influence operating on the end-of-list.

Experiment 5 did not produce any order effects, suggesting that inter-item associations between consecutive items assist the redintegration of item information for the subsequent item. The total item errors revealed that while a frequency effect for both list formats was evident this was reduced for the half lists. The size of the frequency effect for all errors was therefore related to the number of HF or LF words in the list. The category of EEI replicated this pattern. That is, the level of item errors increased for HF words but decreased for LF words when the list included both item types. In the case of omissions, the interaction was such that when the number of HF and LF was equated in the list, levels for each item frequency were equivalent.

Based on Experiments 3-5 the following remarks can be made. The frequency effect, in terms of overall level of correct recall, is a function of list composition, as pure lists exhibit greater frequency effects than mixed lists (Experiments 3 and 5), and mixed lists of the same composition produce total effects of a similar magnitude (Experiment 4). Nonetheless, the recall performance across serial positions of mixed lists reflects the order of list items. The recall of half lists is similar to pure lists for serial positions 1-3, but late positions are recalled at comparable levels that are intermediate to pure lists (Experiment 5). This contrasts with the recall of alternating lists that exhibit intermediate recall across all serial positions (Experiment 3).

There is some evidence of an item-specific contribution to the frequency effect, in terms of the better recall of alternating lists beginning with HF words (Experiments 3 and 4). If the level of serial recall was merely a product of the list-wide level of inter-item associativity or of associativity between consecutive items, alternating lists should produce identical levels of recall, regardless of whether the first item is HF or LF; this is apparently not the case. A second potential source of item-specific influence may come from the recall of items presented at the recency position (Experiments 3-4). Lastly, under some conditions, item-specific contributions have been observed with alternating lists (Experiment 3), in the form of a sawtooth pattern imposed on the serial recall curves, although this effect has not always been observed (Experiment 4), and has been found to be reliant on the strength of effect in the recency position (Experiment 3).

Order effects are inconsistent and small when present, in keeping with the general observations regarding word frequency and its impact on mnemonic information (Hulme et al., 2003; Poirier & Saint-Aubin, 1996; Stuart & Hulme, 2000). As a result of the general insensitivity of order accuracy to word frequency, the patterns of total errors produced mirrors the effect for correct recall, however the details for some error categories are enlightening. In particular the levels of omissions suggest that this category of error is mostly affected by list composition; when HF and LF items occur equally as often, the levels of omission by item type are equivalent (Experiments 3-5). This would imply that omissions are a function of list level properties, possibly non-directional associativity between items during redintegration. However, the results of Experiment 4 imply that item arrangement also has a small influence on the numbers of omissions produced, perhaps reflecting relative differences between list formats in the resources required to sustain item information; the pattern of omissions suggest that half-lists assist the retention of information more than alternating lists.

The patterns for EEI suggest that this error category may be influenced by compositional differences in lists, but not item arrangement in mixed lists. These errors reduce for LF items and increase for HF items in mixed rather than pure lists (Experiments 3 and 5) but an effect remains regardless of whether lists are alternating or half in format (Experiments 3-5). It is possible that EEI rates could result from a combination of list-level and item-specific effects. That is, while the mixing of HF and LF items in the same list leads to intrusion rates for HF and LF words that are closer than those encountered with pure lists, HF items retain a small protective advantage over LF words that might occur because of differences in the frequency of items.

14.5 Conclusion

Experiment 5 identified that the serial recall pattern produced for half lists is distinctly different to the recall of alternating lists. Therefore, in combination with Experiment 4, Experiment 5 produced evidence of a directionally sensitive component to the frequency effect. However, this influence appears not to be effective after the change in item type within half lists, as no difference between HF and LF items was observed for half lists in the last three serial positions. Furthermore, a non-significant difference in the recency position, suggesting that inter-item associative effects might operate in the second half of lists could also be attributed to a more general item-specific effect at the recency position. Therefore, in order to test whether recall produces a frequency effect in cases where items of the same frequency are sequenced after a HF to LF transition, a fourth experiment testing the recall of pure, alternating, half and a new list format ‘sequence’ (HHLLLH or LLHHHL), in a between-subject design, was conducted. In sequence lists the third item of the sequence corresponded to serial position 5, instead of the recency position that might be prone to contamination by other effects. Furthermore, analyses of the serial recall data involved the use of conditional probabilities to test whether transitions between HF and LF, and LF and HF, items would result in the same likelihood, providing a stronger test of the directional inter-item associativity hypothesis. Finally, in this experiment item errors were also examined by list half, in order to better diagnose the manner in which error types occur across list formats.

Chapter 15

Experiment 6: A Between-Subjects Comparison of Serial Recall Performance for Mixed Lists with Equal Numbers of HF and LF Items

15.1 Introduction

Experiment 6 was a between-subjects design comparing the within-subjects serial recall of two list types across four list formats. The recall of pure, alternating and half lists were once more investigated in this study. Additionally, a third mixed list format matching other mixed lists on composition, the ‘sequence’ list (HLLLHH or LHHHLL) was created to examine in greater detail the serial recall behaviour in mixed lists after a transition between HF and LF words had occurred.

To preserve continuity the analyses used on the data in Experiments 3-5 were conducted in this experiment. However, to better explore the consequences for recall performance when transitions between HF and LF words were encountered, this experiment incorporated an additional method of analysis using conditional probabilities. If associativity is bi-directionally equivalent, and if the recall of an item is a function of the inter-item associativity that it shares with the previous item in the list (Howard & Kahana, 2002b), then the likelihood it will be recalled, given that the previous item has been recalled, should be the same for HF to LF and LF to HF transitions. The standard serial recall analysis masks these sensitivities as all instances of successful serial recall at each serial position are accumulated, hence for this study recall in each position was also examined when conditionalised upon successful recall of the prior list item.

A second point of clarification addressed in Experiment 6 relates to the item error analyses. In Experiment 5 it was not possible to tell whether some advantages in error categories coincident with half lists occurred because of the stronger inter-item associations between consecutive HF words, or because this list format encouraged a positionally insensitive item-specific contribution to recall. Therefore in the reporting of this experiment, in addition to the analysis of item errors across the entire list, a breakdown between first and second halves of the list was performed. If differences in the frequency effect coincident with item arrangement are due to the associativity between consecutive list items, then these should be reflected as the presence of a greater effect in the first half of the list, as the serial recall data of Experiments 3-5 had identified that the first half of the list produces differences in performance.

A final difference in the data collection between Experiment 6 and Experiments 3-5 involved the omission of speech rate measures. As Experiments 3-5 had yielded negligible effects of articulatory duration these measures were not collected.

15.2 Method

15.2.1 Participants

A total of 103 undergraduate University of Wollongong students participated in the experiment for course credit. The data from 7 participants were excluded from analysis because they were not native Australian-English speakers (6), or were visually impaired (1). The remaining 96 participants were allocated to one of four conditions. The pure list participants (21 female, 3 male) had a mean age of 23.8 years (*SD* = 8.6 years), the alternating list participants (18 female, 6 male) had a mean age of 22.2 years (*SD* = 8.0 years), the half list participants (22 female, 2 male) had a mean age of 22.8 years (*SD* = 5.8 years), while the sequence list participants (19 female, 5 male) had a mean age of 21.0 years (*SD* = 3.8 years).

15.2.2 Materials

The same stimuli that were used in Experiment 3 were employed in Experiment 6. Script files contained 64 six-word trials and tested one of four list format conditions - pure frequency lists, alternating lists, half lists and sequence lists. The sequence list format was a variation on the half composition, where the sequence of items of the same frequency type in the second half of the list was brought forward by one serial position, and the displaced item type from the first half sequence was moved to the last serial position. Each item of the HF and LF word sets was presented twice within script files – once within each of the list types for the script file condition. Allocation of items to serial positions was again random within the constraints of each list format, and the presentation of individual trial types was random.

15.2.3 Procedure

Participants were assigned to one of the four list format conditions on a rotating basis, according to the order of their testing. The procedure for administering the experiment was the same as for Experiment 3, except that in this experiment speech rate measures were not taken.

15.3 Results

The data was scored according to the strict serial recall criterion. The mean number of correctly recalled items by serial position and condition is shown in Figure 15.1, where the distinct effects of list composition are readily observed. The patterns for the pure, alternating and half formats replicated those obtained in the previous experiments, while the novel sequence format demonstrated consistency with the half condition, in that the convergence of the curves occurred at the point of change of item type. The means for formats collapsed across list types suggested that the variation in the overall level of performance for each condition was small; descriptively, the items in pure lists were recalled the least well (M = .535), followed by the items in the alternating lists (M = .541), then by items in the half condition (M = .559), while the items in the sequence lists were recalled the best (M = .583). However, given that list format was the between-groups variable in this experiment this variation may be participant related. The overall means for the list types of each format revealed that lists beginning with HF words were better recalled than lists beginning with LF words; pure lists – HF (M = .634) versus LF (M = .436); alternating – HLHLHL (M = .556) versus LHLHLH (M = .526), half – HHHLLL (M = .576) versus LLLHHH (M = .541), and sequence – HHLLLH (M = .601) versus LLHHHL (M = .566).

In summary, it would appear that HF items in mixed lists beginning with LF words were better recalled than their counterparts in only a handful of serial positions. Furthermore, the differences where performance was superior in these lists did not compensate for the advantage that lists beginning with HF words realised in the other serial positions.

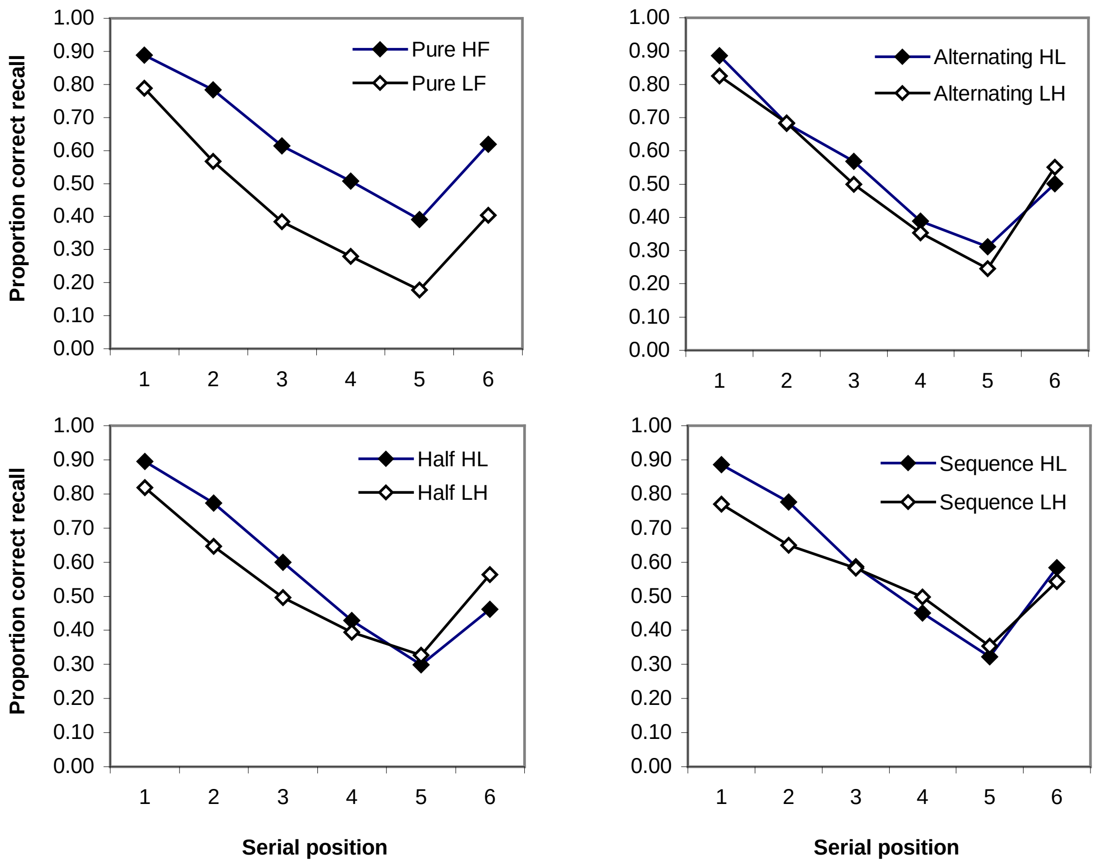


Figure 15.1. Serial recall of words as a function of frequency and list composition in Experiment 6. Panels represent the between-subjects conditions.

15.3.1 Serial recall

The criterion for significance was the same as for Experiment 3. The data were subjected to a 4 x 2 x 6 mixed analysis of variance where list format (pure, alternating, half, and sequence) was the between-subjects factor and list type (lists commencing with HF or LF words) and serial position (1 - 6) were the within-subjects factors. This analysis revealed a main effect of list type, $F(1,92) = 138.30$, $p < .001$, $MSE = .012$, confirming that lists beginning with HF words were better recalled than lists beginning with LF words, and serial position, $F(5,460) = 363.07$, $p < .001$, $MSE = .033$, but the effect of format was non-significant, $F(3,92) = 0.82$, $p = .480$, $MSE = .169$. Thus, differences in mean performances between formats were not reliable. The list type by list format interaction was significant, $F(3,92) = 42.80$, $p < .001$, $MSE = .012$, reflecting primarily the greater difference in recall of pure list types in comparison to differences between list types in the other formats. The list type by serial position interaction was also significant, $F(5,460) = 9.96$, $p < .001$, $MSE = .006$, driven by the asymmetries in the list types for the half and sequence formats. However, the interaction of format by serial position was non-significant, $F(15,460) = 1.07$, $p = .382$, $MSE = .033$, identifying that there was no difference across serial position in the average recall between formats. This result was qualified by a significant three-way interaction (format by list type by serial position), $F(15,460) = 9.70$, $p < .001$, $MSE = .006$, which determined that when performance was considered by list type, differences in serial recall curves between lists beginning with HF items and those beginning with LF items varied across the four formats.

This interaction was explored further by analysing the data for each format separately as four 2 x 6 repeated measures analyses of variance. For the pure lists there was a significant main effect of list type, $F(1,23) = 199.19$, $p < .001$, $MSE = .014$, replicating the standard frequency effect, and a significant main effect of serial position, $F(5,115) = 111.23$, $p < .001$, $MSE = .028$. The list type by serial position interaction was also significant, $F(5,115) = 5.52$, $p < .001$, $MSE = .005$, indicating the effect became larger over the first few serial positions. To resolve whether the interaction was a result of a ceiling effect operating on the first position for HF lists, the last five serial positions were re-analysed. The main effects were once again significant – list type, $F(1,23) = 219.10$, $p < .001$, $MSE = .013$, and serial position, $F(4,92) = 61.80$, $p < .001$, $MSE = .026$, but the interaction was non-significant, $F(4,92) = 0.13$, $p = .971$, $MSE = .005$, supporting a ceiling effect interpretation of the interaction.

A 2 x 6 repeated measures analysis of variance on the alternating list data produced a significant main effect of list type, $F(1,23) = 5.63$, $p = .026$, $MSE = .011$. Thus lists that began with a HF item were recalled modestly, but reliably, greater than those beginning with an LF item. The main effect of serial position was significant, $F(5,115) = 82.82$, $p < .001$, $MSE = .045$, as was the list type by serial position interaction, $F(5,115) = 6.30$, $p < .001$, $MSE = .006$. The presence of an interaction was due presumably to the subtle sawtooth pattern present in both list types, indicating the possibility of small item-specific effects. To investigate this pattern a contrast of the type conducted in Experiments 3 and 4 was run on this data, where the differences between alternating serial positions was compared. The contrast across all serial positions was significant $F(1,23) = 27.82$, $p < .001$, $MSE = .007$. Furthermore, the result persisted when the recency position was omitted from the analysis, $F(1,23) = 7.167$, $p = .013$, $MSE = .010$, thus demonstrating a result consistent with an item-specific effect not reliant on the behaviour of the final item in the list. While in Experiments 3 and 4 the corresponding analyses were also conducted on speech rate adjusted data, and this adjustment would reduce the effect otherwise obtained in the raw data, the magnitude of the *F* ratios found in the current case suggest that the significance of the effect would have survived any control for the speech rate of items. Bonferroni adjusted simple effects supported the interpretation of the item-specific contribution as a significant frequency effect was found for positions one, three and five.

An equivalent analysis of the half list format data resulted in main effects of list type, $F(1,23) = 11.79$, $p = .002$, $MSE = .008$, and serial position, $F(5,115) = 127.56$, $p < .001$, $MSE = .025$. Furthermore, the list type by serial position interaction was significant, $F(5,115) = 16.91$, $p < .001$, $MSE = .005$, highlighting the superiority of lists beginning with HF words. Bonferroni adjusted simple effects identified a significant frequency effect for positions one, two three and six. Therefore at the point of change in item frequency within list types (the fourth serial position) performance converged, remaining similar for the fifth position, and then diverged at the recency position where HF items were better recalled than LF items.

Lastly, within-subjects analysis of the sequence list data revealed main effects of list type, $F(1,23) = 6.51$, $p = .018$, $MSE = .013$, and serial position, $F(5,115) = 63.13$, $p < .001$, $MSE = .046$, while the list type by serial position interaction was also significant, $F(5,115) = 9.73$, $p < .001$, MSE = .007. The interaction followed a corresponding pattern to the half list format for the initial serial positions, confirming the better recall of the HHLLLH list type to the LLHHHL list type. Simple effects analysis using Bonferroni adjustment revealed that significant frequency effects were present in only the first two serial positions. At the first point of change in list composition (position three) performance converged, however in this condition the difference in the level of recall between list types did not change for the remainder of the serial positions (3-6).

15.3.2 Using conditional probabilities to examine inter-item effects

To explore the possibility that serial recall is driven, at least in part by some item-to-item associative mechanism, the data in Experiment 6 was rescored according to a conditional recall criterion. Should this be the case, it would be expected that the conditional likelihood that an item is recalled would be the same for transitions between HF and LF items in mixed lists as these have been argued to possess inter-item association of intermediate strength (Hulme et al., 2003).

Despite concerns regarding the accuracy of dependency measures (see Henson et al., 1996), it was thought, given the use of open stimulus sets in this instance that the use of transitional shift probabilities, i.e., conditionalised probabilities based on the recall status of the previous item only should be sufficient to indicate dependency due to the relationship between adjacent items. However, it is acknowledged that effects of disruption in recall later into the list may mask actual dependencies due to item-to-item association. Accordingly, it is appropriate to remain mindful that measures of this sort for later serial positions might contain influences from a number of sources.

A strict position interpretation was adopted for this data due to the short length of the supraspan lists. This method compares with the relative position scoring methods used to score much longer lists (Howard & Kahana, 2002a, 2002b; Klein et al., 2005). That is, for the current data the proportion of instances where an item presented in serial position *j* was recalled in position *j* on the condition that the previous items in position *j*-1 had also been recalled correctly was recorded. This data would identify any instances where inter-item associations between list items could explicitly operate. To account for the changing size of the sample space when calculating the conditional probabilities associated with these proportions, the following formula was used. Namely,

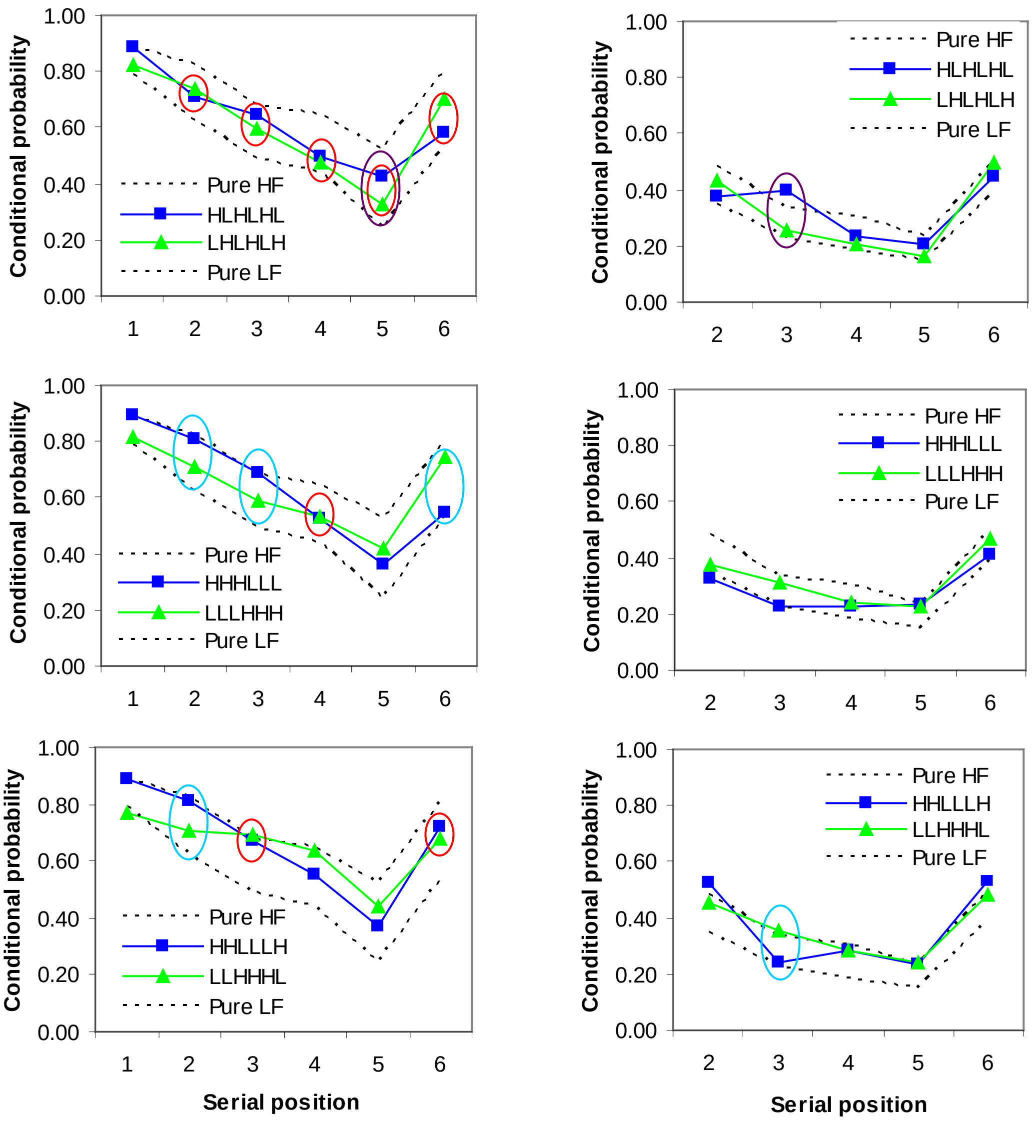
$$P(j \mid j-1) = \frac{P(j \cap j-1)}{P(j-1)} \qquad ,$$

where $P(j \mid j-1)$ is the probability of recall of an item in position *j* subject to the correct recall of the previous item $j-1$, $P(j-1)$ is the probability of correct recall of the prior item, and $P(j \cap j - 1)$ is the probability of the coincidence of events j and $j - 1$ (Larsen & Marx, 1981). The term $P(j - 1)$ becomes an estimate for the adjusted sample space. These probabilities are presented for each mixed list condition in Figure 15.2A. The data for the first serial position are the proportions recalled as per the original serial recall scoring, as there is no previous event for this case. The equivalent data for the pure lists is presented for each mixed list condition, and act as an envelope for the mixed list values.

The identification of recall events in this way allowed the data to be fractionated into the ‘continuous’ recall as described above, and those instances where recall occurred despite the previous item being in error, this was termed ‘recovery’ recall. In this situation recall might reflect an item-specific influence and indicate how well recall can recover from disruption at output for serial positions 2-6. The conditional probabilities for the recovery data are given in Figure 15.2B. The continuous and recovery recall data sets were examined separately.

Simple effects on the conditionalised data were conducted to determine whether they conformed to the patterns that directional inter-item associativity would anticipate; in terms of continuous recall, frequency effects would be expected in serial positions where corresponding sequences of HF and LF items were presented in a condition, while no difference in effect would occur at transition points in lists. Specifically, sequences of HF items should be better recalled than sequences of LF words because HF items have stronger pre-experimental association and HF to LF and LF to HF transitions in lists should result in the same level of recall. Therefore it would be predicted that positions two and three, and five and six in half lists, and positions two, four and five in sequence lists would produce differences. In contrast, no difference between the conditional probabilities of continuous recall should exist for all positions in alternating lists (2-6), while position four in the half lists and positions three and six in the sequence lists should also be equivalent. In addition, under the generous assumption that the fractionation of continuous and recovery recall accurately separates inter-item and item-specific effects, it would be expected that if item-specific effects do not influence recall, then no differences in the recovery data should exist for any of the conditions.

The analysis found that all points of transition between HF and LF items, except for position five in the alternating lists, produced non-significant differences in conditional probabilities for continuous recall; this exception was found to be a marginal result. Furthermore, significant differences were found for positions two and three and six in half lists, and position two in sequence lists. In summary, 1 out of the



A

B

Figure 15.2. The conditional probabilities of continuous (A) and recovery (B) recall between consecutive items for list types of mixed list conditions in Experiment 6. The dashed lines are the envelope formed from the conditional probabilities of the pure lists. Red circles identify whether inter-item associativity would predict no difference in likelihood of recall (serial positions 2-6). Blue circles show where effects exist, according to Bonferroni adjustment (serial positions 2-6). Purple circles indicate marginal effects (serial positions 2-6).

8 positions where no difference was predicted produced a marginal effect, while 4 out of the 7 positions predicted to produce a frequency effect did so. Therefore the continuous data, particularly across the first four serial positions, were consistent with a directional inter-item associativity explanation of serial recall.

The recovery recall identified that a significant difference occurred in position three of sequence lists. That is, at the point of transition in the list HF words were recovered better than LF words after the previous item was not correctly recalled. None of the other comparisons for this data reached significance, although position three for alternating lists was marginal. In general then, according to this analysis, recovery episodes were free of the influence of frequency.

15.3.3 Item Analysis

Item errors, as previously defined, were identified for item types within each list format and are presented in Table 15.1. Order errors were conditionalised to indicate the extent to which recalled items were positioned incorrectly relative to presentation order. This process identified that the adjusted error rate for HF items in pure lists was .127, for LF items in pure lists was .149, for HF items in alternating lists was .119, for LF items in alternating lists was .115, for HF words in half lists was .110, for LF words in half lists was .122, for HF words in sequence lists was .100 and finally for LF words in sequence lists was .099. In addition, the transposition curves for this experiment are presented in Figure 15.3. They highlight that any differences in order memory do not appear to be associated with arrangement of items in the list by frequency type.

A series of analyses of variance were conducted on the error data. Specifically, a 4 x 2 (list format by item type) mixed ANOVA was performed on the conditionalised order data and identified non-significant effects of frequency $F(1,92) = 2.35$, $p = .129$, $MSE = .001$, and list format, $F(3,92) = 1.86$, $p = .142$, $MSE = .007$, and a non-significant frequency by format interaction, $F(3,92) = 1.82$, $p = .149$, $MSE = .001$. Thus there was no difference between HF and LF stimuli in the proportion of all items remembered that were recalled in the wrong position.

The analysis of the total item error data revealed an effect of frequency, $F(1,92) = 254.85$, $p < .001$, $MSE = .002$, and a significant frequency by list format interaction,

$F(3,92) = 51.26$, $p < .001$, $MSE = .002$, but the main effect of list format was not significant, $F(3,92) = 0.52$, $p = .672$, $MSE = .022$. Across conditions, HF words incurred fewer item errors than LF words. The interaction reflected the greater frequency effect for pure than mixed lists. Bonferroni adjusted tests on each list format identified however that all conditions produced significant frequency effects; pure lists [$t(23) = 16.59$, $p < .001$], half lists [$t(23) = 5.78$, $p < .001$], sequence lists [$t(23) = 4.55$, $p < .001$], and alternating lists [$t(23) = 3.51$, $p = .002$].

Table 15.1

Mean proportion of items correctly recalled and mean proportions of different error categories as a function of item type and list format in Experiment 6

	Correct	Errors			
		Order			Item
		Total	ILI	Trans	
HF in pure lists	.634 (.125)	.090 (.044)	-	.090 (.044)	.277 (.115)
LF in pure lists	.434 (.101)	.073 (.035)	-	.073 (.035)	.494 (.095)
HF in alt. lists	.559 (.128)	.071 (.029)	.056 (.026)	.015 (.012)	.376 (.119)
LF in alt lists	.524 (.126)	.063 (.030)	.050 (.024)	.013 (.010)	.407 (.110)
HF in half lists	.592 (.113)	.071 (.033)	.024 (.016)	.047 (.024)	.337 (.101)
LF in half lists	.525 (.136)	.069 (.034)	.024 (.015)	.045 (.030)	.406 (.124)
HF in seq. lists	.613 (.107)	.065 (.033)	.028 (.016)	.036 (.022)	.328 (.090)
LF in seq. lists	.554 (.135)	.054 (.024)	.022 (.012)	.032 (.018)	.386 (.118)
	Item errors by category				
	Omissions	EEI	ISI	IEI	Other
HF in pure lists	.189 (.127)	.041 (.031)	.028 (.019)	.012 (.009)	.007 (.009)
LF in pure lists	.314 (.133)	.102 (.053)	.030 (.020)	.038 (.020)	.011 (.009)
HF in alt. lists	.265 (.115)	.054 (.027)	.032 (.018)	.019 (.012)	.007 (.007)
LF in alt lists	.265 (.122)	.075 (.033)	.025 (.014)	.033 (.025)	.008 (.007)
HF in half lists	.245 (.089)	.041 (.027)	.028 (.021)	.016 (.011)	.008 (.007)
LF in half lists	.268 (.122)	.075 (.032)	.024 (.014)	.024 (.014)	.015 (.011)
HF in seq. lists	.214 (.091)	.050 (.031)	.032 (.028)	.023 (.015)	.010 (.008)
LF in seq. lists	.230 (.120)	.081 (.037)	.029 (.019)	.034 (.022)	.013 (.011)

Note. ILI – Intra-list intrusions, Trans – transposition to a same frequency position in the list, EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Other – repetitions and nonwords. Standard deviations are given in brackets.

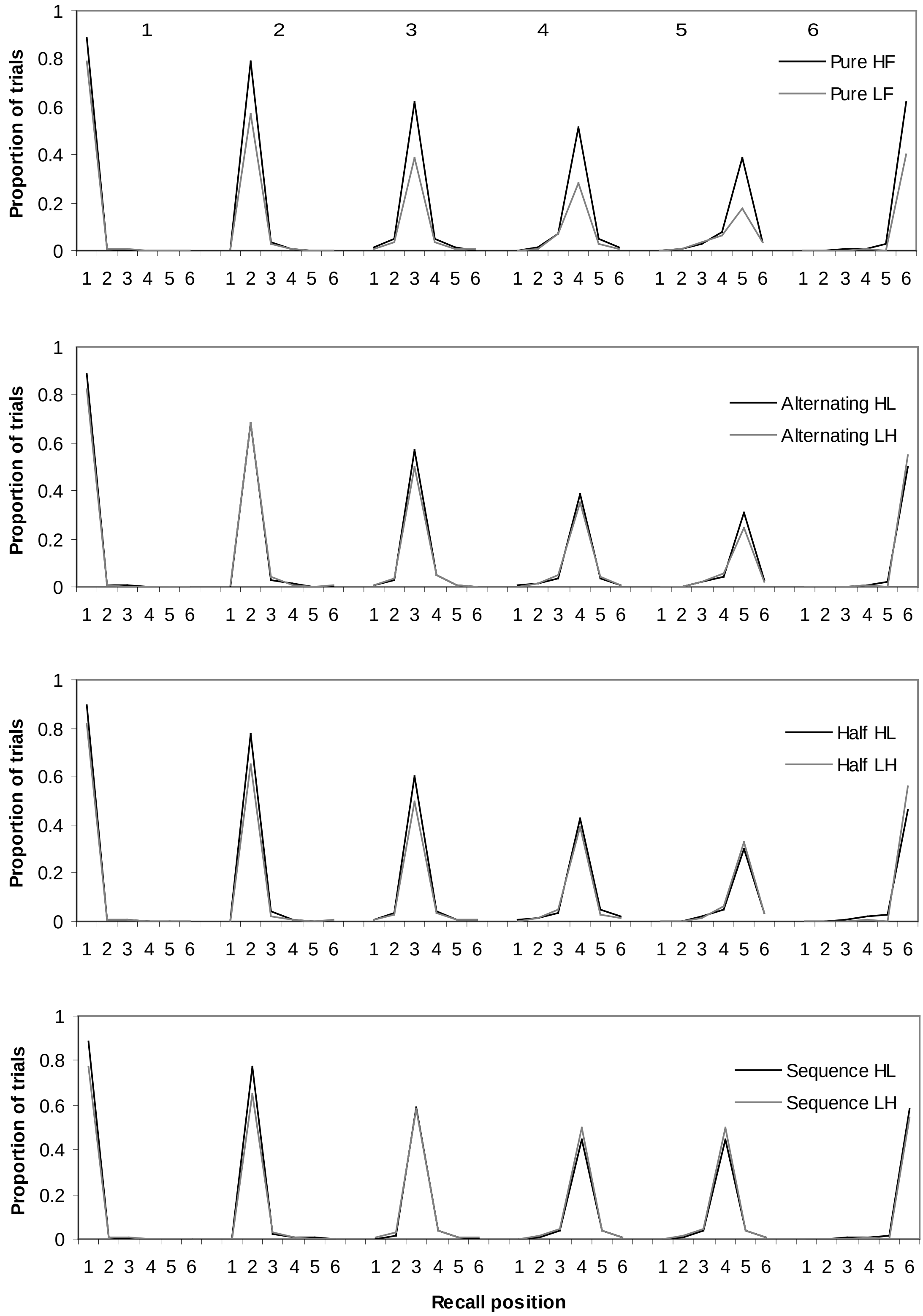


Figure 15.6. Transposition gradients for serial recall as a function of word frequency and list type in Experiment 6. Between-subjects conditions are depicted separately.

An examination of the omission patterns across formats determined that frequency had a significant effect, $F(1,92) = 56.23$, $p < .001$, $MSE = .001$, but not format, $F(3,92) = 0.67$, $p = .573$, $MSE = .025$. The frequency by format interaction however was significant, $F(3,92) = 26.72$, $p < .001$, $MSE = .001$. Bonferroni adjusted multiple comparisons showed that while there was a significant difference in omissions between HF and LF positions for pure lists [$t(23) = 12.12$, $p < .001$], other conditions did not differ; half [$t(23) = -1.89$, $p = .071$], sequence [$t(23) = -1.26$, $p = .220$], and alternating [$t(23) = -0.05$, $p = .961$].

An analysis on the EEI data found that frequency, $F(1,92) = 152.55$, $p < .001$, $MSE = 4E\text{-}4$, was significant but list format, $F(3,92) = 0.72$, $p = .543$, $MSE = .001$, was non-significant. The interaction for this category however was also significant, $F(3,92) = 8.02$, $p < .001$, $MSE = 4E\text{-}4$, with the difference between item types being greatest for the pure list condition. All multiple comparisons revealed significant differences between item types, pure [$t(23) = -10.44$, $p < .001$], half [$t(23) = -4.80$, $p < .001$], sequence [$t(23) = -7.35$, $p < .001$], alternating [$t(23) = -3.40$, $p = .002$]. Low frequency items were more likely to be replaced at recall by a word outside of the experimental sets than HF words. Therefore, it would appear that the differences found in the total item errors were mostly comprised of differences from the EEI error category.

The remaining error categories explained 4-8% of all item outcomes. Differences between frequency types for ISI were modest, if present at all. In addition the trend for IEI across formats indicated that more items from the HF pool intruded into LF positions in the lists, than vice versa.

Finally, in order to determine distributions of item errors according to approximate location in the lists (first half versus second half), the item error data was partitioned and is represented in Tables 15.2 and 15.3. The data identified that while item error differences between HF and LF items in pure lists were generated in both list halves, in mixed lists small differences tended to be present for the first half of the list only. The main category of item errors, omissions, produced greater rates in the second than first half of lists. Omissions furthermore exhibited a weak trend consistent with a directionally sensitive inter-item associativity mechanism present in the first half of the list, but the magnitudes of effect were small. In the first half of the lists, sequences of HF items found in pure and half lists were associated with lower omission rates resulting in small frequency effects, while the corresponding

Table 15.2

Mean proportion of items correctly recalled and mean proportions of order and item errors as a function of item type, list format, and list half in Experiment 6

			Correct	Order Errors	Item Errors
Pure					
	First half				
		HF	.381 (.058)	.033 (.020)	.077 (.048)
		LF	.290 (.059)	.026 (.017)	.172 (.048)
	Second half				
		HF	.253 (.081)	.057 (.031)	.199 (.077)
		LF	.143 (.059)	.048 (.024)	.321 (.063)
Alternating					
	First half				
		HF	.357 (.068)	.029 (.015)	.116 (.056)
		LF	.335 (.070)	.025 (.012)	.135 (.058)
	Second half				
		HF	.202 (.076)	.042 (.020)	.261 (.077)
		LF	.189 (.071)	.038 (.023)	.272 (.065)
Half					
	First half				
		HF	.378 (.052)	.028 (.018)	.091 (.044)
		LF	.327 (.067)	.022 (.016)	.147 (.059)
	Second half				
		HF	.214 (.076)	.043 (.023)	.246 (.070)
		LF	.198 (.075)	.047 (.025)	.258 (.073)
Sequence					
	First half				
		HF	.374 (.056)	.023 (.016)	.103 (.047)
		LF	.334 (.078)	.020 (.013)	.140 (.069)
	Second half				
		HF	.239 (.074)	.042 (.020)	.225 (.068)
		LF	.219 (.074)	.035 (.017)	.245 (.064)

Note .Standard deviations are given in brackets.

differences for alternating and sequence lists were negligible. Furthermore, in mixed lists a frequency effect was not present in the second half of the lists. Therefore it seems that item arrangement affects omission levels early in the list, to some degree, while list composition (pure versus mixed) determines whether an effect for

Table 15.3

Mean proportions of item errors by item error category as a function of item type, list format, and list half in Experiment 6

			Omissions	EEI	ISI	IEI	Other
Pure							
	First half						
		HF	.041 (.040)	.018 (.016)	.012 (.010)	.005 (.006)	.001 (.003)
		LF	.086 (.048)	.051 (.027)	.014 (.011)	.018 (.008)	.002 (.004)
	Second half						
		HF	.148 (.093)	.023 (.019)	.016 (.011)	.007 (.006)	.006 (.009)
		LF	.228 (.100)	.050 (.030)	.016 (.017)	.019 (.015)	.008 (.007)
Alternating							
	First half						
		HF	.065 (.041)	.023 (.016)	.017 (.013)	.009 (.008)	.002 (.003)
		LF	.074 (.049)	.034 (.016)	.014 (.011)	.013 (.012)	.000 (.001)
	Second half						
		HF	.200 (.084)	.031 (.010)	.015 (.010)	.010 (.008)	.005 (.006)
		LF	.192 (.082)	.041 (.022)	.012 (.009)	.020 (.017)	.008 (.006)
Half							
	First half						
		HF	.057 (.037)	.016 (.013)	.011 (.013)	.005 (.007)	.001 (.002)
		LF	.086 (.054)	.037 (.021)	.009 (.009)	.013 (.008)	.003 (.003)
	Second half						
		HF	.188 (.067)	.025 (.017)	.016 (.012)	.011 (.006)	.007 (.007)
		LF	.182 (.078)	.038 (.020)	.015 (.012)	.012 (.009)	.012 (.011)
Sequence							
	First half						
		HF	.057 (.034)	.020 (.019)	.014 (.015)	.010 (.009)	.002 (.003)
		LF	.070 (.062)	.037 (.022)	.013 (.009)	.015 (.012)	.004 (.007)
	Second half						
		HF	.157 (.074)	.029 (.016)	.018 (.017)	.013 (.009)	.008 (.009)
		LF	.159 (.073)	.042 (.023)	.016 (.013)	.018 (.014)	.009 (.009)

Note. EEI – Extra-experimental intrusions, ISI – Intra-set intrusions, IEI – Intra-experimental intrusions, Other – repetitions and nonwords. Standard deviations are given in brackets.

omissions is present in the latter half of the list. The EEI produced rates that were typically greater for the second than first list halves, and greater for LF than HF words. Once more, a weak tendency towards a consecutive inter-item contribution is observable, particularly in the contrast between the differences for first and second list halves for half and alternating lists. In addition, and more generally, the frequency effect in EEI was greater for pure than mixed lists implicating list composition in the manifestation of this type of error.

15.4 Discussion

Experiment 6 was designed to explore the impact of shifting a three-item sequence, as found in the second half of half lists, one position forward. This arrangement served to avoid the coincidence of the third item in sequence with the recency position, and test whether frequency effects between list items were possible in the second half of the list after a transition between HF and LF items had been made. In addition Experiment 6 was performed as a replication of the conditions in Experiments 3-5 using a between-subjects design that would eliminate any carryover effects, should they exist.

The analysis of the correct serial recall data for pure lists confirmed a frequency effect of the form encountered earlier. The frequency by serial position interaction was found to be a function of a ceiling effect acting on the first serial position. This result concurs with the pure lists results reported in Experiments 3 and 5, as well as the pattern observed by Hulme et al. (2003, Experiment 2). The results for alternating lists indicated once more that lists beginning with HF items experience a small advantage in correct recall, possibly as a consequence of an item-specific contribution that operates at the start-of-list position. Furthermore, in this experiment and as found in Experiment 3, a small list-wide item-specific effect was observed as a subtle sawtooth in the serial recall curves.[1] Evidence for a directionally sensitive contribution of inter-item associativity was again produced in the recall of half lists, as in Experiment 5 recall mimicked pure lists across the first three serial positions. However, recall beyond the point of change in item composition did not produce a reliable frequency effect, except for the recency position, which, as previously discussed, may also respond to the item-specific properties of items.[2]

---

[1] Alternatively slight differences in the directional associations between HF and LF, and LF and HF words, respectively might be responsible for this pattern.

[2] Note however that while effects at recency have been consistent with item type in Experiments 3-6, Hulme et al. (2003) found a marginally significant negative frequency effect, that is, better recall for LF than HF words, in the recency position with alternating lists of their Experiment 2.

The analysis of correct recall for sequence lists failed to find a significant frequency effect for the third HF and LF items in sequence (the fifth serial position). Therefore, according to this data, influences of inter-item associativity between list items are insufficient to support differences in the recall of mixed lists in the second half of the list. While non-significant trends indicating better recall for HF than LF words exist, more generalised contributions, for example the non-directional associativity between list items (Hulme et al., 2003; Stuart & Hulme, 2000), might determine the overall success of recall in late serial positions.

The possibility that inter-item associativity between consecutive list items had been obscured by the absolute levels of recall of the previous items in correct serial recall scoring, and masked in the second half of the lists by instances of recovery after a failure to recall the previous item, was considered by fractionating the data into continuous and recovery recall events, and expressing these as conditional probabilities. Tests on these probabilities, gauging how well the continuous data conformed to a directional inter-item associativity explanation, indicated that the first three to four list positions were well accounted for, although a significant difference in the recovery data occurred for sequence lists, where recovery of HF words in the third serial position was greater than for LF words. Accordingly, a second analysis reinforced the proposition that inter-item effects of a directional nature operate for the primacy portion of the list, these effects are less influential for late list positions, as identified by the finding that only 1 out of 4 positions where sequences of HF and LF items occurred in the second half of the lists produced a significant effect, and this instance coincided with the recency position.

This experiment did not find any order effects related to word frequency, placing the locus of frequency effect(s) once more with differences in the retention of item information (Hulme et al., 2003; Poirier & Saint-Aubin, 1996; Stuart & Hulme, 2000). The pattern of total item errors revealed that list composition influenced the degree to which item recall for HF words was superior to recall for LF words; the effect was greater for pure than mixed lists. The inclusion of HF and LF words in the same list reduces errors for LF words and increases them for HF words, relative to pure list performance (Hulme et al., 2003). However all list formats produced frequency effects for the total error data. Therefore, by some means, and despite differences in item arrangement, LF words were associated with greater item error rates than HF words across mixed list conditions. These results contrast with those of

Hulme et al. (2003, Experiment 2) who found a reversed frequency effect for item errors in alternating lists.

The results for the levels of omissions identified that while the failure to recall an item of either frequency type was the same for each mixed list, the inability to recall an item was more likely for LF than HF items in pure lists. Therefore, differences in omissions for HF and LF words would appear to be a function of list composition, and possibly reflect the operation of a non-directional, list-level mechanism that naturally advantages pure HF over pure LF lists.

The pattern observed for the total item errors was shadowed by that for EEI, the intrusion of non-experimental words into the recall of list items. This occurred more for LF than HF items, but the effect was greater for pure than mixed lists. Effects for all mixed list conditions however were still evident and these were independent of item arrangement; this would imply that EEI are a result of item-specific properties in combination with list-level influences.

However, the categories of error examined by list half suggested the attributions of the various facets of the frequency effect to each error category might not be as straightforward as described above. As would be anticipated by the pattern of correct recall, the division of total item errors into first and second halves suggested that differences between HF and LF words in each of the mixed lists were accounted for primarily by differences in error rates in the first half of the list, implicating item arrangement as a contributing factor. However, while it is possible that omissions and EEI contributed to this trend, the effects associated with these rates were small.

Therefore based on these observations, it is proposed that the serial recall of lists and the generation of item errors are dominated by a composite of associative mechanisms that are directional and non-directional in nature, and operate in different regions of the list. The recall of items presented in early positions is affected by the level of association between consecutive items, while recall for later items is dominated by the level of non-directional associativity drawn from all list items. A small start-of-list item-specific effect in the recall of alternating lists is consistently reported. However, other potential sources of item specific effect (the sawtooth in alternating lists, and differences found in the recency position) are less consistent and may be dependent on specific differences in testing.

Accordingly, lists where HF to LF transitions occur in the first few serial positions (e.g. alternating lists) will produce recall performance consistent with the moderate strength of association between items. Mixed lists that contain sequences of HF items at the start-of-list will be reliably better recalled than those containing sequences of LF words (half and sequence lists) in these serial positions, while the equivalence of these mixed lists in terms of non-directional associativity will produce similar levels of recall for the later list items.

The incidence of item errors is compatible with this view. These correspond with variations in correct recall for early and late list portions across list conditions. Item-to-item associativity is seemingly reflected in the rates of omissions and EEI experienced in the first half of lists and suggests that the better recall of early HF than LF sequences in mixed lists correlates with better retention of item information, at least in a partial state, and possibly better protection against intrusions from non-experimental items.

Earlier versions of the redintegration hypothesis envisioned this process acting individually and serially on temporary traces retrieved from the short-term store (Hulme et al., 1997; Schweickert, 1993). However, Hulme et al. (2003; Stuart & Hulme, 2000) argued that the serial recall of alternating lists demonstrated that the impact of LTM knowledge on STM performance could not be at the level of the individual item; the relationship between list items would determine how well an item was recalled. More, specifically the inter-item associative links between all presented items would determine the availability of each LTM representation underpinning the reconstructive process. While a two-component model of this kind can conceivably maintain separate short-term and long-term stores that interact only during the recall of individual items, it cannot exist in this form if the recall of list items is dependent on the inter-item associative links across the list set. Somehow this activation pattern must be acquired prior to the recall. Hulme et al. (2003) stated that "…if redintegration is an important process in serial recall it must be conceived of as operating upon (and being influenced by) the composition of a retrieval set that is created in memory following list presentation" (p. 514). The question then arises as to how this retrieval set is formed.

Recently Stuart and Hulme (2009) outlined in greater detail the mechanism of non-directional inter-item associativity they thought would underpin redintegration and produce the context-based variability found with the frequency effect (Hulme et al., 2003; Stuart & Hulme, 2000). They argued that each time a stimulus is *presented* the activation of its LTM representation is raised. Furthermore when an item is presented in the context of a list, this activation travels along pathways connecting it to its activated cohort. Activation is influenced by the pre-existing strength of connection between items as a function of word co-occurrence in language. Importantly, this interpretation of redintegration admits interaction between the STM and LTM from the point of presentation onwards, thus discarding the notion of a separate short-term store that references LTM at a late-stage, and bringing it closer to the assumptions held by language-based models.

Ironically, the pattern of directional associativity followed by nondirectional associativity for the final items could be approximated by a strictly late-stage interaction with LTM where the short-term traces of each item access LTM in turn. With early list items the size of the 'net' would be small and this would place relative emphasis on the activation between list neighbours as a function of co-occurrence in natural language. Late items would be supported by the maturing activation developing from associativities across the list cohort during recall, and this would be nondirectional in kind. However, this late-stage proposal neglects the increasing evidence that lexical-semantic variables influence STM performance prior to recall (e.g. Jefferies et al., 2006b; Romani et al., 2008) making the possibility LTM has such a restricted role unlikely.

Yet explanations that admit interaction with LTM from the point of presentation onwards (e.g. N. Martin & Saffran, 1997; R.C. Martin et al., 1999; Stuart & Hulme, 2009) must be able to explain how associativity evolves from a directional to a nondirectional influence. Unless the development of an activation net based on input is slow and requires the duration of the presentation and recall periods to fully develop in a nondirectional sense, it would be anticipated that the cohort would act as a mutually supportive set of activations at the point of recall, and that HF and LF items regardless of serial position should be recalled according to the level of activation determined by this set. Without some form of re-instantiation for early list items maintaining a directional bias in the support of item information, these explanations are incomplete.

Accordingly, some other factor is likely to determine the change in supportive activation across serial positions. For example, the part of the list observed to reflect item-to-item associative effects in serial recall corresponds to the subspan of items that are cumulatively rehearsable within the inter-stimulus interval at presentation

(Page & Norris, 1998; Tan & Ward, 2008)[3]. It is possible that the extent that item-to-item associations facilitate the recall of words reflects the degree to which associativity promotes the retention of those items across the period when rehearsal is typically conducted and additional mnemonic material is presented. Alternatively, Allen and Hulme (2006) found that serial recall was most closely associated with how well the semantic representation of items can elicit the speech output of the corresponding words. A second possibility is that the observed change in directional to nondirectional associativity reflects the development of an associative net that is specific to speech output, and is created independently from existing activation in the language processor during overt recall. Lastly, the recall of an early list item is likely to harm the recall of items later in the list (Cowan, Saults, Elliott, & Moreno, 2002; Nairne, 1990; Oberauer, 2003; Tan & Ward, 2007), however Nairne, Ceo, and Reysen (2007) have shown that the preceding item can act as a cue to recall of the next item, and that this cueing has a tendency to diminish for items in the recency portion of the curve. As a result, output effects might contribute to the pattern of associativity found in the recall of mixed lists, and half lists in particular. These possibilities are considered further in the General Discussion (Chapter 16).

On a related note, although coarticulatory fluency is not thought to be an active feature in the current investigation, it is worth considering whether the presumed effects of this variable (Woodward et al., 2008) are consistent with the patterns of results observed across the experiments performed. A simple interpretation of this approach would predict that recall would be a function of the relative difficulty of the coarticulation of phonemes constituting word boundaries as indexed by word frequency. Consequently, sequences of HF items should be more reproducible in terms of speech programming than sequences of LF words, and this does occur for items in the first half of the list; this difference is as great as it is for pure lists (Experiment 5). However the absence of a frequency effect between sequences of HF and LF words in the second half of the list suggests that influences wider than the coarticulation at word boundaries determine recall levels for these serial positions. It could be argued for example, in the case of half lists, that the reduction in recall of HF items in the second half of the list when compared to the recall of pure lists would be

[3] Although the presentation rate used in Experiments 3-6 was 1 word per second, all items had a CVC structure and so could be rehearsed rapidly. Post hoc speech rate measures in Experiments 3-5 estimated this to be approximately 3 items per second, although the test environment might encourage faster production again.

due to the impairment to speech planning brought about by the relative difficult in sequencing three LF words in the first half of the list. Similarly, the recall for LF items in the second half of the list, observed to be greater than the recall for pure LF lists, could be explained as the benefit afforded these items by the more efficient speech programming for the first half of the list. An unambiguous test of this proposal would be to conduct replications of these studies using articulatory suppression. If the recall patterns persist for mixed lists, as more generally the frequency effect has been seen to endure under articulatory suppression (Gregg et al., 1989; Tehan & Humphreys, 1988), this would act as clear evidence that coarticulatory fluency at word boundaries is not the mechanism responsible for the frequency effect.

Saint-Aubin and LeBlanc (2005) argued that the frequency effect could be explained in terms of the distinctiveness of list items combined with the resources required to reconstruct items of a particular frequency level. High frequency items were assumed to be more distinctive from other HF items than LF items were to other LF items. This was used to explain the standard frequency effect in pure lists. The recall of an LF word could be enhanced if it was presented amongst HF words as this increased its distinctiveness from other list items, although recall of an LF isolate was not better than the recall of HF words in pure lists. High frequency isolates were better recalled than the LF items they were presented with, however they were not recalled as well as or better than HF words in a pure list. To accommodate these observations it was reasoned that the greater processing costs associated with recalling a list constructed predominantly of LF words would be responsible for the difference in recall between HF isolates and HF words in pure lists.

Furthermore, this argument could be extended to explain the serial recall pattern for alternating lists found by Hulme et al. (2003). The abolition of the frequency effect for these lists would be due to the homogenisation of distinctiveness resulting from the mixing of LF and HF words. The intermediate level of performance against pure lists would be a function of the processing required in the reconstruction of items, given the mixed composition of the lists.

If these were the only facets of the stimuli responsible for the frequency effect, then, in the simplest terms, mixed lists possessing the same number of HF and LF items should produced the same level of recall, as they are matched in terms of the distinctiveness of items and the overall resources required to process them. This is clearly not the case and contrasts with the results obtained for word length. Hulme et al. (2004) determined in a serial order reconstruction task that the performance on randomly ordered lists of three long and three short words was not different to either alternating lists of long and short words or pure lists of short words. While these results were not derived from a serial recall task, they corresponded to an earlier serial recall experiment comparing recall for alternating and pure lists of long and short words (Hulme et al., 2004), suggesting that the effect of word length is not affected by item arrangement. Therefore, it would appear that the primary mechanism responsible for the word length effect is markedly different to the mechanism(s) determining the frequency effect.

However, it is also apparent that in contexts where the numbers of HF and LF items are not balanced, an effect functionally equivalent to distinctiveness might operate in short-term recall, as the results reported by Saint-Aubin and LeBlanc (2005) are not fully explained by the associative mechanisms described above. Clearly this will require further investigation to determine how associative effects might alter with a change in list composition and whether any higher order mechanisms might be responsible for these changes.

15.5 Conclusion

This chapter reported on an experiment contrasting the serial recall performance of three forms of mixed list that were matched on the numbers of HF and LF items and performance with pure lists. Analyses suggested that in mixed lists the recall of items in early serial positions is influenced by a directional form of inter-item associativity supporting the retention of item information, while recall performance in the second half of the list is dominated by non-directional asssociativity derived from the associations of items across the list set. Furthermore, across Experiments 3-6 support for Hulme et al.’s (2003) claim that recall might be a product of either wholly directional or wholly nondirectional associative influence was not found. These results are not readily explained by any of the currently endorsed explanations of the frequency effect without modification. The implications for future research and development of theory are examined in more detail in the General Discussion (Chapter 16).

Chapter 16

General Discussion

16.1 Introduction

The aim of the research in this thesis was to contribute to a better understanding of the frequency effect in serial recall. This was accomplished by investigating two components of the processing of word frequency. Experiments 1-2 explored how word frequency relates to another lexical-semantic property of words, namely concreteness, and yields further insight into the relationships between linguistic attributes as they contribute to serial recall. These experiments identified similarities between the processing of these variables that seem best explained within a language-based framework (e.g. R.C. Martin et al., 1999; Romani et al., 2008).

A second strand of experimentation (Experiments 3-6) examined the behaviour of mixed frequency lists. More specifically, Experiments 4-6 included conditions where items of the same frequency type were sequenced in lists. These experiments provide complementary information to the results for alternating lists previously reported by Hulme et al. (2003; Morin et al., 2006) and in Experiment 3 of this thesis, and give a clearer indication of how adjacent items early in the list relate to each other as a function of word frequency. In addition, this research refutes the proposition that short-term recall is supported solely by nondirectional and nonspecific activation from all list items across serial positions (Hulme et al., 2003; Stuart & Hulme, 2000, 2009).

This chapter presents a brief historical overview of the various explanations of the frequency effect in serial recall and attempts to position the current findings in the contexts of the most recently endorsed frameworks. It also considers those aspects of the current research that are not well accommodated for by current theory, and offers suggestions for further research to better specify these outcomes and appropriately inform the development of theory.

16.2 STM through the lens of the frequency effect: results in context

Early conceptions of STM were drawn from a unitary perspective that emphasised the timescale over which material was retained for recall (Melton, 1963). As such, from a temporal perspective, STM represented one end of the memory continuum of a single system. This view contrasted with the position that STM is a kind of memory distinct from LTM, is subserved by specialized systems or capacities and is subject to different processes (e.g. decay) from LTM (Baddeley & Hitch, 1974).

The unit-sequence interference hypothesis (Postman, 1961, 1962; Turnage, 1967; Underwood & Postman, 1960), drawn from unitary accounts of memory, argued that the serial recall of HF items should be harmed more than LF items because stronger pre-existing associations between HF words, as a function of association in language (Deese, 1959, 1960; Postman, 1962; Underwood & Postman, 1960), would disrupt memory for the sequence as presented. However, Baddeley and Scott (1971) demonstrated that such a disruption was absent from serial recall when it was performed within the timescale of STM operation.

Memory theory evolved to view STM as a specialised entity with dedicated storage and a reliance on phonological processing (Atkinson & Shiffrin, 1968, 1971; Baddeley & Hitch, 1974). Atkinson and Shiffrin's (1968, 1971) framework argued that the key role of STM processing was to transfer information to the LTS through rehearsal. Within this approach, an LTM variable such a frequency might influence the degree to which the transfer of information could be accomplished, and the consequent durability of an episodic trace. This interpretation would be supported by results displaying frequency effects for items spending larger amounts of time in the STS (e.g. M.J. Watkins; O.C. Watkins & Watkins, 1977).

The phonological loop account of verbal STM also became an influential framework through which short-term serial recall results were interpreted (e.g. Baddeley et al., 1975). It took the position that STM acted as a self-contained system that managed speech-based material and emphasised the role of subvocal rehearsal in the maintenance of to-be-remembered items; those items that were more efficiently rehearsed, as indicated by differences in speech rate or other physical discrepancies (e.g. numbers of syllables), were less likely to experience decay in the fixed capacity STS, and were therefore better recalled. Consequently, variable manipulations that could be reduced to differences in rehearsal rate, consistent with the direction of effect, could be explained by this model. Wright (1979) demonstrated that HF words were articulated faster than LF words, and this formed the basis of a reinterpretation of the frequency effect as a reflection of relative rehearsal efficiencies between HF and LF words.

However, when speech rate differences were accounted for, either in terms of experimental design (Gregg et al., 1989; Tehan & Humphreys, 1988) or statistical adjustment (Roodenrys et al., 1994; Hulme et al., 1997; Hulme et al., 2003), a frequency effect persisted, and posed a serious challenge to the explanatory power of the phonological loop. While it was possible that STM was sensitive to differences in frequency by virtue of differences in rehearsal (Gregg et al., 1989), a contribution from a second memory system, also sensitive to frequency was evident (Hulme et al., 1997; Roodenrys et al., 1994; Tehan & Humphries, 1988). Hulme et al. (1991) had shown that the lexicality effect also produced a non-speech rate residual span and had nominated phonological LTM as its source. Words possessed an advantage over nonwords because they were represented in LTM, but lexical representations for nonwords by definition were absent. These long-term representations would assist the late-stage reconstruction of short-term phonological traces at the point of recall in much the same way that a speech error might be corrected (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Roodenrys et al., 1994; Schweickert, 1993). Accordingly, the frequency effect was presumed to reflect differences in the accessibility of LTM representations between HF and LF words, a property determined by relative use. This argument therefore assumed that the recall of an item was dependent solely on its frequency.

Walker and Hulme (1999) maintained that a late-stage redintegration explanation was not only a suitable framework for the frequency effect, but also the concreteness effect. However, the manner in which frequency (Hulme et al., 1997) and concreteness (Walker & Hulme, 1999) interacted with serial position was argued to be qualitatively different, as were performances on the backward recall task. These served as motivation to suggest that two redintegrative processes operate in serial recall; specifically, independent reconstructive processes would access separate phonological and semantic LTM systems at the point of recall.

Stuart and Hulme (2000) were the first to demonstrate the possibility of pre-existing associativity as an active influence in the development of the frequency effect in serial recall, and argued that a set of presented items might form a pool of activation that supports late-stage redintegration. Although disputed by others (Saint-Aubin & LeBlanc, 2005; Saint-Aubin & Poirier, 2005) this interpretation found support in the work of Hulme et al. (2003; Morin et al., 2006) who determined that in alternating lists HF and LF words are recalled at similar levels. The frequency effect in serial recall is therefore context dependent; how well an item is recalled is a function of the list in which it is presented. On the basis of the initial research this contextual influence was considered to be nondirectional, although Hulme et al. (2003) raised the possibility that an item-to-item associative effect would also explain their results. Allen and Hulme (2006) discarded a redintegration explanation of serial recall altogether, arguing that a unitary language-based model (N. Martin & Saffran, 1997) best accommodated the relationships between the effects of semantic representations on speech production and serial recall, and the association between frequency-based differences in speech input and serial recall found in their work.

Consequently, the presumed nature of the redintegrative contribution from LTM (Hulme et al., 2003; Morin et al., 2006; Stuart & Hulme, 2000, 2009) has undergone substantial revision in conceptual models of STM to the point of deletion in some instances. In its latest guise, redintegration is based on the development of an activated cohort that arises from activations of the lexical representations of items at presentation and spreading activation along the pathways connecting them (Stuart & Hulme, 2009). The strength of connections between items is a function of their co-occurrence in natural language. This activation net supports the reconstruction of items at recall. Importantly, this view places the locus of LTM interaction at the commencement of the serial recall task, thus abandoning a strictly two component view of redintegration, and bringing this approach more in line with language-based models (N. Martin & Saffran, 1997; R.C. Martin et al., 1999).

Language-based models of STM have developed from the convergence of the language processing and STM literatures (Allen & Hulme, 2006; Majerus, 2009; N. Martin & Saffran, 1992, 1997; R.C. Martin et al., 1999; Romani et al., 2008; Roodenrys et al., 2002). This development has been supported by the noted similarities in performance of neurological patients on language processing and STM tasks and evidence of dissociation between patients (e.g. Majerus, Van der Linden, Poncelet, & Metz-Lutz, 2004; N. Martin & Saffran, 1997; R.C. Martin et al., 1999), as well as observed sensitivity of STM performance to the manipulation of language representations (e.g. when serial recall of nonwords follows the incidental learning of an artificial grammar - Majerus, Van der Linden, Mulder, Meulemans, & Peters, 2004), and neuroimaging studies that identify brain regions associated with language processing are engaged during STM tasks (Cameron et al., 2005; Majerus, Poncelet, Van der Linden, et al., 2006; Ruchkin et al., 1999; Ruchkin et al., 2003). These models assume that multiple forms of representations (phonological, lexical and semantic) are activated from the point of encoding onwards and assist in the maintenance of information over the retention period through interactive activation. Therefore, they can more easily accommodate effects that implicate ongoing involvement of LTM variables in STM tasks. Within a language-based view, and in a manner similar to the most recent instantiation of redintegration (Stuart & Hulme, 2009), associativity would arise from activation that spreads along connections between lexical representations with the strength of activation influenced by the degree of association items have in natural language.

Experiments 1 and 2 of this thesis examined the relationship between frequency and concreteness in serial recall tasks that contained a factorial manipulation of these variables. An interaction was observed where the frequency effect decreased as the concreteness of items increased. In these experiments the frequency by serial position interaction and concreteness by serial position interaction were found to be similar. An order effect was found in one experiment, and although consistent with other experiments manipulating lexical-semantic variables that have found order effects, the effect size was small. Therefore, the point of influence of these variables on recall performance focuses on the retention of item information.

Experiments 3-6 directly addressed the nature of associativity underpinning the frequency effect in serial recall. It was found that in mixed lists with equal numbers of HF and LF words, HF words can be recalled better than LF words if (i) items of the same frequency type form a sequence within the list (e.g. HHHLLL vs LLLHHH, or HHLLLH vs LLHHHL), and (ii) these sequences occur in the early portion of the lists. It is proposed that item-to-item associativity is active for the early portion of the list, while nondirectional, list-level activation supports recall in the final list positions, and this support relates to the retention of item information.

16.3 The nature of LTM contribution

16.3.1 Late-stage redintegration

The results of Experiments 3-6 can be accommodated within a strictly late-stage interpretation of redintegration, that is, reconstruction based on the associative links between LTM representations at the point of item recall only (Hulme et al., 1997; Schweickert, 1993). The change from directional to nondirectional associative support across list positions might reflect the growing net of activation that results from the successive entry of degraded short-term traces to LTM. For early list positions supportive activation across all possible pairwise associations should be dominated by the connections between adjacent items. As recall progresses, the associative support matures to include all possible pairwise combinations between list items, and therefore becomes nondirectional.

A frequency by concreteness interaction can be explained by the dual-redintegration approach of Walker and Hulme (1999) if it assumed that HF and highly concrete items are more likely to be successfully reconstructed from both systems and leads to redundancy in item retrieval. However, this position runs into difficulty when the forms of the 'by position' interactions are considered. Experiments 1-2 did not find evidence of distinct signatures for each variable, and a review of the relevant literature caste doubt on the theoretical basis of the frequency by serial position interaction reported by Hulme et al. (1997) on which this assertion was founded. The effect in this case is more likely to be associated with stimulus set size than the frequency of items. Additionally, other data taken as evidence for a late-stage effect of lexical-semantic variables (e.g. matching span for concrete and abstract words, Walker & Hulme, 1999), has since been criticised for its lack of generalisability; whether or not lexical-semantic effects occur in serial recognition is determined by the strength of manipulation of item memory factors (Jefferies et al., 2006b; Romani et al., 2008).

16.3.2 Accounts involving LTM variables prior to retrieval

The most recent model of redintegration (Stuart & Hulme, 2009) and language-based models (N. Martin & Saffran, 1997; R.C. Martin et al., 1999) assume that LTM involvement, and therefore associative influence, is possible from the point of encoding in the serial recall task. Experiments 1-2 investigated the behaviour of frequency with a second lexical semantic variable, word concreteness in serial recall. Therefore, from the perspective of language-based models, these experiments examined the relationship between lexical and semantic representations within the language processor. Consistent with the interaction of LTM representations outlined in language-based models, these experiments confirmed an interaction between frequency and concreteness. Lexical nodes that are more strongly activated (in this case nodes corresponding to HF words) will return greater levels of feedback activation to the connected nodes at the phonological level.[1] The degree of item concreteness indexes the number of semantic nodes that connect with a lexical node (R.C. Martin & Lesch, 1996). The activation of a node associated with a semantic feature produces feedback activation to those lexical nodes linked with the semantic feature, including any semantic competitors at the lexical level. Accordingly, items with more semantic features will be advantaged by interactive activation within the language processor, promoting the retention of the correct phonological representations in LTM. The nature of the interaction found in these experiments, namely a reduction in the frequency effect for high concreteness words suggests that the level of activation of item nodes has an upper limit, and the theoretical level of activation achievable with HF, high concreteness items surpasses this limit. Furthermore, if the effects of frequency and concreteness combine within a single interactive system, similarity between their interactions with serial position would be anticipated, and this was demonstrated, most clearly in Experiment 2.

However, in light of the results of Experiments 3-6, the challenge for these types of models is to identify how positional effects in recall might interact with the maturation of associativity observed as recall proceeds. As specified, these models would predict that the supportive activation of items at the commencement of recall should be nondirectional, as the activations from each list item would have contributed to the associative net at this point in the task. Therefore, these explanations are insufficient as currently instantiated to identify a basis for the evolution of associativity as a list of items is recalled in order.

Experiments 3-6 suggest that the nature of activity in the language processor is complex, is dynamic over the course of the serial recall task, and may reflect the interaction between item and order memory mechanisms. The recall of early list items suggests that variation in the integrity of item information arises from activation

[1] Note that in pure lists the locus of effect of associativity in word frequency (directional or nondirectional) is undetectable because the strength of association between adjacent items and the pairwise average across the list set should be equivalent. Therefore all lexical nodes corresponding to items in pure HF lists should receive higher activation than nodes that correspond to items in pure LF lists.

levels determined by the strength of association between the lexical representations of adjacent items in the language processor. The level of co-occurrence in natural language between items determines the degree to which lexical nodes will supply feedback activation to associated phonological nodes. In contrast, the retention of item information for late list items appears to be supported by nondirectional activation drawn from the entire set of items in the list, as a function of pairwise co-occurrence (Hulme et al., 2003; Stuart & Hulme, 2000, 2009). This activation (both directional and nondirectional) is presumed to travel along pathways between lexical representations in LTM (Stuart & Hulme, 2009).

This pattern of activation suggests that the order of items either maintains or reinstates, prior to recall, activation within the item information system that preserves directional associativity for early but not late list items. If this is the case, then closer examination of the possible sources of refreshment within the item information system might lead to a better understanding of how this effect arises. Among these include the effect of rehearsal on the maintenance of the sequence of memory events forming the to-be-remembered list, effects of speech production at the point of recall, and output interference and cueing effects during recall.

The hybrid model of Majerus (2009) offers a framework in which to position these potential effects as it includes separate item and order memory mechanisms that interact in the execution of verbal STM tasks. While speech production effects would be presumed to act from within the item memory system, rehearsal and output interference influences would be a result of the order memory system interacting with item memory in the service of maintaining and recalling a sequence of items.

16.4 Recommendations for future investigation: The formation of directional and nondirectional associativity in serial recall

16.4.1 A role for rehearsal?

Participants often rehearse items in a cumulative forward-order manner during the presentation of items in early list positions (Bhatarah, Ward, Smith, & Hayes, 2009; Page & Norris, 1998; Tan & Ward, 2008). This differential treatment across serial positions suggests that rehearsal might play some part in the evolution of inter-item associativity in serial recall. This section examines recent research investigating the effects of rehearsal on serial recall and its potential role in STM and memory more generally.

A line of investigation exploring similarity of processes between serial and free recall from the perspective of episodic memory has been conducted by Ward and colleagues (Bhatarah et al., 2009; Bhatarah, Ward, & Tan, 2006; 2008; Tan & Ward, 2007, 2008; Ward et al., 2009). These researchers have focused on the relationships between item encoding, rehearsal strategy, output order (in terms of task requirement) and the resultant features in serial position curves. Like others (e.g. G.D.A. Brown et al., 2007; Howard & Kahana, 2002a, 2002b), they assume a recency-based approach to memory; the relative differences between primacy and recency effects in free and serial recall are a function of the same recency-sensitive mechanisms interacting with the degree of constraint on the order in which items are recalled. They also suggest, as do others (Howard & Kahana, 2002a, 2002b), that these general memory mechanisms possess a tendency to engage in forward ordered recall, as is observed in free recall, and as is demanded in serial recall tasks (Bhatarah et al., 2008).

Accordingly, these researchers argue that the encoding of items and the rehearsal strategies used in free and serial recall are indifferent, and task demands determine variations in the likelihood that items from different portions of the serial position curve are recalled. This position is in opposition to the stance taken by Baddeley and Hitch (1974; Baddeley & Hitch, 1999) on the basis of their seminal findings from concurrent task experiments (see also Bhatarah et al., 2006). These revealed that a free recall recency effect in proportion to the total level of recall persists even when a concurrent 6-digit serial recall task is inserted between the presentation of free recall items, and inserted after presentation of the last item. It was argued that the capacity responsible for the recency effect in free recall could not be the same as the short-term facility responsible for the retention of the digits in the concurrent task (Baddeley & Hitch, 1974).

This problem was pursued further by Bhatarah et al. (2008) who examined whether the encoding of information was likely to be different for the two tasks. They compared free and serial recall under pre- and post- cueing conditions. In pre-cueing conditions, participants knew prior to list presentation the nature of test, while post-cueing conditions advised participants whether they were to recall the list according to free or serial recall instruction after presentation. Differences between the cueing conditions therefore would indicate if task-specific encoding occurred in normal pre-cued activity.

In addition, these tasks were performed on list lengths more suitable to serial recall (8 items). For each condition Bhatarah et al. (2008) conducted conditional response probability analysis (lag-CRP, Howard & Kahana, 1999, 2002a, 2002b; Kahana, 1996) on the entire range of lags, avoiding the incompleteness of earlier analyses applied to longer lists (Farrell & Lewandowsky, 2008). Bhatarah et al. (2008) found that the serial position curves for free and serial recall exhibited the respective features common to curves observed with these tasks (the typical U-shape for free recall and extended primacy with limited recency effects for serial recall). Differences between serial position curves generated under pre- and post-cue conditions were minimal. These results suggest that differences in encoding for the two tasks are unlikely. Furthermore, forward serial recall was observed to be a prevailing feature in both free and serial recall conditions suggesting common underlying mechanisms.

Another point of comparison between free and serial recall tasks involves rehearsal patterns (Bhatarah et al., 2009). Using overt rehearsals, the authors tracked the strategies participants used in rehearsing to-be-remembered material from eight-item lists, in free or serial recall. They explored whether rehearsal strategy changed with test expectancy and in separate experiments examined how variables argued to moderate rehearsal, namely presentation rate and word length, and articulatory suppression and word length, impacted recall in both forms of task. Lastly, they tested whether the word length effects observed with eight-item lists were generalisable to six and twelve item lists.

Bhatarah et al. (2009) found that rehearsal patterns were similar for free and serial recall; items in the first half of the list were more likely to be rehearsed than later list items. Rehearsal patterns were also similar across the manipulation of test expectancy and indicated there was no strategic shift in rehearsal dependent upon whether participants had knowledge of the type of recall to be performed or not. Rehearsals decreased in a parallel fashion across the two tasks when items possessed greater word length, or when the items were presented at a fast rate. Taken together, the authors argued that these similarities were consistent with the engagement of a common rehearsal process in free and serial recall.

Despite fundamental differences in the respective serial position curves generated by free and serial tasks, there were similar effects on recall when word length was manipulated, and these generalised across list lengths of six, eight, and twelve items (Bhatarah et al., 2009). The variation of presentation rates was found to impact free and serial recall, although presentation rate interacted with different variables in each task. In free recall, the effect of serial position but not word length was dependent on presentation rate, while in serial recall the effect of word length but not serial position was a function of presentation rate. These were argued to be a consequence of the different output orders arising with each task. Articulatory suppression reduced, but did not eliminate entirely, word length effects in both tasks. Conditions that reduced the opportunity to rehearse, namely faster presentation rates, greater word length and articulatory suppression were found to suppress primacy effects in both tasks.

These results were interpreted by Bhatarah et al. (2009) to be consistent with the view that the role of rehearsal is to maintain the accessibility of as many items as possible at the point of recall, and that this function is common between serial and free recall. Furthermore, recall in both tasks is governed by the same mechanisms.

Accordingly, Bhatarah et al. (2009) endorse a unitary approach to memory where experienced events are positioned on an episodic continuum in a chronological order that extends from the most to the least recently experienced events. Within this continuum, the presentation of items and their rehearsals constitute individual events, and the act of rehearsal re-instantiates an item to an event closer to the point of test. This maintains an item’s accessibility, leading to the possibility of further rehearsal, and the greater likelihood of recall.

The respective forms of the serial position curves are a function of the relative accessibility of items in combination with task requirements. Strong recency effects are observed in free recall because in the absence of any order constraint, individuals will output those items that are the most readily accessible and these will tend to be items presented at the end of list. Items from the start of list that have been rehearsed as presentation proceeds will be the next most accessible producing the smaller primacy effect in free recall. The remainder of the items, recalled from events receding on the continuum because of the recall of items, will be recalled the least well.

In contrast, the extended primacy effect found in serial recall is a consequence of the requirement to reproduce the item sequence in presentation order. Typically, early list items have been rehearsed more often and rehearsed during the interstimulus intervals in the first half of the list, placing associated events closer to the point of test, i.e. making them more accessible than medial list items. As recall proceeds across from the first to the medial serial positions, performance decreases and recovers slightly for the final item, as it benefits in a relative sense from being the most recent event prior to the commencement of recall.

Consistent with this explanation, conditions that impact the degree to which rehearsal can occur will harm the recall of early list items in free and serial recall tasks when compared with control conditions. Conversely, late list items are immune from such manipulations, as are control conditions that do not readily facilitate rehearsal (e.g. lists of long words presented with and without articulatory suppression) (Bhatarah et al., 2009).

In the current work Experiments 4-6 identified that directional associativity, dependent on the level of co-occurrence in natural language as indexed by word frequency, exists for early but not late serial positions; the recency portion of the serial position curve exhibited recall consistent with nondirectional associative support for items. Given the tendency for participants to engage in cumulative forward-order rehearsal for early list items (Bhatarah et al., 2009; Page & Norris, 1998; Tan & Ward, 2008) a role for rehearsal in the development of inter-item associativity in serial recall is possible. In addition, given the evidence for a common rehearsal mechanism and comparable effects on serial position curves when variables influencing rehearsal are manipulated in free and serial recall (Bhatarah et al., 2008; Bhatarah et al., 2009; Ward et al. 2009), rehearsal may underlie the observed tendency for forward order recall to vary according to inter-item associativity in free recall (Howard & Kahana, 2002b) and form the link between serial and free recall referred to by Hulme et al. (2003).

Tan and Ward (2008) also used overt rehearsal to show that increased cumulative forward-order rehearsals were associated with the longer presentation rates. If rehearsal has some relationship to the formation of directional associative effects in serial recall, specifically in terms of how far into the list this form of associativity penetrates, this should be identified by altering the presentation rate in Experiments 3-6. Alternatively, varying the rehearsal of participants (fixed versus cumulative strategies) and verifying these patterns with overt rehearsals should result in an alteration of the degree of directional associativity in recall, if rehearsal is an underlying factor in the development of associative influences within the frequency effect.

16.4.2 Influences from speech production

Another issue raised by the current work involves the identification of the point (or points) at which associative effects are active within the language processor. As highlighted in Chapter 7, recent studies (e.g. Jefferies et al., 2006b; Romani et al., 2008) have established that lexical semantic variables influence how well items are remembered prior to overt output. Experiments 4-6 raise the possibility that the pattern of activation generated by the presentation of items might also be more complex than previously assumed and not of a form where all items are supported to an equal extent at the completion of list presentation. Given the elapsed time between presentation and recall the prospect that this activation retains directional sensitivity is questionable, although as discussed in the previous section rehearsal may be a means to refresh associative links between adjacent items for early list positions. Alternatively, if activation from the input of items into the language processor results in a mutually supportive network of list items, then how the patterns of associativity observed in recall arise demands some resolution.

Most language-based models assume separate, though interconnected, input and output pathways (Allen & Hulme, 2006; N. Martin & Saffran, 1997; R.C. Martin et al., 1999; Monsell, 1987). Allen and Hulme (2006) determined that serial recall performance was strongly associated with performance in speech production, implicating the speech output pathway as a critical contributor to the successful output of items. Accordingly, the act of producing an item might serve to create a reinstantiation of the initial pattern of activation within the speech motor processes and underpin the development of directional through to nondirectional associativity as recall proceeds across list positions. This possibility is similar to the multiple processes point of view espoused by Thorn and colleagues (Thorn et al., 2005; Thorn et al., 2009, see section 7.6.2).

A direct comparison between serial recall performance and performance on a second task minimising output requirements (e.g. serial recognition) should ascertain whether the patterns of associativity across serial positions are dependent on task requirement. If they were observed to be similar, this would suggest that directional associativity for early list items is preserved within the language processor, and in particular the input pathway, and point to the existence of mechanisms that could reproduce such a pattern (e.g. rehearsal). If the pattern of associativity observed for serial recognition was found to be different to the pattern obtained with serial recall, then output-specific processes would be implicated. As the HF and LF words used in the current experiments were drawn from open sets, it would be anticipated that serial recognition experiments based on these stimuli would produce a strong manipulation of item information, and therefore provide an appropriate test of lexical-semantic factors prior to recall.

16.4.3 Effects of recall processes

Output interference has been identified as one source of the primacy effect in serial recall, where recall of an early list item is seen to reduce the likelihood of recall of items later in the list (Cowan et al., 2002; Nairne, 1990; Oberauer, 2003; Tan & Ward, 2007), although this form of interference is more prevalent for visually presented material (Cowan et al., 2002). In addition, Nairne et al. (2007) conducted experiments that examined the effects of output position and prior recall of another list item on the recall of a target item. These experiments also controlled the time at which recall of the target item occurred and tested whether there was an effect of response set (the number of remaining list items to be recalled). They found that, relative to the recall of the target item alone, the recall of the target item was enhanced when the first item preceded the target item by one position, while recall of the target was harmed when the first item belonged to the serial position subsequent to the target [2], although there was a suggestion that forward initial recalls produced an impairment in the recency portion of the curve. Furthermore the authors found that this effect was likely to be due to the mere presentation of the initial item before the target item (tested in one condition by requesting participants copy the initial item down) rather than its recall *per se*. Accordingly, Nairne et al. (2007) argued that the

[2] This associative asymmetry has been reported before in free recall (Howard & Kahana, 2002a, 2002b; Kahana, 1996) and in the serial recall of longer lists (Kahana & Caplan, 2002).

act of recalling an item is not the defining element in providing an effective cue for a target, but its incidence within the retention interval.

In these experiments output interference effects were observed as the recall of a target item alone was greater than recall of the target after an initial item (Nairne et al., 2007). However output interference was seen to vary across serial positions, with larger effects observed for later list items.

The directionally sensitive effects of associativity seen in Experiments 4-6 occur for early rather than late list items, and in these experiments input and output positions are confounded. It is possible that the recall process is responsible for the shift from directional to non-directional associativity for late list items. As items are recalled and interference is created for the next item in sequence, direct associations might become less useful, and nondirectional associative activation may be relied upon to support recall in the final instances. Alternatively, it could be that the potential for an item to act as a cue for recall of the next item changes across serial positions (Nairne et al., 2007), and the nature of this change is frequency dependent.

The possibility that output interference impacts the pattern of associativity found in the mixed frequency experiments could be examined by conducting probed recall experiments, where a probe consisting of a list item cues recall of the next item in the list (Kahana & Caplan, 2002). This task eliminates any effects of overt output prior to the target item and would indicate whether the physical production of speech affects the evolution of directional to nondirectional associative effects in serial recall. Alternatively, a method similar to that used by Cowan et al. (2002) could be considered, where participants are required to produce whole list or partial list reports that commence at different points within the list. Should the positioning of directional and nondirectional associative effects alter with the point of recall in the case of probed recall, or the position where serial recall commences with the Cowan et al.'s (2002) method, this would indicate that output processes are influential in determining the nature of associativity across serial positions. A comparison of serial recall performance with the free position recall performance (Crowder, 1969; Tan & Ward, 2007) could highlight any impacts made by the imposition of recall order. Free position recall is a relaxed form of ordered recall where participants can recall items in any order but must nominate the position of the item, and therefore would identify the recall features of the items most accessible at the point of test (typically, recency items). Should directional associative effects be available at the commencement of recall for late list items it would be anticipated that this comparison would detect these effects.[3] Lastly, performances on auditory and visual presentation modalities could be contrasted across repetitions of Experiments 4-6 to determine whether associative effects are sensitive to the degree to which item encoding is resistant to output interference (Cowan et al., 2002).

16.5 Further investigation within the item memory system

The following sections consider how the current research could be extended to explore some features of serial recall formerly examined by other researchers but not directly addressed by the experiments presented in this thesis.

16.5.1 Phoneme binding in lists of mixed frequency

Jefferies and colleagues (Jefferies et al., 2006a, 2006b; Jefferies et al., 2009; Jefferies et al., 2004, 2005) have examined the role that lexical-semantic variables play in promoting the binding or coherence of the phonological traces of items within linguistic LTM (Knott et al., 1997; Patterson et al., 1994), and argue that the pattern of loss of coherence observed in neuropsychological patients with impoverished semantic knowledge (Jefferies et al., 2004, 2005; Knott et al., 1997; Patterson et al., 1994) can be generated in healthy participants when the semantic content of items is varied (Jefferies et al., 2006a). They found in normal samples that HF words promoted greater coherence of the phonological trace than LF words, and high imageability words promoted greater coherence than low imageability words (Jefferies et al., 2006a). Hoffman, Jefferies, Ehsan, Jones, and Lambon Ralph (2009; Jefferies et al., 2006a) have argued that semantic binding is the basis upon which better item memory is attained.

In the semantic binding account of STM the phonological representations of items are activated on presentation (Hoffman et al., 2009). The constituent phonological elements of a word are strongly associated with each other because they are consistently co-activated when a word is perceived or produced. Unlike interactive activation accounts (N. Martin & Saffran, 1997; R.C. Martin et al., 1999; Romani et

[3] Ward et al. (2003) examined the free recall of mixed frequency lists, but these were pseudo-random in terms of item arrangement and much longer (20 items).

al., 2008) however, semantic representations are also assumed to be activated at presentation; it is spreading activation between semantics and phonology that constrains or binds the phonological elements in their correct sequence.

Jefferies et al. (2006a) investigated phonological coherence in the contexts of pure (words) and mixed (words and nonwords) lists. Although Jefferies et al. (2006a) manipulated frequency and imageability in both of their experiments they did not report any analyses indicating that they examined this interaction. Hoffman et al. (2009) combined the presentation of pure and mixed lists within the same experiment and tested healthy controls and six semantic dementia patients. For item recall they reported a frequency by imageability interaction with list type, where pure lists exhibited a greater imageability effect for HF than LF lists, but mixed lists produced a smaller imageability effect for HF than LF lists. The result for pure lists stands in contrast to the outcomes reported in Experiments 1-2, however it was derived from the total sample that included the semantic dementia patients, and is therefore not a direct comparison with the current studies. It would be anticipated on the basis of findings from Experiments 1-2 that a replication of their study with healthy participants alone would find a frequency effect in item recall that diminishes for high imageability words.

In pure lists of words, phoneme migrations occur more for LF than HF items (Hoffman et al., 2009; Jefferies et al., 2006a), and occur more for low imageability than high imageability items (Hoffman et al., 2009). However, to date, an interaction between frequency and imageability in the rate of phoneme migrations with pure lists has not been reported; it is possible that this is due to the difficulties in generating sufficient levels of migration to produce statistically robust results in contexts where phonological coherence is relatively intact.

An alternative approach to using mixed lists of words and nonwords to compromise the semantic content of the list, involves the use of pure word lists presented in noise (Surprenant, 1999). Degradation of input would increase the difficulty of processing and make phonological coherence from additional sources (i.e. lexical-semantic properties) more important. Increased opportunity for phoneme migration would make any variations of rates by stimulus set more explicit and be more likely to provide conditions that determine whether the effects observed at the level of correct recall translate to phoneme movement within the trace.

The experiments reported in this thesis do not test the semantic binding hypothesis as the stimuli were not constructed to control for the presentation of phonemes within a trial; accordingly analysis of phoneme migrations in the data from Experiments 3-6 would encounter interpretation difficulties. However, if Experiments 3-6 were replicated under conditions where the presentation of phonemes was controlled and these included a factor of signal quality, it would be anticipated that the coherence of mixed lists should be intermediate with respect to the coherence of pure HF and pure LF lists respectively. Secondly, the coherence of HF and LF items should be different in pure and mixed lists, as how well items in a list remain bound is a function of the binding capacity of other items in the list. Lastly, on the basis of correct recall determined in the current work, it would be useful to test whether coherence effects can be directional in nature and therefore dependent upon the sequencing of items.[4] In addition, an experiment using the pure list stimuli of Jefferies et al. (2006a) and varying signal quality could investigate whether the interaction of frequency and concreteness found in serial recall is reflected in the stability of the phonological trace at the phoneme level.

16.5.1.1 Speech production and semantic binding

Jefferies et al. (2009) raised the possibility that recall might be influenced by an additional phoneme binding effect that occurs during overt output. They argued that the act of articulating an item necessarily identifies its constituent phonemes, and this could influence the recall of later items. Within the context of mixed frequency lists this could be tested by comparing the nature of effects observed for both serial recall and serial recognition using suitably controlled stimulus sets (see section 16.5.1). Therefore, if differences existed between the effects observed for each task, a second process involving overt output, and possibly relating to associations between items in terms of speech programming, might be active.

[4] As currently instantiated (e.g. Hoffman et al., 2009), semantic binding is assumed to operate at the list level and therefore the presence of directional effects would challenge this position. Results with mixed lists of words and nonwords supported the concept of semantic binding at the list level (Jefferies et al., 2009), but the sequencing of the same item type was limited to two items.

16.5.2 Coarticulatory influences in lists of mixed frequency

Woodward et al. (2008) proposed that the frequency effect could be explained by the differences in the coarticulation of word boundaries between HF and LF words. In this sense, differences in recall amount to variation in the efficiencies of the assemblage of articulatory gestures of items in rehearsal and recall. Post-hoc analyses of the stimulus sets in Experiments 3-6 suggested that there was no apparent bias in the co-articulation of word boundaries for HF and LF sets. However, the recall of items in early list positions did conform to a pattern anticipated by the coarticulation hypothesis. To more fully investigate this proposal it is recommended, as discussed in section 15.4, that the experiments performed with these stimulus sets be repeated with articulatory suppression and use written recall. These design features should disambiguate whether the sequencing of articulatory gestures, particularly at word boundaries makes any sizeable contribution to serial recall when the familiarity of items is manipulated.

16.6 Directional and nondirectional associativity … not the whole story

Hulme et al. (2003; Stuart & Hulme, 2000) had identified that list composition was an important factor in determining the level of performance that recall of a HF or LF item could achieve. However, their manipulation of composition was constrained to mixed lists with balanced numbers of HF and LF words[5]. Saint-Aubin and LeBlanc (2005, see also Roodenrys, unpublished – section 6.3.5) identified that isolate effects occurred when a single HF or LF item was inserted into a pure list of the other kind; HF words were recalled better than LF words in pure lists, but not as well as HF words in pure lists, while LF isolates were recalled as well as HF words in pure lists. Saint-Aubin and LeBlanc (2005) argued for a distinctiveness plus processing account to explain their results.

It is worth considering whether these effects might also be explained by a combination of item-to-item and nondirectional list-level associativity. The results outlined by Saint-Aubin and Poirier (2005) could be accommodated if it was assumed that recall was influenced by whatever form of associativity provided the greater level

[5] As is the case with the present thesis.

of support. Specifically, the higher nondirectional associativity of the list containing the LF isolate could offset the intermediate item-to-item associativity contribution from the transition between HF and LF words. The better recall of the HF isolate could reflect the absence of strong nondirectional support to supplement the item-to-item associative contribution.

However, this explanation neglects the apparent regions of influence for inter-item and nondirectional associativity as identified in Experiments 3-6; the LF isolate effects (i.e. poorer recall for LF isolates) were not present early in the list and HF isolate effects were present in the later list positions. Clearly the failure to find a set of mechanisms (associative- or distinctiveness-based) that can explain the frequency effect across differing list compositions (pure, mixed, and isolate) underscores its intricate nature.

Consequently, to answer whether there is a discrete or continuous change in the frequency effect from isolate conditions through to pure lists, and to identify what type of additional mechanism(s) might be responsible, future research needs to examine the effect of compositional change in a more continuous fashion. This could involve an extension to the design of Saint-Aubin and LeBlanc (2005) to include not only single isolates but longer sequences inserted into the body of lists.

16.7 Conclusion

Hulme et al.’s (1997) assertion that the frequency effect in STM was complex and not well understood remains an accurate one. The experiments in this thesis have raised as many questions as they may have hoped to resolve, in particular with respect to the influences that operate between list items in serial recall. The nature of recall observed with mixed lists reveals that performance cannot be explained by ‘settled’ activation patterns within the language processor, and demand further examination to determine how the effects of associations between items change as the serial recall task is executed. Speculatively, this raises the possibility that the frequency effect in STM is influenced by more general memory mechanisms, for example rehearsal and/or recall processes, that dictate how word frequency influences the recall of items as recall proceeds across list positions. Such possibilities may lead to a better understanding of how linguistic information more typically interacts with order memory.

In the course of study of the frequency effect in serial recall, it is apparent that old elements are considered in new ways as a means to explain the latest patterns in experimental data. Pre-existing associations appear to matter (Hulme et al, 2003; Morin et al., 2006; Stuart & Hulme, 2000, 2009), but not in the manner predicted by Underwood and Postman (1960); rehearsal may yet have an influential role (Bhatarah et al., 2009; Tan & Ward, 2008; Woodward et al., 2008) but not as conceived by Baddeley et al. (1975; Baddeley, 1986; Baddeley & Hitch, 1974); and LTM representations have emerged as fundamental to the operation of STM (Hulme et al., 1997; Hulme et al., 2003; Stuart & Hulme, 2000, 2009) but are likely to be activated well before the 'point of retrieval' nominated by late-stage theorists (Allen & Hulme, 2006; Majerus, 2009; Romani et al., 2008). More generally, the concept of a short-term storage system for the retention of speech-based material (Baddeley, 1986; Baddeley & Hitch, 1974) has evolved into a multi-representational scheme based on LTM knowledge interacting with a non-linguistic sequencing mechanism that has particular utility for memory over the short-term (Majerus, 2009). This latter formulation contains similarities with unitary accounts of memory (e.g. the form of order mechanism in SIMPLE, G.D.A. Brown et al., 2007, or the recency-based approach of Ward et al., 2009), and offers the possibility of a more unified account between general memory and psycholinguistic approaches to STM phenomena.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64-88.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighbourhood conflicts. *Psychonomic Bulletin & Review, 4,* 439-461.

Altarriba, J., Bauer, L.M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31*, 578-602.

Altmann, E.M. (2009). Evidence for temporal decay in short-term episodic memory. *Trends in Cognitive Sciences*, *13*, 279.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2., pp 89-195). New York: Academic Press.

Atkinson, R.C., & Shiffrin, R.M. (1971). The control of short-term memory. *Scientific American*, *225*, 82-90.

Baayen, R.H., Piepenbrock, R., & Van Rijan, H. (1993). *The Celex Lexical Database*. Philadelphia: Linguistic data consortium, University of Pennsylvania.

Baddeley, A.D. (1966a). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*, 302-309.

Baddeley, A.D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*, 362-365.

Baddeley, A.D. (1972). Retrieval rules and semantic coding in short-term memory. *Psychological Bulletin, 78*, 379-385.

Baddeley, A.D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A.D. (1997). *Human memory: Theory and practice* (Revised edition). Sussex: Psychology Press Ltd.

Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A.D. (2002). Is working memory still working? *European Psychologist, 7*, 85-97.

Baddeley, A.D. (2007). *Working memory, thought, and action*. New York: Oxford University Press.

Baddeley, A.D., & Ecob, J.R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature*, *227*, 288-289.

Baddeley, A.D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8., pp 47-90). New York: Academic Press.

Baddeley, A.D., & Hitch, G. (1994). Developments in the concept of working memory. *Neuropsychology, 8,* 485-493.

Baddeley, A.D., & Levy, B.A. (1971). Semantic coding and short-term memory. *Journal of Experimental Psychology, 89*, 132-136.

Baddeley, A.D., & Scott, D. (1971). Word frequency and the unit-sequence interference hypothesis in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*, 35-40.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14,* 575-589.

Balota, D.A., & Neely, J. (1980). Test expectancy and word frequency effects in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6,* 576-587.

Bhatarah, P., Ward, G, Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition, 37*, 689-713.

Bhatarah, P., Ward, G., & Tan, L. (2006). Examining the relationship between free recall and immediate serial recall: The effect of concurrent task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 215-229.

Bhatarah, P., Ward, G., & Tan, L. (2008). Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. *Memory & Cognition, 36*, 20-34.

Bird, K., Hadzi-Pavlovic, D., & Isaac, A. (2000). *PSY*. Sydney: University of New South Wales.

Bireta, T.J., Neath, I., & Surprenant, A.M. (2006). The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review, 13*, 434-438.

Bradley, M.M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective rating.* Technical Report C-1. The NIMH Center for Emotion and Attention, University of Florida.

Brener, R. (1940). An experimental investigation of memory span. 467-482. *Journal of Experimental Psychology, 26*, 467-482.

Broadbent, D.E. (1987). *Perception and communication.* New York: Oxford University Press.

Brown, G.D.A. (1984). A frequency count of 190,000 words of the London-Lund corpus of English conversation. *Behavior Research Methods, Instruments, & Computers, 16*, 502-532.

Brown, G.D.A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language*, *34*, 594-621.

Brown, G.D.A., Neath, I., & Chater, N. (2007). A Temporal Ratio Model of Memory. *Psychological Review, 114*, 539-576.

Brown, G.D.A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127-181.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12-21.

Brown, J. (1964). Short-term memory. *British Medical Bulletin, 20*, 8-11.

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

Burgess, N., & Hitch, G.J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551-581.

Cairo, T.A., Liddle, P.F., Woodward, T.S., & Ngan, E.T. (2004). The influence of memory load on phase specific patterns of cortical activity. *Cognitive Brain Research, 21*, 377-387.

Cameron, K.A., Haarmann, H.J., Grafman, J., & Ruchkin, D.S. (2005). Long-term memory is the representational basis for semantic verbal short-term memory. *Psychophysiology, 42*, 643-653.

Caza, N., & Belleville, S. (1999). Semantic contribution to immediate serial recall using an unlimited set of items: Evidence for a multi-capacity view of short-term memory. In I. Neath, G.D.A. Brown, M. Poirier, & C. Fortin (Eds), *Short-term/working memory: A special issue of the International Journal of Psychology* (pp. 334-338).

Caza, N., Belleville, S., & Gilbert, B. (1999). How loss of meaning with preservation of phonological word form affects immediate serial recall performance: A linguistic account. *Neurocase, 8*, 255-273.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel word lists. *Memory & Cognition, 21*, 539-545.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33A*, 497-505. Retrieved from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75-84.

Conrad, R., & Hull, A.J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology, 55*, 429-432.

Cowan, N. (1992). Memory span and the timing of spoken recall. *Journal of Memory and Language*, *31*, 668-684.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2005). *Working memory capacity*. NewYork: Psychology Press.

Cowan, N., Baddeley, A.D., Elliott, E.M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review*, *10*, 74-79.

Cowan, N., Day, L., Saults, S., Keller, T.A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory.

Journal of Memory and Language, 31, 1-17.

Cowan, N., Saults, J.S., Elliott, E.M., & Moreno, M.V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*, 153-177.

Crowder, R.G. (1969). Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 524-528.

Davenport, M., & Hannahs, S.J. (1998). *Introducing phonetics and phonology.* Great Britain: Arnold.

DeCara, B. & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers, 34*, 416-423.

DeLosh, E.L., & McDaniel, M.A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1136-1146.

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports, 5*, 305-312.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association, *Psychological Reports, 7*, 337-344.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287-314.

Dent, K. Johnston, R.A., & Humphreys, G.W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 282-301.

Dewhurst, S.A., Brandt, K.R., & Sharp, M.S. (2004). Intention to learn influences the word frequency effect in recall but not in recognition memory. *Memory & Cognition, 32*, 1316-1325.

Diana, R.A., & Reder, L.M. (2006). The how-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*, 805-815.

Drewnowski, A., & Murdock, B.B. Jr. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 319-332.

Ebbinghaus, H. (1913). *Memory: a contribution to experimental psychology.* New York, NY: Teachers College Press. Electronic reproduction. Washington, D.C.: American Psychological Association, 2004. Available via the World Wide Web.

Engle, R.W., Nations, J.K., & Cantor, J. (1990). Is “working memory capacity” just another name for word knowledge? *Journal of Educational Psychology, 82*, 799-804.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19*, 624-634.

Fallon, A. B., Groves, K. & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task:When CAT helps RAT, but not MAN. *International Journal of Psychology, 34,* 301-307.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59-79.

Farrell, S., & Lewandowsky, S. (2008). Empirical and theoretical limits on lag recency in free recall. *Psychonomic Bulletin & Review*, *15,* 1236-1250.

Forde, E.M.E., & Humphreys, G.W. (2002). The role of semantic knowledge in short-term memory. *Neurocase*, *8*, 13-27.

Forster, K. I., & Forster J. C. (1999). *DMDX [computer software]*. Tucson: University of Arizona.

Freedman, J.L., & Loftus, E. F. (1971). Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 107-115.

Frick, R. (1995). On accepting the null hypothesis. *Memory & Cognition*, *23*, 132-138.

Gathercole, S.E., Frankish, C., Pickering, S.J., & Peaker, S.M. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 84-95.

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

Geiger, S., & Lewandowsky, S. (2008). Temporal isolation does not facilitate forward serial recall- or does it? *Memory & Cognition, 36*, 957-967.

Glanzer, M., & Adams, J.K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 5-16.

Geffen, G., & Luszcz, M.A. (1983). Are the spoken durations of rare words longer than those of common words? *Memory & Cognition*, *11*, 13-15.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183-216). Great Britain: John Wiley & Sons.

Gregg, V., Freedman, C.M, & Smith, D.K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80*, 363-374.

Gregg, V.H, Montgomery, D.C., & Castano, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior, 19*, 240-245.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267-333.

Hamilton, A.C., & Martin, R.C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective and Behavioral Neuroscience, 5*, 1-13.

Hamilton, A.C., & Martin, R.C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex, 43*, 112-123.

Harley, T.A., & Bown, H.E. (1998). What causes tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89*, 151-174.

Hebb, D. O. (1961). The organization of behaviour. New York: John Wiley & Sons, Inc.

Henson, R.N.A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73-137.

Henson, R.N.A., Norris, D.G., Page, M.P.A., & Baddeley, A.D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology, 49A*, 80-115.

Hoffman, P., Jefferies, E., Ehsan, S., Jones, R.W., Lambon Ralph, M.A. (2009). Semantic memory is key to binding phonology: Converging evidence from immediate serial recall in semantic dementia and healthy participants. *Neuropsychologia, 47*, 747-760.

Howard, M.W., & Kahana, M.J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 923-941.

Howard, M.W., & Kahana, M.J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269-299.

Howard, M.W., & Kahana, M.J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*, 85-98.

Howes, D. H. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29*, 296-305.

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 447-463.

Hulme, C., Maughan, S., & Brown, G.B.A. (1991). Memory for familiar and unfamiliar: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685-701.

Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A.M., & Brown, G.D.A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 586-594.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527-536.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G.D.A, Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1217-1232.

Hulme, C., Stuart, G., Brown, G.D.A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*, 500-518.

Hulme, C., Surprenant, A.M., Bireta, T.J., Stuart, G., & Neath, I. (2004). Abolishing the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 98-106.

James, W. (1950). *The Principles of Psychology*. (Vol. 1). USA: Dover Publications Inc.

Jefferies, E., Frankish, C., & Lambon Ralph, M.A. (2006a). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, *54,* 81-98.

Jefferies, E., Frankish, C., & Lambon Ralph, M.A. (2006b). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *The Quarterly Journal of Experimental Psychology, 59*, 949-964.

Jefferies, E., Frankish, C., & Noble, K. (2009). Lexical coherence in short-term memory: Strategic reconstruction or semantic glue? *The Quarterly Journal of Experimental Psychology, 62*, 1967-1982.

Jefferies, E., Jones, R.W., Bateman, D., & Lambon Ralph, M.A. (2004). When does word meaning affect serial recall in semantic dementia? *Cognitive, Affective and Behavioral Neuroscience, 4*, 20-42.

Jefferies, E., Jones, R.W., Bateman, D. & Lambon Ralph, M.A. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology, 22*, 183-212.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, *24*, 1-19.

Kahana, M.J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*, 103-109.

Kahana, M.J., & Caplan, J.B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition, 30*, 841-849.

Kausler, D.H., & Puckett, J.M. (1979). Effects of word frequency on adult age differences in word memory span. *Experimental Ageing Research, 5*, 161-169.

Keppel, G., & Underwood, B.J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153-161.

Kintsch, W., & Buschke, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology, 80*, 403-407.

Klein, K.A., Addis, K.M., & Kahana, M.J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition, 33*, 833-839.

Knott, R., Patterson, K, & Hodges, J.R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology, 14*, 1165-1216.

Knott, R., Patterson, K, & Hodges, J.R. (2000). The role of speech production in auditory-verbal short-term memory: Evidence from progressive fluent aphasia. *Neuropsychologia, 38*, 125-142.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English.* Providence, Rhode Island: Brown University Press.

La Pointe, L.B., & Engle, R.W. (1990). Simple and complex working spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1118-1133.

Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Landauer, T.K., Foltz, P.W, & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259-284.

Larsen, R.J., & Marx, M.L., (1981**)**. *An introduction to mathematical statistics and its applications*. Englewood Cliffs, NJ: Prentice-Hall.

Lee, C.L. & Estes, W.K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior, 16*, 396-418.

Lee, C.L. & Estes W.K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 149-169.

Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: A dynamic network model. *International Journal of Psychology, 34*, 434-446.

Lewandowsky, S., Brown, G.D.A., Wright, T, & Nimmo, L.M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language, 54*, 20-38.

Lewandowsky, S., Duncan, M., & Brown, G.D.A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11*, 771-790.

Lewandowsky, S. & Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research, 63*, 163-173.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B.H. Ross (Ed). *The psychology of learning and motivation: Advances in research and theory* (Vol 49). (pp. 1-48). San Diego, CA, US: Elsevier Academic Press.

Lewandowsky, S., Nimmo, L.M., & Brown, G.D.A. (2008). When temporal isolation benefits memory for order. *Journal of Memory and Language, 58*, 415-428.

Lewandowsky, S., & Murdock, B.B. Jr. (1989). Memory for serial order. *Psychological Review, 96*, 25-57.

Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review, 15*, 875-888.

Lewandowsky, S., Oberauer, K., & Brown, G.D.A. (2008). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13,* 120-126.

Lewandowsky, S., Oberauer, K., & Brown, G.D.A. (2009). Response to Altmann: Adaptive forgetting by decay or removal of STM contents? *Trends in Cognitive Sciences, 13,* 280-281.

Li, S., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 837-847.

Loftus, E.F., & Freedman, J.L. (1972). Effect of category-name frequency on the speed of naming an instance of the category. *Journal of Verbal Learning and Verbal Behavior, 11*, 343-347.

Lorch, R.F., & Myers, J.L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 149-157.

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1-36.

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity Neighborhoods of spoken words. In G.T.M. Altmann (Ed.). *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122-147). Cambridge, MIT Press.

McGeoch, J.A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*, 352-370.

Mackintosh, B., & Mathews, A. (2003). Don’t look now: Attentional avoidance of emotionally valenced cues. *Cognition and Emotion, 17*, 623-646.

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

Majerus, S., Norris, D., & Patterson, K. (2007). What does a patient with semantic dementia remember in verbal short-term memory? Order and sound but not words. *Cognitive, Neuropsychology, 24*, 131-151.

Majerus, S., Poncelet, M., Elsen, B., & Van der Linden, M. (2006). Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. *European Journal of Cognitive Psychology, 18,* 848-873.

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*, 95-119.

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V. et al. (2006). The left intraparietal sulcus and verbal short-term memory: focus of attention or serial order? *Neuroimage, 32*, 880-891.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B.S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, *107,* 395-419.

Majerus, S., & Van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, *21*, 303-310.

Majerus, S., Van der Linden, M., Braissand, V., & Eliez, S. (2007). Verbal short-term memory in individuals with chromosome 22q11.2 deletion: Specific deficit in serial order retention capacities? *American Journal on Mental Retardation, 112*, 79-93.

Majerus, S., Van der Linden, M., Mulder, G., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language, 51*, 297-306.

Majerus, S., Van der Linden, M., Poncelet, M., & Metz-Letz, M.-N. (2004). Can phonological and semantic short-term memory be dissociated? Further evidence from Landau-Kleffner syndrome. *Cognitive Neuropsychology, 21*, 491-512.

Martin, N. (2009). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn and M. Page (Eds)., *Interactions between Short-term and Long-term Memory in the Verbal Domain* (pp. 220-243). Hove, UK: Psychology Press.

Martin, N., & Saffran, E.M. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language, 43*, 240-274.

Martin, N. & Saffran, E. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641-682.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52,* 83-113.

Martin, R.C., & Hamilton, A.C. (2007). Implications from cognitive neuropsychology for models of short-term and working memory. In N. Osaka, R.H. Logie, and M. D'Esposito (Eds)., *The Cognitive Neuroscience of Working Memory* (pp. 181-196). Hove, UK: Psychology Press.

Martin, R.C., & Lesch, M.F. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S.E. Gathercole (Ed)., *Models of Short-term Memory* (pp. 149-178). New York, USA: Oxford University Press.

Martin, R.C., Lesch, M.F., & Bartha, M.C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*, 3-29.

May, R.B., & Tryk, H.E. (1970). Word sequence, word frequency, and free recall. *Canadian Journal of Psychology, 24*, 299-304.

Mayzner, M.S., & Tresselt, M.E. (1958). Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology, 56*, 376-379.

Melton, A.W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1-21.

Merrit, P.S., DeLosh, E.L., & McDaniel, M.A. (2006). Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition, 34*, 1615-1627.

Milner, B. (1968). Disorders of memory after brain lesions in man. Preface: Material-specific and generalized memory loss. *Neuropsychologia*, *6*, 175-179.

Miller, L.M. (2004). *Semantic neighbourhood effects in serial recall*. Unpublished Honours thesis, University of Wollongong.

Miller, L.M., & Roodenrys. S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition, 37*, 850-865.

Monnier, C., & Syssau, A. (2008). Semantic contribution to verbal short-term memory: Are pleasant words easier to remember than neutral words in serial recall and serial recognition? *Memory & Cognition, 36,* 35-42.

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. MacKay, W. Prinz, E. Scheerer (Eds). *Language perception and production.* (pp. 273-311). Washington, DC, US: American Psychological Association.

Monsell, S. (1991). The nature and locus of word frequency effect. In D. Besner & G.W. Humphreys (Eds.), Basic processes in reading: Visual word recognition (pp. 148-197). Hillsdale, NJ: Erlbaum.

Monsell, S., Doyle, M.C., & Haggard, P.N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118*, 43-71.

Morin, C., Poirier, M., Fortin, C., & Hulme, C. (2006). Word frequency and the mixed-list paradox in immediate and delayed serial recall. *Psychonomic Bulletin & Review*, *13*, 724-729.

Mueller, S.T., Seymour, T.L., Kieras, D.E., & Meyer, D.E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1353-1380.

Mulligan, N. (2001). Word frequency and memory: Effects on absolute versus relative order memory and on item memory versus order memory. *Memory & Cognition, 29,* 977-985.

Murdock, B.B.Jr. (1961). The retention of individual items. *Journal of Experimental Psychology, 62,* 618-625.

Murdock, B.B.Jr. (1967). Recent developments in short-term memory. *British Journal of Psychology, 58,* 421-433.

Murdock, B.B. Jr. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General, 105,* 191-216.

Murray, D.J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*, 679-684.

Murray, A., & Jones, D.M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 594-598.

Nairne, J.S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251-269.

Nairne, J.S. (2001). A functional analysis of primary memory. In H.L. Roediger III, J.S. Nairne, I. Neath, A.M. Surprenant (Eds). *The nature of remembering: Essays in honor of Robert G. Crowder.* (pp. 283-296). Washington, DC, US: American Psychological Association.

Nairne, J.S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53-81.

Nairne, J.S., Ceo, D.A., & Reysen, M.B. (2007). The mnemonic effects of recall on immediate retention. *Memory & Cognition, 35,* 191-199.

Nairne, J.S., & Kelly, M.R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language*, *50*, 113-133.

Nairne, J.S., Riegler, G.L. & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 702-709.

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition*, *25*, 256-263.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7*, 403-423.

Neath, I., & Brown, G.D.A. (2006). SIMPLE: Further applications of a local distinctiveness model in memory. In B.H. Ross (Ed.), *The psychology of learning and motivation* (pp. 201-243). San Diego, CA: Academic Press.

Neath, I., & Nairne, J.S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, *2*, 429-441.

Nimmo, L.M., & Lewandowsky, S. (2005). From brief gaps to very long pauses: Temporal isolation does not benefit serial recall. *Psychonomic Bulletin & Review, 12*, 999-1004.

Nimmo, L.M., & Lewandowsky, S. (2006). Distinctiveness revisited: Unpredictable temporal isolation does not benefit short-term serial recall of heard or seen effects. *Memory & Cognition, 34*, 1368-1375.

Nimmo, L.M., & Roodenrys, S. (2004). Investigating the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language, 50,* 245-258.

Nosofsky, R. (1985). Luce's choice model and Thurstones's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics, 37,* 89-91.

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49,* 469-483.

Page, M.P.A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761-781.

Paivio, A. (1986). *Mental representations: A dual-coding theory approach*. New York: Oxford University Press, Inc.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255-287.

Patterson, K., Graham, N., & Hodges, J.R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience, 6,* 57-69.

Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*, 398-422.

Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48A*, 384-404.

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology, 50,* 408-412.

Posner, M.I., & Konick, A.F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology, 72*, 221-231.

Postman, L. (1961). Extra-experimental interference and the retention of words. *Journal of Experimental Psychology, 61*, 97-110.

Postman, L. (1962). The effects of language habits on the acquisition and retention of verbal associations. *Journal of Experimental Psychology, 64,* 7-19.

Postman, L., Turnage, T.W., & Silverstein, A. (1964). The running memory span for words. *Quarterly Journal of Experimental Psychology, 16,* 81-89.

Raymond, B. (1969). Short-term storage and long-term storage in free recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 567-574.

Raney, G.E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49*, 151-172.

Reitman, J.S. (1971). Mechanisms of forgetting in short-term memory. *Cognitive Psychology, 2,* 185-195.

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

Roodenrys, S., Hulme, C., Alban, J., Ellis, A.E., & Brown, G.D.A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22*, 695-701.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L.M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1019-1034.

Roodenrys, S., & Miller, L.M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Memory & Cognition, 36*, 578-587.

Roodenrys, S., & Quinlan, P.T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 8,* 71-78.

Ruchkin, D.S., Berndt, R.S., Johnson, R. Jr., Grafman, J., Ritter, W., & Canoune, H.L. (1999). Lexical contributions to retention of verbal information in working memory: Event-related brain potential evidence. *Journal of Memory and Language*, *41*, 345-364.

Ruchkin, D.S., Grafman, J., Cameron, K., & Berndt, R.S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavorial and Brain Sciences*, *26*, 709-777.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101-105.

Saint-Aubin, J., & LeBlanc, J. (2005). Word frequency effects in immediate serial recall of pure and mixed lists: Tests of the associative link hypothesis. *Canadian Journal of Experimental Psychology, 59*, 219-227.

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin & Review*, *12*, 171-177.

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology*, *52A,* 367-394.

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, *7*, 332-340.

Saint-Aubin, J., & Poirier, M. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. *Memory, 13,* 325-332.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory, and Cognition*, *21*, 168-175.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 419-425.

Schweickert, R., Chen, S., & Poirier, M. (1999). Redintegration and the useful lifetime of the verbal memory representation. *International Journal of Psychology*, *34*, 447-453.

Sears, C.R., Siakaluk, P.D., Chow, V.C., Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect?, *Journal of Psycholinguistic Research, 37*, 269-291.

Serra, M., & Nairne, J.S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21,* 34-40.

Shiffrin, R. M. (1993). Short-term memory: A brief commentary. *Memory and Cognition, 21*, 193-197.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 796-802.

Stuart, G.P, & Hulme, C. (2009). Lexical and semantic influences on immediate serial recall: A role for redintegration. In A. Thorn and M. Page (Eds), *Interactions between Short-term and Long-term Memory in the Verbal Domain* (pp. 157-176). Hove, UK: Psychology Press.

Sumby, W.H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1,* 443-450.

Surprenant, A.M. (1999). The effect of noise on memory for spoken syllables. *International Journal of Psychology, 34,* 328-333.

Surprenant, A.M., & Neath, I. (2009). The nine lives of short-term memory. In A. Thorn and M. Page (Eds), *Interactions between Short-term and Long-term Memory in the Verbal Domain* (pp. 16-43). Hove, UK: Psychology Press.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1589-1625.

Tan, L., & Ward, G. (2007). Output order in immediate serial recall. *Memory & Cognition, 35,* 1093-1106.

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15,* 535-542.

Tehan, G., & Humphreys, M.S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society*, *26*, 293-296.

Thorn, A.S.C., Frankish, C.R., & Gathercole, S.E. (2009). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn and M. Page (Eds), *Interactions between Short-term and Long-term Memory in the Verbal Domain* (pp. 198-219). Hove, UK: Psychology Press.

Thorn, A.S.C., Gathercole, S.E., & Frankish, C.R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50,* 133-158.

Thorndike, E.L., & Lorge, I. (1944). The Teachers Workbook of 30,000 Words. New York, Teachers College, Columbia Press.

Thorndike, E.L. (1923). *Educational psychology volume I: The original nature of man.* Teachers College, Columbia University. Downloaded from http://www.archive.org/details/originalnatureof011709mbp

Tolan, G., & Tehan, G. (1999). Determinants of short-term forgetting: Decay, retroactive interference, or proactive interference. *International Journal of Psychology*, *34*, 285-292.

Toro, J.M., Nespor, M., Mehler, J., & Bonatti, L.L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, *19*, 137-144.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 145-152.

Tse, C-S, & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, *15*, 675-690.

Tse, C-S, & Altarriba, J. (2009). The word concreteness effect occurs for positive, but not negative, emotion words in immediate serial recall. *British Journal of Psychology*, *100*, 91-109.

Turnage, T.W. (1967). Unit-sequence interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *6*, 61-65.

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgetting. *Psychological Review*, *67*, 73-95.

Vitevitch, M.S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 735-747.

Vitevitch, M.S. & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*, 325-329.

Vitevitch, M.S., & Luce, P.A. (1999). Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374-408.

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

Watkins, M.J. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529-534.

Watkins, M.J., LeCompte, D.C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 239-245.

Watkins, O.C., & Watkins, M.J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 712-718.

Ward, G., Tan, L., & Bhatarah, P. (2009). The roles of short-term and long-term verbal memory in free and serial recall: Towards a recency-based perspective. *Interactions between Short-term and Long-term Memory in the Verbal Domain* (pp. 44-62). Hove, UK: Psychology Press.

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 186-210.

Waugh, N.C. (1960). Serial position and the memory span. *American Journal of Psychology, 73*, 68-79.

Waugh, N.C., & Norman, D.A. (1965). Primary memory. *Psychological Review, 72*, 89-104.

Whiteman, H.L., Nairne, J.S., & Serra., M. (1994). Recognition and recall-like processes in the long-term reconstruction of order. *Memory, 2*, 275-294.

Winnograd, E., & Conn, C.P. (1971). Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning and Verbal Behavior, 10*, 702-706.

Woodward, A.J., Macken, W.J., & Jones, D.M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency?, *Journal of Memory and Language, 58*, 48-65.

Wright, C.E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition, 7*, 411-419.

Appendices

Appendix A: Excel Visual Basic code for the Feature model (Nairne, 1990) – Alternating HL condition

```
Option Explicit
'*********************************************************************************
'   Feature Model
'*********************************************************************************
'   Define variables
'*********************************************************************************

    Dim oMD(6, 20) As Integer               ' Vector set modality dependent features
    Dim oMI(6, 20) As Integer               ' Vector set modality independent features
    Dim MD(6, 20) As Integer                ' Modality dependent features - coded 1 or -1
    Dim MI(6, 20) As Integer                ' Modality independent features - coded 1 or -1
    Dim PMMD(6, 20) As Integer             ' Degraded modality dependent traces
    Dim PMMI(6, 20) As Integer             ' Degraded modality independent traces
    Dim IG(20) As Integer                   ' Internally generated trace for interference with
                                            ' MI features of last item
    Dim ROrder(2, 6) As Single              ' Random numbers
    Dim Order(6) As Integer                 ' Order of vectors in trial
    Dim Prob(6, 6) As Single               ' Probabilities of secondary memory items being
                                            ' recalled given degraded primary memory trace
    Dim ProbStore(1000, 6)                  ' Store to assess the average probability by item
    Dim Recover(6) As Integer               ' Records the number of times an item has been
                                            ' previously recovered
    Dim W(6) As Single                      ' Response bias weights
    Dim Bins(2, 6)                          ' Identifying item and cumulative probability bins
    Dim a As Single                         ' Scalar in distance measure
    Dim c As Single                         ' Scalar in recovery equation
    Dim F As Single                         ' Probability of overwriting
    Dim bMD(20) As Single                   ' Attention Parameter for MD features
    Dim bMI(20) As Single                   ' Attention Parameter for MI features
    Dim D(6, 6) As Single                   ' Distance measures between items (diagonal matrix)
    Dim S(6, 6) As Single                   ' Similarity measures for items
    Dim Start(20) As Integer                ' Vector of features with equal -1 and 1 elements
    Dim Index(20, 2) As Single              ' Array that allocates features to positions
    Dim Indexk As Integer
    Dim Sum, SumI, SumD As Single          ' Sum buffer for distance and similarity calculations
    Dim Switch As Single                    ' Buffer to switch values when reordering list items
    Dim p, CountPlus, CountNeg  As Integer  ' Position code for reordering
    Dim Count2Plus, Count2Neg  As Integer
    Dim Choose As Single                    ' Randomly generated choice of list item according to
                                            ' a normal distribution
    Dim Small As Single                     ' Comparison value to initiate sort
    Dim Flag As Integer                     ' Flag for repeat attempt and item recovery
    Dim ProbR As Single                     ' Probability of successful recovery
    Dim Recall(1000, 6) As Single           ' Vector of final recall order of items (0 = omission)
                                            ' 1000 trials
    Dim CIP(6)                    ' Order memory
    Dim i, j, k, m, n As Integer            ' Counters

    Sub Feature()
'*********************************************************************************
'   Create the vector set
'*********************************************************************************
    For i = 1 To 20
        bMD(i) = 1
        bMI(i) = 1
        IG(i) = 0
        For j = 1 To 6
```

```
            oMD(j, i) = 0
            oMI(j, i) = 0
            PMMD(j, i) = 0
            PMMI(j, i) = 0
        Next j
    Next i

    For j = 1 To 6
        CountPlus = 0
        CountNeg = 0
        For i = 1 To 20
10          Choose = Rnd(1)
            If (Choose <= 0.5 And CountPlus < 10) Then
                oMD(j, i) = 1
                CountPlus = CountPlus + 1
            ElseIf (Choose <= 1# And CountNeg < 10) Then
                oMD(j, i) = -1
                CountNeg = CountNeg + 1
            Else
                GoTo 10
            End If
        Next i
    Next j

    For j = 1 To 3
        CountPlus = 0
        CountNeg = 0
        Count2Plus = 0
        Count2Neg = 0
        For i = 1 To 20
50          Choose = Rnd(1)
            If (Choose <= 0.25 And Count2Plus < 5) Then
                oMI(j, i) = 2
                Count2Plus = Count2Plus + 1
            ElseIf (Choose <= 0.5 And CountPlus < 5) Then
                oMI(j, i) = 1
                CountPlus = CountPlus + 1
            ElseIf (Choose <= 0.75 And CountNeg < 5) Then
                oMI(j, i) = -1
                CountNeg = CountNeg + 1
            ElseIf (Choose <= 1# And Count2Neg < 5) Then
                oMI(j, i) = -2
                Count2Neg = Count2Neg + 1
            Else
                GoTo 50
            End If
        Next i
    Next j

    For j = 4 To 6
        CountPlus = 0
        CountNeg = 0
        For i = 1 To 20
100         Choose = Rnd(1)
            If (Choose <= 0.5 And CountPlus < 10) Then
                oMI(j, i) = 1
                CountPlus = CountPlus + 1
            ElseIf (Choose <= 1# And CountNeg < 10) Then
                oMI(j, i) = -1
                CountNeg = CountNeg + 1
```

```
            Else
              GoTo 100
            End If
        Next i
    Next j

'**********************************************************************************
'   Initialise everything
'**********************************************************************************
    For m = 1 To 1000
        For i = 1 To 20
          bMD(i) = 1
          bMI(i) = 1
          IG(i) = 0
        Next i
        For i = 1 To 6
            For j = 1 To 6
                D(i, j) = 0
                S(i, j) = 0
                Prob(i, j) = 0
            Next j
            W(i) = 1
            Recover(i) = 0
        Next i

        a = 7
        c = 2
        F = 0.8

'**********************************************************************************
'   Place the vectors in order for this trial
'**********************************************************************************
        For j = 1 To 3
            ROrder(1, j) = j
            ROrder(2, j) = Rnd(1)
        Next j

        For j = 1 To 3
            Switch = ROrder(2, j)
            Small = ROrder(1, j)
            For k = j To 3
                If (ROrder(2, k) < Switch) Then
                    Switch = ROrder(2, k)
                    Small = ROrder(1, k)
                    ROrder(2, k) = ROrder(2, j)
                    ROrder(1, k) = ROrder(1, j)
                    ROrder(2, j) = Switch
                    ROrder(1, j) = Small
                    p = k
                End If
            Next k
            ROrder(2, p) = 1
            ROrder(1, j) = Small
            ROrder(2, j) = Switch
        Next j
        For j = 4 To 6
            ROrder(1, j) = j
            ROrder(2, j) = Rnd(1)
        Next j
```

```
For j = 4 To 6
    Switch = ROrder(2, j)
    Small = ROrder(1, j)
    For k = j To 6
        If (ROrder(2, k) < Switch) Then
            Switch = ROrder(2, k)
            Small = ROrder(1, k)
            ROrder(2, k) = ROrder(2, j)
            ROrder(1, k) = ROrder(1, j)
            ROrder(2, j) = Switch
            ROrder(1, j) = Small
            p = k
        End If
    Next k
    ROrder(2, p) = 1
    ROrder(1, j) = Small
    ROrder(2, j) = Switch
Next j

For i = 1 To 3
    For j = 1 To 20
        MI(2 * i - 1, j) = oMI(ROrder(1, i), j)
        MD(2 * i - 1, j) = oMD(ROrder(1, i), j)
        MI(2 * i, j) = oMI(ROrder(1, i + 3), j)
        MD(2 * i, j) = oMD(ROrder(1, i + 3), j)
    Next j
Next i

11      Indexk = Int(Rnd(1) * 5) + 1
        If Indexk = 0 Then GoTo 11

        For j = 1 To 20
'*********************************************************************************
'   Have modelled no rehearsal effect on last item - enhances recency closer to that shown in
'   Nairne (1990)
'*********************************************************************************
            IG(j) = MI(Indexk, j)
        Next j

'*********************************************************************************
'   Determine the degraded primary memory vectors probability of overwriting is set to 1
'*********************************************************************************
      For i = 1 To 5
          For j = 1 To 20
              If (Rnd(1) <= F And (MI(i + 1, j) = MI(i, j))) Then
                  PMMI(i, j) = 0
              Else
                  PMMI(i, j) = MI(i, j)
              End If
              If (Rnd(1) <= F And (MD(i + 1, j) = MD(i, j))) Then
                  PMMD(i, j) = 0
              Else
                  PMMD(i, j) = MD(i, j)
              End If
          Next j
      Next i

      For j = 1 To 20

          If (Rnd(1) <= F And (MI(6, j) = IG(j))) Then
```

```
            PMMI(6, j) = 0
        Else
            PMMI(6, j) = MI(6, j)
        End If
        PMMD(6, j) = MD(6, j)
    Next j

'*******************************************************************************
'  Determine distance between each vector and from these values generate similarities
'*******************************************************************************

    For i = 1 To 6
        For k = 1 To 6
            SumI = 0
            SumD = 0
            For j = 1 To 20
                If (PMMI(i, j) <> MI(k, j)) Then
                    SumI = SumI + bMI(j)
                End If
                If (PMMD(i, j) <> MD(k, j)) Then
                    SumD = SumD + bMD(j)
                End If
            Next j
            D(i, k) = a * (SumI + SumD) / 40
            S(i, k) = Exp(-D(i, k))
        Next k
    Next i

'*******************************************************************************
'  Determine the probability of an secondary memory item being recalled given the degraded
'  primary memory trace
'*******************************************************************************

    For i = 1 To 6
        Sum = 0
        For j = 1 To 6
            Sum = Sum + W(j) * S(i, j)
        Next j
        For j = 1 To 6
            Prob(i, j) = W(j) * S(i, j) / Sum
        Next j
    Next i
    For i = 1 To 6
        ProbStore(m, i) = Prob(i, i)
    Next i

'*******************************************************************************
'  For each list, determine the boundaries for cumulative probability
'  Randomly generate a number between 0 and 1 and select the corresponding item according
'  to the cumulative bins scale
'*******************************************************************************

    For i = 1 To 6
        Flag = 0
        For j = 1 To 6
            Bins(1, j) = j
            Bins(2, j) = Prob(i, j)
        Next j
        For j = 2 To 6
            Bins(2, j) = Bins(2, j) + Bins(2, j - 1)
```

```
            Next j
'*********************************************************************************
'   Random number generator different to one used by Nairne (1990), Neath & Nairne (1995)
'   Mapped onto (0,1]
'*********************************************************************************
            Choose = 1 - Rnd(1)
             j = 1
            Do While Choose > Bins(2, j)
               j = j + 1
            Loop
            ProbR = Exp(-c * Recover(Bins(1, j)))
20          Choose = 1 - Rnd(1)
            If (Choose <= ProbR) Then
               Recall(m, i) = Bins(1, j)
'*********************************************************************************
'   This is not consistent with the referenced papers, but consistent with code of Feature model
'   Recovery is not indexed by the number of times an item was recalled, but is set to 1.
'   Replaced Recover(Bins(1, j)) = Recover(Bins(1, j)) + 1
'*********************************************************************************
               Recover(Bins(1, j)) = 1
            Else
               If Flag = 0 Then
                  Flag = 1
                  GoTo 20
               Else
                  Recall(m, i) = 0
               End If
            End If
         Next i
   Next m

   For j = 1 To 6
         Sum = 0
         SumI = 0
         For i = 1 To 1000
              SumI = SumI + ProbStore(i, j)
              If Recall(i, j) = j Then
                 Sum = Sum + 1
              End If
         Next i
         Recall(0, j) = Sum / 1000#
         ProbStore(0, j) = SumI / 1000
   Next j

'*********************************************************************************
'   Output results
'*********************************************************************************
   Sheets("Sheet1").Select
   Range("A1").Select
   ActiveCell.Offset(1, 1).Range("A1").Select

   For i = 1 To 6
         ActiveCell.Value = Recall(0, i)
         ActiveCell.Offset(0, 1).Range("A1").Select
   Next i
   ActiveCell.Offset(1, -6).Range("A1").Select
   For i = 1 To 6
         ActiveCell.Value = ProbStore(0, i)
         ActiveCell.Offset(0, 1).Range("A1").Select
   Next i
```

```
Sheets("Sheet2").Select
Range("A1").Select
ActiveCell.Offset(1, 1).Range("A1").Select

For i = 1 To 1000
    For j = 1 To 6
        ActiveCell.Value = Recall(i, j)
        ActiveCell.Offset(0, 1).Range("A1").Select
    Next j
    ActiveCell.Offset(1, -6).Range("A1").Select
Next i
Range("A1").Select
ActiveCell.Offset(1, 10).Range("A1").Select
For i = 1 To 1000
    For j = 1 To 6
        ActiveCell.Value = ProbStore(i, j)
        ActiveCell.Offset(0, 1).Range("A1").Select
    Next j
    ActiveCell.Offset(1, -6).Range("A1").Select
Next i

End Sub
```

Appendix B: Stimulus sets used in Experiments 1-2.

							Phonological similarity		
Condition	Item	Conc.	Freq.	Phonemes	Letters	PNS	Onset	Nucleus	Coda
LFLC	verb	337	3	3	4	6	0.28	0.27	0.20
	truce	335	3	4	5	9	0.30	0.35	0.20
	grief	303	15	4	5	12	0.34	0.23	0.20
	hint	312	26	4	4	12	0.29	0.25	0.25
	guess	247	31	3	5	11	0.35	0.27	0.20
	myth	334	28	3	4	16	0.34	0.25	0.20
	blame	293	23	4	5	5	0.29	0.23	0.26
	pause	306	36	3	5	59	0.24	0.39	0.33
	fate	255	35	3	4	36	0.27	0.23	0.20
	theme	336	37	3	5	11	0.28	0.23	0.26
	proof	328	33	4	5	5	0.27	0.35	0.20
	harm	244	33	3	4	23	0.29	0.41	0.26
	M	302.50	25.25	3.42	4.58	17.08	0.30	0.29	0.23
	SD	35.63	12.10	0.51	0.51	15.83	0.03	0.07	0.04
HFLC	hope	261	163	3	4	25	0.36	0.28	0.24
	cause	287	174	3	5	59	0.32	0.31	0.21
	truth	261	134	4	5	6	0.30	0.34	0.18
	risk	290	85	4	5	6	0.32	0.31	0.27
	deal	342	193	3	4	30	0.35	0.34	0.22
	chance	254	178	4	6	6	0.37	0.38	0.26
	rate	308	211	3	4	46	0.32	0.29	0.20
	rule	286	128	3	4	30	0.32	0.34	0.22
	cost	348	204	4	4	18	0.32	0.38	0.24
	style	344	107	4	5	10	0.34	0.31	0.22
	force	331	242	3	5	27	0.35	0.31	0.20
	claim	331	100	4	5	7	0.31	0.29	0.30
	M	303.58	159.92	3.50	4.67	22.50	0.33	0.32	0.23
	SD	34.98	49.19	0.52	0.65	17.21	0.02	0.03	0.03

Condition	Item	Conc.	Freq.	Phonemes	Letters	PNS	Phonological similarity		
							Onset	Nucleus	Coda
LFHC	tail	613	36	3	4	35	0.24	0.29	0.27
	barn	614	12	3	4	33	0.25	0.35	0.21
	steak	646	12	4	5	20	0.28	0.29	0.36
	sheep	622	20	3	5	23	0.37	0.36	0.24
	rope	608	44	3	4	38	0.40	0.29	0.24
	chalk	634	9	3	5	22	0.37	0.30	0.26
	thumb	638	27	3	5	15	0.26	0.32	0.22
	fox	605	16	4	3	26	0.28	0.37	0.31
	broom	613	7	4	5	10	0.31	0.39	0.22
	pond	623	19	4	4	16	0.23	0.37	0.26
	crane	606	5	4	5	15	0.36	0.29	0.21
	bell	620	42	3	4	29	0.25	0.31	0.27
	M	620.17	20.75	3.42	4.42	23.50	0.30	0.33	0.26
	SD	13.17	13.55	0.51	0.67	8.88	0.06	0.04	0.04
HFHC	heart	605	164	3	5	28	0.27	0.32	0.20
	bird	602	103	3	4	46	0.23	0.29	0.17
	head	603	310	3	4	38	0.27	0.28	0.17
	land	604	272	4	4	18	0.35	0.29	0.24
	sun	617	152	3	3	37	0.35	0.27	0.21
	meal	602	91	3	4	37	0.31	0.36	0.23
	glass	635	144	4	5	3	0.41	0.32	0.25
	ball	615	112	3	4	40	0.23	0.35	0.23
	dog	610	116	3	3	19	0.26	0.36	0.38
	skin	614	102	4	4	12	0.33	0.33	0.21
	film	604	122	4	4	6	0.28	0.33	0.26
	horse	613	133	3	5	23	0.27	0.34	0.25
	M	610.33	151.75	3.33	4.08	25.58	0.30	0.32	0.23
	SD	9.52	68.99	0.49	0.67	14.24	0.05	0.03	0.05

Appendix C: Means and standard deviations of correct recall in Experiments 1-6.

Table C1

Mean proportion and standard deviation of correct recall by serial position and condition in Experiment 1.

	Serial position					
	1	2	3	4	5	6
HFHC	.927 (.125)	.844 (.156)	.769 (.191)	.627 (.227)	.415 (.245)	.440 (.240)
HFLC	.921 (.142)	.829 (.188)	.748 (.213)	.604 (244)	.373 (.210)	.394 (.214)
LFHC	.902 (.151)	.812 (.166)	.664 (.221)	.576 (.233)	.302 (.203)	.361 (.205)
LFLC	.848 (.141)	.733 (.156)	.606 (.201)	.404 (.232)	.208 (.195)	.288 (.175)

Note. HFHC - High frequency, high concreteness, HFLC – High frequency, low concreteness, LFHC - Low frequency, high concreteness, LFLC – Low frequency, low concreteness. Standard deviations are given in brackets.

Table C2

Mean proportion and standard deviation of correct recall by serial position and condition in Experiment 2.

	Serial position					
	1	2	3	4	5	6
HFHC	.929 (.110)	.854 (.156)	.687 (.225)	.581 (.241)	.614 (.216)	.898 (.108)
HFLC	.911 (.112)	.781 (.198)	.599 (.232)	.512 (.269)	.547 (.232)	.917 (.081)
LFHC	.906 (.124)	.772 (.170)	.572 (.227)	.491 (.238)	.510 (.185)	.890 (.106)
LFLC	.880 (.150)	.658 (.228)	.472 (.240)	.351 (.203)	.406 (.191)	.802 (.136)

Note. HFHC - High frequency, high concreteness, HFLC – High frequency, low concreteness, LFHC - Low frequency, high concreteness, LFLC – Low frequency, low concreteness. Standard deviations are given in brackets.

Table C3

Mean proportion and standard deviation of correct recall by serial position and list type in Experiment 3.

	Serial position					
	1	2	3	4	5	6
Pure HF	.917 (.091)	.785 (.153)	.655 (.168)	.523 (.244)	.427 (.226)	.627 (.211)
Pure LF	.800 (.147)	.627 (.193)	.458 (.236)	.363 (.227)	.255 (.176)	.446 (.173)
Alt'g HL	.873 (.114)	.708 (.166)	.568 (.207)	.438 (.224)	.368 (.217)	.545 (.172)
Alt'g LH	.833 (.125)	.684 (.183)	.509 (.203)	.420 (.244)	.314 (.223)	.599 (.207)

Note. Alt'g – alternating lists. Standard deviations are given in brackets.

Table C4

Mean proportion and standard deviation of correct recall by serial position and list type in Experiment 4.

	Serial position					
	1	2	3	4	5	6
Half HL	.872 (.097)	.792 (.144)	.651 (.198)	.503 (.201)	.339 (.187)	.509 (.172)
Half LH	.807 (.104)	.681 (.152)	.583 (.173)	.439 (.226)	.354 (.191)	.571 (.184)
Alt'g HL	.859 (.128)	.738 (.169)	.632 (.212)	.481 (.227)	.351 (.194)	.538 (.147)
Alt'g LH	.830 (.129)	.714 (.192)	.531 (.205)	.436 (.221)	.311 (.178)	.554 (.187)

Note. Alt'g – alternating lists. Standard deviations are given in brackets.

Table C5

Mean proportion and standard deviation of correct recall by serial position and list type in Experiment 5.

	Serial position					
	1	2	3	4	5	6
Pure HF	.925 (.077)	.821 (.131)	.688 (.170)	.556 (.251)	.448 (.266)	.665 (.197)
Pure LF	.832 (.101)	.649 (.179)	.479 (.195)	.349 (.226)	.264 (.197)	.450 (.191)
Half HL	.924 (.060)	.821 (.109)	.670 (.185)	.448 (.213)	.340 (.215)	.536 (.167)
Half LH	.830 (.111)	.663 (.160)	.516 (.203)	.410 (.203)	.340 (.200)	.602 (.215)

Note. Standard deviations are given in brackets.

Table C6

Mean proportion and standard deviation of correct recall by serial position and list type in Experiment 6.

	Serial position					
	1	2	3	4	5	6
Pure HF	.888 (.092)	.784 (.118)	.613 (.176)	.508 (.206)	.391 (.181)	.618 (.147)
Pure LF	.789 (.100)	.568 (.159)	.384 (.164)	.280 (.146)	.177 (.131)	.404 (.129)
Alt'g HL	.887 (.090)	.682 (.160)	.569 (.196)	.388 (.190)	.311 (.190)	.501 (.162)
Alt'g LH	.826 (.112)	.684 (.168)	.500 (.196)	.353 (.193)	.246 (.159)	.551 (.192)
Half HL	.896 (.088)	.773 (.108)	.599 (.170)	.428 (.182)	.299 (.151)	.462 (.178)
Half LH	.819 (.094)	.647 (.165)	.496 (.183)	.394 (.180)	.328 (.170)	.564 (.146)
Seq'd HL	.885 (.083)	.776 (.150)	.587 (.173)	.451 (.188)	.322 (.179)	.583 (.190)
Seq'd LH	.770 (.168)	.650 (.169)	.582 (.154)	.497 (.149)	.354 (.189)	.543 (.158)

Note. Alt'g – alternating lists, Seq'd – sequenced lists. Standard deviations are given in brackets.

Appendix D: HF and LF stimulus sets used in Experiments 3-6

					Phonological similarity		
Condition	Item	log(Freq.)*	PNS	Conc.*	Onset	Nucleus	Coda
LF	babe	0.30	19	562	0.18	0.25	0.17
	barb	0.30	21	527	0.18	0.34	0.17
	barn	1.08	33	614	0.18	0.34	0.21
	bet	1.11	37	403	0.18	0.26	0.17
	bib	0.30	21	548	0.18	0.29	0.17
	bin	0.78	39	598	0.18	0.29	0.21
	bite	1.15	43	509	0.18	0.25	0.17
	bud	0.85	40	549	0.18	0.28	0.16
	cane	1.04	47	590	0.33	0.25	0.21
	cape	1.20	24	581	0.33	0.25	0.18
	carp	0.48	20	613	0.33	0.34	0.18
	cart	1.08	32	576	0.33	0.34	0.17
	coil	0.85	26	490	0.33	0.27	0.23
	cone	0.70	39	573	0.33	0.31	0.21
	cork	0.78	34	608	0.33	0.32	0.35
	cowl	0.00	23	456	0.33	0.30	0.23
	dame	0.60	25	528	0.20	0.25	0.22
	deed	1.00	31	410	0.20	0.33	0.16
	dell	0.30	27	513	0.20	0.26	0.23
	dim	1.23	28	402	0.20	0.29	0.22
	dime	0.70	28	582	0.20	0.25	0.22
	done	1.00	44	217	0.20	0.35	0.21
	dot	1.04	29	530	0.20	0.35	0.17
	dumb	1.04	35	340	0.20	0.28	0.22
	fawn	0.30	41	581	0.25	0.32	0.21
	fell	0.95	30	407	0.25	0.26	0.23
	foal	0.30	39	420	0.25	0.31	0.23
	foil	0.60	21	509	0.25	0.27	0.23
	hawk	0.85	30	623	0.29	0.32	0.35
	haze	0.78	53	509	0.29	0.25	0.24
	hide	0.70	42	451	0.29	0.25	0.16
	hood	0.78	26	547	0.29	0.35	0.16
	hop	1.00	24	494	0.29	0.35	0.18
	hose	0.60	49	596	0.29	0.31	0.26

					Phonological similarity		
Condition	Item	log(Freq.)*	PNS	Conc.*	Onset	Nucleus	Coda
LF *cont.*	howl	0.90	26	434	0.29	0.30	0.23
	keel	0.30	37	515	0.33	0.33	0.23
	kite	0.70	35	592	0.33	0.25	0.17
	knoll	0.30	36	486	0.27	0.35	0.23
	lace	1.15	30	545	0.28	0.25	0.26
	lard	0.30	44	517	0.28	0.34	0.16
	lark	0.60	27	578	0.28	0.34	0.35
	lease	0.78	30	371	0.28	0.33	0.26
	lice	0.48	25	543	0.28	0.25	0.26
	lime	0.95	26	590	0.28	0.25	0.22
	mall	1.04	34	417	0.25	0.32	0.23
	mat	1.15	41	513	0.25	0.29	0.17
	moat	0.60	32	453	0.25	0.31	0.17
	mole	0.78	45	590	0.25	0.31	0.23
	moss	0.85	28	575	0.25	0.35	0.26
	nip	0.30	25	515	0.27	0.29	0.18
	noose	0.00	16	542	0.27	0.39	0.26
	noun	0.30	11	387	0.27	0.30	0.21
	numb	0.70	20	379	0.27	0.28	0.22
	pall	0.30	42	362	0.19	0.32	0.23
	pat	0.95	35	400	0.19	0.29	0.17
	pearl	1.08	31	597	0.19	0.29	0.23
	peck	0.60	28	432	0.19	0.26	0.35
	peep	0.48	40	388	0.19	0.33	0.18
	pep	0.00	15	314	0.19	0.26	0.18
	pine	1.23	34	592	0.19	0.25	0.21
	poll	1.56	31	515	0.19	0.35	0.23
	pope	0.78	25	593	0.19	0.31	0.18
	puck	0.00	30	472	0.19	0.28	0.35
	pup	0.00	20	544	0.19	0.28	0.18
	rack	1.04	39	535	0.22	0.29	0.35
	rake	0.30	39	597	0.22	0.25	0.35
	ram	0.78	36	541	0.22	0.29	0.22
	rap	1.00	31	452	0.22	0.29	0.18
	reap	0.30	32	373	0.22	0.33	0.18

					Phonological similarity		
Condition	Item	log(Freq.)*	PNS	Conc.*	Onset	Nucleus	Coda
LF *cont.*	rhyme	0.70	31	434	0.22	0.25	0.22
	rim	0.95	31	511	0.22	0.29	0.22
	ripe	0.95	27	360	0.22	0.25	0.18
	rum	0.78	32	600	0.22	0.28	0.22
	sane	0.90	45	290	0.26	0.25	0.21
	sap	0.30	28	540	0.26	0.29	0.18
	sod	0.60	36	569	0.26	0.35	0.16
	sop	0.00	25	373	0.26	0.35	0.18
	tame	0.70	24	335	0.21	0.25	0.22
	toad	0.60	37	568	0.21	0.31	0.16
	toil	0.48	17	386	0.21	0.27	0.23
	toll	0.95	40	424	0.21	0.35	0.23
	veal	0.70	24	528	0.24	0.33	0.23
	veil	1.20	33	537	0.24	0.25	0.23
	vile	0.60	23	379	0.24	0.25	0.23
	wad	0.60	41	479	0.34	0.35	0.16
	weep	0.60	30	439	0.34	0.33	0.18
	weird	0.85	32	253	0.34	0.30	0.16
	whack	0.00	29	409	0.34	0.29	0.35
	whale	1.30	50	533	0.34	0.25	0.23
	whiff	0.48	30	413	0.34	0.29	0.27
	whip	1.18	39	570	0.34	0.29	0.18
	whirl	0.30	30	402	0.34	0.29	0.23
	whoop	0.00	18	383	0.34	0.35	0.18
	worm	1.23	19	611	0.34	0.29	0.22
	wreck	0.95	32	505	0.22	0.26	0.35
	wren	0.70	33	629	0.22	0.26	0.21
	M	0.70	31.38	491.30	0.25	0.30	0.22
	SD	0.36	8.36	92.80	0.05	0.03	0.05

Note. log(Freq.) – log (base 10) of word frequency; Conc. – concreteness, and PNS – phonological neighbourhood size.

* Frequency values were adjusted for the effects of homophones. Concreteness values are the weighted averages by frequency count across homophones.

					Phonological similarity		
Condition	Item	log(Freq.)*	PNS	Conc.*	Onset	Nucleus	Coda
HF	ball	2.06	40	611	0.19	0.33	0.24
	base	1.94	26	448	0.19	0.25	0.24
	beach	1.96	21	592	0.19	0.30	0.38
	bed	2.43	44	635	0.19	0.27	0.15
	bill	1.88	39	528	0.19	0.29	0.24
	bird	2.01	46	602	0.19	0.27	0.15
	board	2.03	61	565	0.19	0.33	0.15
	boat	1.89	35	637	0.19	0.30	0.16
	bone	1.85	42	588	0.19	0.30	0.20
	book	2.64	22	609	0.19	0.35	0.34
	card	1.85	42	565	0.34	0.33	0.15
	case	2.69	26	548	0.34	0.25	0.24
	cup	1.89	19	539	0.34	0.31	0.19
	cut	1.92	30	430	0.34	0.31	0.16
	dark	2.29	21	497	0.20	0.33	0.34
	date	1.89	28	514	0.20	0.25	0.16
	dead	2.26	28	429	0.20	0.27	0.15
	deal	2.29	30	342	0.20	0.30	0.24
	farm	1.95	14	565	0.25	0.33	0.23
	fat	1.96	31	540	0.25	0.32	0.16
	feel	2.48	36	324	0.25	0.30	0.24
	feet	2.53	28	636	0.25	0.30	0.16
	fight	2.01	39	455	0.25	0.25	0.16
	firm	1.99	17	400	0.25	0.27	0.23
	form	2.55	22	438	0.25	0.33	0.23
	full	2.44	20	378	0.25	0.35	0.24
	girl	2.64	22	607	0.33	0.27	0.24
	gun	1.99	29	612	0.33	0.31	0.20
	hall	2.15	39	555	0.29	0.33	0.24
	hard	2.48	45	425	0.29	0.33	0.15
	head	2.49	38	603	0.29	0.27	0.15
	heart	2.21	28	605	0.29	0.33	0.16
	heat	2.09	31	472	0.29	0.30	0.16
	hell	1.97	33	355	0.29	0.27	0.24

					Phonological similarity		
Condition	Item	log(Freq.)*	PNS	Conc.*	Onset	Nucleus	Coda
HF *cont.*	hill	2.07	39	588	0.29	0.29	0.24
	hope	2.21	25	261	0.29	0.30	0.19
	hot	2.16	31	507	0.29	0.38	0.16
	job	2.52	24	432	0.40	0.38	0.18
	keep	2.36	29	339	0.34	0.30	0.19
	kid	1.94	32	536	0.34	0.29	0.15
	lead	2.27	54	543	0.28	0.30	0.15
	learn	2.10	19	370	0.28	0.27	0.20
	leg	2.24	15	626	0.28	0.27	0.33
	light	2.56	40	550	0.28	0.25	0.16
	line	2.47	40	477	0.28	0.25	0.20
	lip	1.89	27	590	0.28	0.29	0.19
	loss	1.99	25	313	0.28	0.38	0.24
	male	2.11	45	552	0.25	0.25	0.24
	mark	1.86	29	464	0.25	0.33	0.34
	mass	2.05	29	397	0.25	0.32	0.24
	meal	1.96	37	602	0.25	0.30	0.24
	meet	2.38	32	417	0.25	0.30	0.16
	mile	2.24	30	460	0.25	0.25	0.24
	mine	2.02	35	452	0.25	0.25	0.20
	mouth	2.17	9	568	0.25	0.29	0.19
	name	2.54	20	405	0.26	0.25	0.23
	neck	1.90	21	587	0.26	0.27	0.34
	nice	2.18	17	279	0.26	0.25	0.24
	night	2.68	37	498	0.26	0.25	0.16
	nine	1.87	30	452	0.26	0.25	0.20
	nose	1.91	38	628	0.26	0.30	0.23
	paid	2.09	38	386	0.20	0.25	0.15
	park	1.89	36	579	0.20	0.33	0.34
	pass	2.08	21	385	0.20	0.33	0.24
	peace	2.42	29	359	0.20	0.30	0.24
	phone	1.86	32	624	0.20	0.30	0.20
	rain	1.99	45	566	0.22	0.25	0.20
	red	2.21	36	501	0.22	0.27	0.15
	road	2.07	50	583	0.22	0.30	0.15

Condition	Item	log(Freq.)*	PNS	Conc.*	Phonological similarity		
					Onset	Nucleus	Coda
HF *cont.*	rock	2.39	33	600	0.22	0.38	0.34
	role	2.08	46	354	0.22	0.30	0.24
	room	2.21	31	566	0.22	0.35	0.23
	rule	2.73	30	286	0.22	0.35	0.24
	seat	2.11	49	568	0.26	0.30	0.16
	sign	2.04	36	516	0.26	0.25	0.20
	soon	2.16	25	261	0.26	0.35	0.20
	sun	2.50	37	617	0.26	0.31	0.20
	take	2.18	28	332	0.21	0.25	0.34
	talk	2.50	30	422	0.21	0.33	0.34
	team	2.11	27	489	0.21	0.30	0.23
	tell	2.01	26	306	0.21	0.27	0.24
	term	2.40	21	374	0.21	0.27	0.23
	top	2.41	27	435	0.21	0.38	0.19
	town	2.39	17	556	0.21	0.29	0.20
	turn	2.35	21	359	0.21	0.27	0.20
	type	2.25	21	376	0.21	0.25	0.19
	walk	2.15	30	452	0.35	0.33	0.34
	wall	1.91	41	589	0.35	0.33	0.24
	week	2.32	33	379	0.35	0.30	0.34
	weight	2.69	40	412	0.35	0.25	0.16
	white	2.19	50	472	0.35	0.25	0.16
	wide	2.59	49	348	0.35	0.25	0.15
	wife	2.13	26	562	0.35	0.25	0.27
	wine	2.39	45	581	0.35	0.25	0.20
	wood	1.89	25	249	0.35	0.35	0.15
	write	2.92	43	377	0.22	0.25	0.16
	M	2.21	31.82	482.96	0.26	0.29	0.22
	SD	0.27	9.79	106.96	0.05	0.04	0.06

Note. log(Freq.) – log (base 10) of word frequency; Conc. – concreteness, and PNS – phonological neighbourhood size.

* Frequency values were adjusted for the effects of homophones. Concreteness values are the weighted averages by frequency count across homophones.