Using Word Association to Understand the Mental Lexicon: The Challenge of Coding

Caroline Joanne Handley

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy

Applied Linguistics

Swansea University

2023

Abstract

Word association tasks have been widely used within applied linguistics and psychology to understand connections between words in the mental lexicon. Many studies have attempted to understand the types of associations people form by coding responses, although results have been inconsistent. The dominant system contrasts paradigmatic and syntagmatic associates, responses that could replace or co-occur with the cue word in a sentence, respectively. In this thesis an alternative classification informed by dual coding theories of word knowledge was tested and then both systems were compared. The new system, based on Santos et al. (2011), posits that word associates are produced via either linguistic (symbolic) processing or conceptual processing grounded in sensorimotor experience of their referents. After a replication of their study, two experiments were conducted using new cue words and eliciting three responses to each cue. Results suggested that word association tasks activate conceptual processing to a greater extent than linguistic processing, contrary to expectations. This supports other research suggesting isolated words and sentences may elicit differential processing. Questions were also raised regarding the nature of conceptual knowledge. Analysis of multiple responses revealed that linguistic associates tended not to be produced after conceptual ones, a trend that was not found when associates were coded as paradigmatic or syntagmatic. In addition, the type of associate produced as the first response influenced that of subsequent ones. Finally, three coding systems were compared on multiple word associations produced by L2 English speakers. The largest difference found between L1 and L2 speakers was the number of linguistic associates produced. Linguistic associates also showed a stronger correlation with a vocabulary knowledge test than paradigmatic ones. Despite coding challenges encountered, the findings suggest that a linguistic-conceptual coding system could constitute a psychologically valid system for coding word association data, providing new insights into the mental lexicon.

Declarations and Statements

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ....[redacted]............... (candidate)

Date ....26 April 2023..................

This thesis is the result of my own investigations, except where otherwise stated. No correction services have been used.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ....[redacted].......... (candidate)

Date ....26 April 2023..................

I hereby give consent for my thesis, if accepted, to be available for electronic sharing after expiry of a bar on access approved by the Swansea University.

Signed ....[redacted]............... (candidate)

Date ....26 April 2023..................

Table of Contents

Acknowledgements

This thesis was made possible by the support of many people at Swansea University, in particular my supervisor, Tess Fitzpatrick. Her advice, criticisms, and knowledge have provided invaluable input to this thesis (the remaining flaws are my own). Her personal support and encouragement have been equally necessary to the completion of this work. I would also like to thank my second supervisor, Vivienne Rogers, who always knows the right questions to ask during important discussions. I would also like to thank everyone at Cardiff University who has continued to support me since I moved to Swansea with my supervisor in my second year. The conversations with faculty members at both universities have been greatly appreciated, especially through the annual Lexical Studies Conference. In particular, I would like to thank Alison Wray, whose intellectual insight and encouragement always inspired me, and Dawn Knight, for her continued encouragement.

I would also like to thank all my fellow students on the Lexical Studies program, both at Swansea and Cardiff, in particular Peter Thwaites for his discussions of analytical issues. Thanks are also due to the many friends and colleagues in Japan who have supported me. John Duplice has consistently motivated me and his practical advice and help have been incredible. I would also like to thank all the students in Wales and Japan who kindly agreed to participate in my experiments and provide the words this thesis is based on. I owe a special thank you to Lawrence Barsalou, not only because his ideas and research provided much of the inspiration for this work, but also for being so incredibly kind and encouraging when I was lucky enough to attend the same conference.

Finally, I would like to thank my friends in the UK and my family, for believing in me even when they thought this was never going to end.

List of Tables

List of Figures

Abbreviations

EAT	Edinburgh Associative Thesaurus
DCT	dual coding theory
LASS theory	Language and Situated Simulation theory
USF norms	University of South Florida norms
VST	Vocabulary Size Test
WA	word association
WAT	word association task

Chapter 1. Introduction

The abstraction of sensory impressions and motor actions into concepts, and the symbolisation of conceptual knowledge into words, are arguably the defining elements of human cognition (Kussmaul, 1877). Both types of knowledge, concepts and words, are often conflated and referred to as semantic knowledge, yet there is no consensus on what semantic knowledge is or how it is stored and processed (e.g., Gainotti, 2017). Given the elusiveness of lexical semantics or, more simply, word meanings (Hanks, 2013; Wittgenstein, 1953/2009), meaning is thought to depend on relations between words (e.g., Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005). This assumption explains why free word association (WA) experiments have been used within psychology and applied linguistics for over a century. Studying the spontaneous and unrestricted associations people form to words should provide informative insights into the mental lexicon and its development. Robust evidence from priming studies supports this idea (e.g., Hutchison, 2003). Yet it remains unclear why, or even whether, WA is an effective method for exploring the mental lexicon (Aitchison, 2003) and WA is still considered a “poorly defined” construct (Estes et al., 2011, p. 253). The rich data produced in these simple tasks seems to be both a strength and a weakness. As De Deyne et al. (2018) state:

> the information captured by an unconstrained word association task does seem to capture the *right kind* of meaning; meaning that is not limited by defining, characteristic, or entity features, but meaning that reflects mental representations that include properties about connotation, scripts and themes, properties notably absent from other subjective measures such as feature norms. (p. 1000)

The main approaches to analysing this “right kind of meaning” in applied linguistics have been grounded in language use, following Osgood et al. (1954). However, as highlighted by De Deyne et al. (2018), words get their meanings from both linguistic and non-linguistic sources, from their referents and contexts of use, not just their textual distributions. In this thesis, an analytical approach is developed that gives equal consideration to both ways of making word meanings. The implications of this approach for interpreting the meanings made in WA tasks are discussed in relation to L1 and L2 WA behaviour, re-evaluating how WA can contribute to understanding the mental lexicon. In line with this view of word meanings, the mental lexicon is viewed as an organised but dynamic, distributed network of flexible (fuzzy) lexical representations or units of form-meaning abstracted from an individual's linguistic and non-linguistic experiences (Aitchison, 2003; Barsalou, 1999, 2008; De Deyne et al., 2016; Evans, 2009; Hanks, 2013; Huth et al., 2012).

1.1 Word Association in Cognitive and Applied Linguistics Research

The word association task (WAT) is a simple experiment in which participants are asked to respond to a cue word, or normally a list of cue words, with another word or words. It is sometimes referred to as free word association (e.g., Nelson et al., 2004), as no restrictions are placed on the response words. Typically, participants are asked to speak or write the first word "that comes to mind" (Clark, 1970, p. 272). WATs have been used in psycho- and cognitive linguistics research to produce stimuli for word priming experiments (e.g., Nelson et al., 2004) and to study relations within the mental lexicon, often conceived as a semantic network (e.g., Crutch & Jackson, 2011; De Deyne & Storms, 2008; Henriksen, 2008). Within applied linguistics, WATs have been used to study language development in children (e.g., Ervin, 1961) and second or foreign language (L2) learners (for a review, see Fitzpatrick & Thwaites, 2020). Despite the evidence that WA data can contribute towards understanding how words are related in the (developing) mental lexicon (e.g., De Deyne et al., 2018; Fitzpatrick & Thwaites, 2020), it seems to have produced more questions than answers (e.g., Fitzpatrick, 2009; Lucas, 2000).

This thesis explores whether bringing together the two areas of psycho- or cognitive linguistic theory and applied linguistic research could yield any new insights into the nature and/or utility of WATs as a tool for understanding the mental lexicon (as suggested by Fitzpatrick, 2007). Within WA research, one commonality between these two fields is their use of WA "norms" (e.g., De Deyne et al., 2018; Kruse et al., 1987; Meyer & Schvaneveldt, 1971; Nelson et al., 2013; Schmitt, 1998b; Wolter, 2002). These are lists of cue words and the response words most frequently elicited by each cue, based on the responses of multiple participants. The proportion of times a cue elicits the same response word from participants is used as a measure of the association strength of the cue-response word pair. Although WA norms have been used in different ways, as illustrated below, they have led to mixed results in both fields.

The use of norms lists in research in cognitive science and psycholinguistics has partly provided the basis for the notion that WA reveals something about how words are stored and processed. The use of cue-response word pairs from WA norms lists produces effective primes (e.g., Meyer & Schvaneveldt, 1971). That is to say, the presentation of the cue word speeds recognition of the dominant or primary response word, as determined by WA norms lists, such as the University of South Florida (USF) norms (Nelson et al., 1998, 2004). However, despite the robustness of this finding (e.g., De Deyne et al., 2018; Hutchison, 2003), there is no accepted explanation for why it occurs or what information WA provides about the mental lexicon (Nelson et al., 2013).

In this sense, the strength of this task is also its weakness. Free WA imposes no restrictions on the links people access to process word meanings, making it an ideal experimental task for examining relations in the mental lexicon. As a result, this simple task yields complex outcomes. WA data is messy. People's responses reflect the multiple ways in which the cue and response words are connected. Conformity and individuality are both undisputed features of WA data and these two aspects vary greatly as a function of the cue words. Some cue words tend to elicit the same response from most people (for example, the cue word "left" elicited the response "right" from over 90% of respondents in the USF norms). However, most cue words show far less consistency. Perhaps for this reason, many authors suggest that semantic knowledge, both conceptual and lexical, is organised primarily by semantic similarity (i.e., the word "tiger" is related to "cat" as both refer to a feline animal) rather than word associativity, despite conflicting evidence (e.g., Lucas, 2000). Semantic similarity is no simple judgement, as reviewed in the next chapter, but it is restricted to categorical relations and shared features. This makes it less complicated compared with explaining associative similarity, which comprises various types of relations between words and the concepts to which they refer (Estes et al., 2011).

Despite the difficulty in explaining WA behaviour, there are advantages to using WATs as a tool for examining the mental lexicon, which may have partly motivated their use within applied linguistic research. WATs are easy to administer, without the need for expensive equipment, and they can be presented as a game, reducing the likelihood that they cause stress to children or L2 language learners (e.g., Entwisle, 1966; Wojcik & Kandhadai, 2020). Moreover, WATs allow participants to demonstrate what they know, rather than assessing language knowledge based on a small number of words. WA norms have been used to provide a comparison of the WAs of L1 and L2 speakers and to trace the development of L2 learners' knowledge as proficiency level increases. For example, Schmitt (1998a, 1998b) and Wolter (2002) both showed that L2 learners tend to produce different responses to L1 speakers. However, neither confirmed the hypothesis that L2 learners' associations would become more native-like (i.e., would include more of the dominant responses given by L1 speakers to the same cue words) as proficiency increased. Given the variability of responses produced by L1 speakers (e.g., Nelson et al., 2004), comparison to WA norms lists may not represent the best approach to understanding the development of the L2 lexicon through WA.

A further approach to assessing the development of lexical knowledge through WA has been to code the relationship of the response word to the cue. This method assumes that the type of association between the cue and response may be informative about lexical knowledge, not just the actual response words. One coding system that has frequently been used to code WA responses is based on a distinction between paradigmatic and syntagmatic associations (Osgood et al., 1954). Based on Saussure (1916/1983), a response is paradigmatically associated to a cue if it might replace it in a sentence and syntagmatically related if it might co-occur with it. Thus, paradigmatic associates are typically of the same word class as the cue, whereas syntagmatic ones are of a different word class (e.g., Osgood et al., 1954).

Early experiments with children suggested their associations showed an age-related shift from syntagmatic to paradigmatic responses (e.g., Ervin, 1961; Entwisle, 1966). However, the application of this coding system to L2 WA has resulted in mixed outcomes, which has been related to issues in analysing the data (Fitzpatrick, 2006; Zareva & Wolter, 2012). The most comprehensive attempt to date to resolve the challenges of applying the paradigmatic-syntagmatic coding system has been the development of a meaning-, position-, and form-based coding system by Fitzpatrick (2006, 2007, 2009). This system has been shown to be reliable (Fitzpatrick et al., 2015) and is stimulating further research (e.g., Kim, 2013). Overall, it would seem that coding WA data may yield further insights into the development of L2 lexical knowledge beyond the information obtained from comparison with L1 speaker norms.

In sum, WA norms predict behavioural responses in word priming experiments, demonstrating the psychological validity of WA (Hutchison, 2003). Furthermore, coding WA data according to the type of association produced by respondents to each cue word can be informative about lexical knowledge and how it develops as a function of language exposure. Nevertheless, interpreting WA data remains a challenge in both areas of research and types of analysis. These observations raise two interesting and related questions: (a) Can a WA coding system developed from cognitive psychological theory provide any insights regarding the nature of the semantic relations accessed in WATs beyond those provided by applying coding systems based in linguistic theories or measures of association strength (cue-response norms)? (b) Can a novel coding system created on the above basis yield new insights into the development of the mental lexicon?

Therefore, the main aims of the research reported in this thesis were (a) to develop a system for coding and analysing data from WATs that is supported by psycho- or cognitive linguistic theory; and (b) to assess whether this system could be used to distinguish the lexical knowledge of L2 learners of different proficiency levels. Due to challenges that arose in developing the new coding system, it was developed iteratively, and tested empirically at each of the three stages of development. Following the third iteration, the new coding system was compared to the two influential coding systems introduced above (paradigmatic-syntagmatic and meaning-, position-, and form-based), which indicated it could contribute new insights into WA behaviour. The fourth and final experiment partially addressed the second aim (however, due to Covid restrictions this analysis was conducted on a single data set, rather than on the multiple data sets originally envisaged).

1.2 Key Terms

The experiments reported in this thesis are informed by psychological and linguistic research and theory. Broadly speaking, psychological theories of language (and concepts) informed the coding system developed here and the assumptions it was based on. The main aims of the experiments build on previous WA research, conducted within the field of applied linguistics. Therefore, for the purposes of this thesis, it is convenient to separate these two influences, despite the considerable shared ground. Regarding this decision, when psychological aspects of language use are emphasised, for simplicity, the term *cognitive science* is used to describe this research area, given the interdisciplinary nature of this field. Cognitive science is used to depict a range of approaches aimed at understanding linguistic and conceptual knowledge. It is recognised that much of this research has been conducted in the fields of psychology, psycholinguistics, cognitive linguistics, and cognitive neuroscience, and that research within all these areas is discussed, particularly in Chapters 2 and 8. These fields are contrasted with WA research exploring the development of lexical knowledge that has been conducted from an applied linguistic approach (e.g., to inform foreign language instruction). Such research is ascribed to the field of applied linguistics, although some of these studies could also be considered psycholinguistic or cognitive linguistic research.

The term *L2 learners* is used to refer to all learners of a foreign language, including multilingual learners and those learning the language predominantly in a classroom setting (studying the L2 as a foreign language) and in immersion contexts (studying the L2 in a country in which it is used as a native and/or official language). Likewise, the term *L1 speaker* is used to refer to speakers who have been raised in an environment dominated by the language in question (but who may also speak other languages). The term *bilingual speaker* is used to refer to speakers who were raised in a context in which two or more languages were used or who regularly use two or more languages. It is recognised that these distinctions do not reflect the continuum of experiences and linguistic abilities that exists (e.g., DeLuca et al., 2019). The term *L1 speaker* is not intended to suggest an ideal language user or a completely monolingual speaker or to imply any superiority.

A few common terms will be used interchangeably when describing word association (WA) behaviour. The words presented to participants are referred to as *cue* or *stimulus words*. The words provided by participants in response to such cue words are referred to as *responses* and *associates*. The particular response word given by the largest proportion of respondents to any one cue word is typically called the *dominant response*, although the term *primary response* is also used. In all four WA experiments conducted multiple associates were collected to each cue word. The term *output position* is used to refer to the order in which responses were produced, typically distinguishing between the first, second, and third responses.

In keeping with the embodied or grounded cognition framework that informs the experiments conducted (Barsalou, 2008; see Chapter 2, Section 2.6), throughout the experimental chapters of this thesis a distinction is made between conceptual and linguistic knowledge. Although both types of knowledge are semantic, that is they are inherently meaningful or informative, they are assumed to be constituted from different sources of semantic knowledge. Conceptual knowledge is considered to be grounded in (abstracted from) multimodal representations or situated simulations. Concepts are rich, although not necessarily conscious, sensorimotor and interoceptive representations that emerge from the brain's interactions with the socio-environment and the body (e.g., Thompson, 2007). In contrast, linguistic knowledge is viewed as symbolic representations that are both abstracted from conceptual knowledge and facilitate access to concepts (Barsalou, 1999; Dove et al., 2022). In addition, the type of conceptual knowledge referred to in this thesis as *situated simulations* or *situational descriptors* is often referred to as *thematic relations* (e.g., Estes et al., 2011; Mirman et al., 2017), so this term is also used, particularly in reference to previous studies in Chapters 2 and 8.

1.3 Thesis Overview

To end this short introduction chapter, a brief outline of the structure of this thesis is presented. Chapter 2 reviews relevant WA research within applied linguistics and psychology, then goes on to explore theories of conceptual and linguistic knowledge within cognitive science. This provides the background for a detailed explanation of the theoretical assumptions that guided the development of the WA coding system which is the focus of this thesis. The theory that was chosen is a weak theory of embodied cognition and language, in which it is posited that lexical or semantic knowledge emerges from the interaction of multimodal conceptual representations and language or symbolic knowledge (e.g., Andrews et al., 2009; Barsalou, 1999; Barsalou et al., 2008; Connell, 2019). This theory provides an appropriate basis for developing a new coding system for WAs due to its focus on the lexical semantics of language and its increasing dominance within cognitive science (Spivey, 2023) [1].

[1] A search on Google scholar for articles including any of the four terms *embodied/grounded + cognition/language* published since 1992 returned approximately 50,000 hits and a similar Scopus search yielded nearly 5000 articles, with a steep increase since 2005, see Appendix 1.

Chapter 3 presents a close replication of a WA experiment conducted with L1 English speakers by Santos et al. (2011). This study introduced a novel coding system which was designed to test a specific theory of word knowledge, Language and Situated Simulation (LASS) theory (Barsalou et al., 2008). This is a weak embodied cognition theory, according to which semantic knowledge is both embodied (conceptual) and symbolic (linguistic). The coding system is based on a distinction between linguistic, taxonomic, and conceptual relations between words. Although similar results were obtained in the replication as in the original study, several important issues were identified with the design and analysis of the original experiment. Therefore, additional, modified analyses are also reported. Overall, the results suggested that the distinction between linguistic and conceptual processing could be valid and could provide a useful basis for an alternative system for analysing WA data.

Chapter 4 reports a second WA experiment in which the main aim was to address limitations in the design of the first (replication) experiment which challenged the reliability of the results, namely the choice of cue words and coding of responses by Santos et al. (2011), and issues that arose in the elicitation of responses in the replication. The subcategories used to select cue words and code associations as linguistic (forward and backward collocations, synonyms, and antonyms) had not been matched to conceptual subcategories by Santos et al. This potential confound was resolved in the second experiment, by expanding the conceptual category through the creation of four conceptual subcategories (object and cue word properties, situational descriptors, and semantic field). In addition, cue words were chosen that were expected to provide participants with equal probabilities of producing a linguistic or conceptual associate. Finally, due to the substantial variation across both participants and cue words in the number of responses produced in the replication experiment, in this WAT a fixed number of response words (three) were collected for each cue word. These changes improved the reliability of the results and the analyses and enabled new analyses, which revealed interesting findings that indicated further support for the new coding system. However, contrary to the predictions and conclusions of Santos et al., the findings indicated that WAs tend to be produced by accessing conceptual rather than linguistic knowledge (at least to the extent that this distinction is valid). This warranted further investigation as it has important implications for the use of WATs to understand the development of word knowledge in L2 learners.

Chapter 5 reports a new WA experiment, again conducted with L1 English speakers, but for which only high frequency words were chosen as cues, so that the same cues could later be used with L2 learners. Taking into account the results of the previous experiment, the new cue words were also selected such that each cue should elicit a higher proportion of either linguistic or conceptual associates. The results of this third experiment indicated it is possible to select cues that will influence response types in the way intended[2]. More importantly, this effect did not obscure the findings of interest detected in the previous experiment. In Chapter 6, the data from this third experiment is used to compare the new coding system with the most frequently used coding system, in which responses are classified as paradigmatic or syntagmatic, as well as a more detailed coding system, with nine meaning-, position-, and form-based categories (Fitzpatrick, 2007).

Chapter 7 provides the final test of the new linguistic-conceptual coding system, by applying it to WA data collected from L2 learners of English (please note that Covid restrictions limited data collection opportunities). The new system was again compared with the paradigmatic-syntagmatic and the meaning-, position-, and form-based coding systems (i.e., all responses were coded in three different ways). The results showed that the greatest difference between the L1 and L2 speaker groups was in the number of linguistic associates produced. Moreover, the number of linguistic associates correlated more reliably with scores on a vocabulary test than the number of any other associate type from any of the three coding systems. Together, these findings suggest that distinguishing between linguistic and conceptual associates may offer new insights into the development of L2 lexical knowledge that are not revealed when using the paradigmatic-syntagmatic coding system.

Chapter 8 interprets the results from all four experiments in view of previous literature and current theories of semantic knowledge. The nature of WA is explored and the insight this task can provide into the mental lexicon is evaluated. The challenges in coding WA data are considered in depth, as well as other limitations of the research reported in this thesis, providing suggestions for future research.

[2] This is very similar to the rationale of norms lists, except the type of association is being examined, rather than the individual response words.

Overall, it is asserted that the linguistic-conceptual coding system may offer an informative interpretation of WA data and warrants further investigation.

Chapter 2. Literature Review

The main aims of this thesis are to develop a new, theoretically motivated system for coding word association (WA) data through experiments with L1 English speakers and then to test it with WA data from second language (L2) learners. In order to establish the foundations for the empirical work that follows, this chapter scrutinises three key areas of investigation: word association research, theories of conceptual or semantic knowledge, and embodied cognition approaches to concepts and language. The ways in which these three areas contribute to the aims of the thesis is highlighted. A brief outline of this chapter is given first.

In Section 2.1 a brief explanation of WA tasks (WATs) and their analysis is provided, as a general orientation to the next two sections. Section 2.2 evaluates the use of WATs within applied linguistics, particularly research related to foreign language learning. This review is focused on the use of coding systems to categorise responses, as this is the approach pursued in this thesis. Section 2.3 reviews two uses of WATs in psycho- and cognitive linguistic research that are pertinent to the aims of this thesis. The first is the use of data collected in WA experiments in priming studies. The second compares WA data with data from corpora and semantic network models. This research is reviewed to establish the extent to which WA is a language task, in the sense that responses reflect knowledge of word co-occurrence, or distributional semantics.

In Section 2.4 a chronological review of theories of conceptual knowledge is presented, leading up to an extended discussion around theories of embodied or grounded cognition (e.g., Varela, et al., 1991; Barsalou, 2008) in Section 2.5. The two most relevant aspects of these theories for the research conducted in this thesis are discussed in more detail in Section 2.6. The first is the distinction made between lexical knowledge and conceptual knowledge, and how these two systems are proposed to interact. The second is the debate regarding the extent to which language is grounded in multimodal experience or is abstracted as a symbolic system.

Finally, in Section 2.7, the literature on conceptual knowledge, embodied cognition and language, and word association research is combined to suggest a new approach to analysing WAs. This approach is based on theories that language is both embodied and symbolic, specifically the theory of perceptual symbol systems (Barsalou, 1999) and language and situated simulation (LASS) theory (Barsalou et al., 2008). This chapter ends by previewing the first experiment, a close replication of a WA experiment designed to test LASS theory (Santos et al., 2011).

2.1 Word Association Tasks and Their Analysis

In WATs a list of cue words is presented to participants who are asked to provide the first word or words they think of in response to each cue (for a detailed overview of the task, see Cramer, 1968). It is a simple experiment to conduct yet it generates a wealth of data for analysis. Importantly, many studies have demonstrated that this data is informative about relations between words in the mental lexicon (as reviewed in Hutchison, 2003; see Section 2.3). WAs collected from one group of participants can be used to predict the behaviour of new groups of participants. Two main approaches have been taken to analysing WA data: producing response norms and coding responses. Both reduce the data to elucidate generalisable patterns within the responses. The key difference is that norming focuses on the individual words produced whereas coding attempts to capture the way in which the response words are associated with each cue, the type of association.

An overview of both these methods of analysing WA data is given next, although it is noted that they differ in their relevance to the experiments presented in this thesis. Observations related to the use of WA norms in psycholinguistic experiments and research comparing WA norms to language use (reviewed in Section 2.3) provided the theoretical motivation for trying to develop a new coding system informed by cognitive psychological theory. WA norms were also used to select the cue words used in the studies. However, native speaker norms are not used as a benchmark for evaluating the stereotypy of WA behaviour in this thesis, so studies that have adopted this approach (e.g., Kruse et al., 1987; Schmitt, 1998b) are not reviewed in this chapter. The focus in this thesis is on designing and testing a coding system for WA data that is informative about the nature of the task and the extent to which it can be used to understand the mental lexicon. Thus, research involving the coding of WA responses, reviewed in Section 2.2, provides crucial comparisons against which the new coding system can be assessed.

2.1.1 Producing Word Association Norms

WA “norms” are essentially lists of the response words given to a set of cue words, ordered by the frequency of response. WA norms are based on the observation that adult L1 speakers often produce the same responses to individual cue words on WATs (Kent & Rosanoff, 1910). For example, when presented with the cue word “black” most people respond with the word “white”. Kent and Rosanoff used this response reliability in studying the WA behaviour of people classified as having mental health problems. They classified WA responses as common, doubtful, and individual reactions, to compare the stereotypy and idiosyncrasy of responses of people with mental health problems to “normal” people. Since their study, several large datasets of WA norms have been collected and made publicly available, which have then been extensively used in further research, such as studies involving priming experiments (see Section 2.3). Three sets of WA norms are briefly described below which have been used in studies reported later in this chapter, as well as a new set of WA norms that is anticipated to become widely employed.

One of the earliest sets of norms are the Minnesota norms (MN; Russell & Jenkins, 1954; see also Jenkins, 1970). These norms were created using the 100 cue words on the Kent-Rosanoff cue word list (Kent & Rosanoff, 1910). The MN consist of data collected from just over 1000 students, who each provided a single response to each cue word. The availability of this large, standardised dataset may partly explain why these cue words (or a subset of them) have been used in many studies, including research with children (e.g., Sandgren et al., 2021; Tresselt & Mayzner, 1964), English language learners (e.g., Kruse et al, 1987, Riegel et al., 1967; Sökmen, 1993), as a test of creativity (e.g., Gough, 1976), and research using translations of the cue words into other languages (e.g., Rosenzweig, 1970).

Two other sets of norms that have been widely used, including in studies reviewed below, are the Edinburgh Associative Thesaurus (EAT; Kiss et al., 1973), maintained by the University of Oxford (http://hdl.handle.net/20.500.12024/1251) and the University of South Florida (USF) norms (Nelson et al., 1998, 2004). The EAT is composed of 100 responses to 8400 cue words, collected from students at British universities (each student provided one associate to each of 100 cues). However, some studies that have used it have only used the responses to the 100 Kent-Rosanoff cue words, in effect using it as a more recent version of the MN.

The norms list used in the experiments reported in this thesis is the University of South Florida (USF) norms (Nelson et al., 1998, 2004), which is publicly available in its entirety (http://w3.usf.edu/FreeAssociation/). At the time the research reported here was conducted, this was the largest and most comprehensive set of norms available. An average of 150 respondents provided one associate to a total of 5018 cue words (the minimum number of respondents was 94, for the cue word "carton" and the maximum was 206, for the cue word "routine"). The norms list contains all response words given by at least two people to each cue word. The USF norms have been a standard source of word association data used to inform research decisions (McRae et al., 2012). Despite their age, they remain an important benchmark (De Deyne et al., 2018) and continue to be used to choose word cues in psycholinguistic studies (e.g., Hoedemaker & Gordon, 2017). A further reason for using the USF norms in the studies reported in this thesis is that they were used by Santos et al. (2011) to select the cue words in the experiment that was replicated here, presented in Chapter 3.

However, it is noted that a new, much larger set of WA norms has recently been released, the Small World of Words (SWOW) norms (De Deyne et al., 2018; https://smallworldofwords.org/en/project/home). The detailed analysis of these norms provided by De Deyne et al. provides interesting information about WA and includes a comparison to the USF norms (Nelson et al., 1998). The SWOW is a multilingual resource, although only the English language dataset (SWOW-EN) is described in this chapter. The SWOW-EN is a set of word associations collected online, which remains open to participation. At the latest update in 2018, it contained responses produced by over 90,000 adult L1 English speakers (mostly university graduates) to 12,292 English cue words. Each participant provides three responses to 18 cue words, although it is possible to participate multiple times. This means that at present, the SWOW-EN contains over 3.5 million response words. Of note, this dataset includes all cue words from the USF norms and the EAT norms (Kiss et al., 1973), in recognition of the importance of these norms lists in psycholinguistic research. More importantly, when De Deyne et al. compared the ability of the three WA norms lists to predict behaviour on psycholinguistic tasks (see Section 2.3.2), they found the three lists were similar in performance. Thus, the larger and much more recent SWOW-EN norms confirm the validity of the USF norms, which were used to choose the cue words for the experiments reported in this thesis.

2.1.2 Coding Word Association Data

An alternative approach to analysing WA responses, and the one used in this thesis, is to code the type of association between the response and the cue word. Various coding systems and variations of systems have been used for this purpose (e.g., Chaffin, 1997; Cremer et al., 2011; Sökmen, 1993; Spätgens & Schoonen, 2020), although most have been composed of two to four main categories (Deese, 1965). In the earliest recorded WAT, Galton (1879) classified his own responses as sense-imagery, histrionic (imagined action), verbal, and proper names. Deese (1965) made a distinction between free association based on structure (internal organisation of thought) and contingency (co-occurrence of stimuli or words), although he viewed responses as "a distribution of *intraverbal* meaning" (p. 42). The two coding systems which have been most widely used in applied linguistics research are described in detail below, namely the paradigmatic-syntagmatic coding system (Osgood et al., 1954) and the meaning-, position-, and form-based coding system (Fitzpatrick, 2006, 2007, 2009). These two systems provide the benchmarks against which the new coding system developed in this thesis is later compared (Chapters 6 and 7).

The paradigmatic-syntagmatic coding system was proposed by Osgood et al. (1954) and has been adopted in many WA studies that have involved the coding of responses. It has also become the most frequently used coding system in research with L2 learners (e.g., Nissen & Henriksen, 2006; Wolter, 2001; Zareva, 2007; see Section 2.2.2). In WA research, responses are typically divided into paradigmatic or syntagmatic associations to the cue words. However, as discussed below, Osgood et al. viewed paradigmatic and syntagmatic similarity as a cline, with some associations entering into both types of relation. These terms were first introduced by Saussure (1916/1983), to distinguish between words that could replace a word in a sentence (paradigmatic) or co-occur with it (syntagmatic). This idea also appears in Firth (1957), who classifies paradigmatic relations as system (vertical) relations and syntagmatic relations as structure (horizontal) relations.

The paradigmatic-syntagmatic coding system, and versions of it, continues to be used in WA research (e.g., Sandgren et al., 2021; Wojcik & Kandhadai, 2020). However, an interesting alternative coding system was proposed by Santos et al. in 2011 based on a distinction between linguistic, taxonomic, and simulation (conceptual) knowledge (see Section 2.7). Although this study has been influential (e.g., Cayol & Nazir, 2020; Connell, 2019; Jones & Golonka, 2012; Lenci et al., 2013), to the best of my knowledge, Santos et al.'s coding system has not been employed in any subsequent WA research. Therefore, it warrants careful scrutiny and is the starting point for the linguistic-conceptual coding system developed later in this thesis. The key difference between the paradigmatic-syntagmatic coding system and the linguistic-conceptual coding system stems from the underlying motivation for their creation. Osgood et al. (1954) first suggested categorising word associations as paradigmatic or syntagmatic in an attempt to create a systematic classification based on "purely linguistic criteria" (p. 115). Despite the awareness that it might not explain all associates (for example, phonological associates), they argued that studies combining linguistic, semantic, and psychological criteria had been unsystematic. The linguistic-conceptual coding system developed in this thesis is an attempt to combine linguistic and psychological categories in a systematic framework based on cognitive science theory. Osgood et al.´s proposal for a paradigmatic-syntagmatic coding system for WA data is examined in more detail next.

Osgood et al.´s (1954) proposal to study WA behaviour by coding responses as paradigmatic or syntagmatic was published in a report of a psycholinguistics seminar. The aims of this seminar were to identify major research problems to be addressed by this newly emerging field of study and suggest productive approaches for doing so. This seminar was heavily influenced by Shannon´s (1948) theory of communication, more commonly referred to as information theory (see Pierce, 1961/1980 for a non-mathematical introduction). This was viewed as one of the three elements of psycholinguistics, along with linguistics and behaviourist learning theory. The proposals regarding WA research are within a section on the "sequential or transitional structure of language behaviour" (Osgood et al., 1954, p. 93). At that time, WA was seen as a simpler method for understanding the predictive value (informativeness) of specific words than collecting large amounts of natural speech. In other words, the rationale for WA research was that WA behaviour was easier to obtain than corpus data and assumed to be highly similar (this hypothesis can now be tested, see section 2.3.2).

The paradigmatic-syntagmatic coding system was intended to capture two different types of similarity relations between the cue and response words, namely, that responses could replace the cue word or co-occur with it. Perhaps due to the influence of behaviourism, Osgood et al. (1954) suggest that the probability of both types of response can be ascertained from their distribution in language. This contrasts with Saussure (1916/1983), who stated that paradigmatic relations were associative, mental relations arising from memory. According to Saussure (p. 171), syntagmatic relations are *in praesentia* in discourse, whereas paradigmatic relations are *in absentia*, held in memory based on form and/or meaning. For Osgood et al., the purpose of the paradigmatic-syntagmatic coding system was to analyse the transitional probabilities of words, such that WA data could aid in the creation of texts for testing comprehension. The focus was on the extent to which cue words elicited common responses, much like the motivation for producing WA norms. In this respect, Osgood et al. made two key observations about this coding system which, as noted by Fitzpatrick and Thwaites (2020), have largely been overlooked by subsequent research.

First, Osgood et al. (1954) stated that paradigmatic associates are the same word class as the cue, but syntagmatic associates could be of the same or a different word class. As a result, responses of the same word class as the cue should be more frequent because they "can be similar both paradigmatically and syntagmatically" (p. 116). They gave the example of the cue-response "table-chair" as an example of this. Although double coding was intended to constitute an important feature of this coding system, in practice it is rarely performed (two exceptions are Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015). It is also worth noting that Osgood et al. applied dual coding to coordinates such as "table-chair" and contrasts such as "black-white", which feature prominently on the Kent-Rosanoff cue word list used in many studies (e.g., Postman & Keppel, 1970; Riegel et al., 1967; Sandgren et al., 2021). Based on these examples, almost half (48) of the dominant responses in the MN (Russell & Jenkins, 1954) have both paradigmatic and syntagmatic associations with their cue words. Moreover, in line with Osgood et al.´s observation, dual association responses in the MN tend to be highly dominant cue-response pairs (i.e., those given by most respondents).

Second, Osgood et al. (1954) stated that paradigmatic and syntagmatic were not absolute values but rather a continuum, given the variability of the strength of the relationship between the cue word and a particular associate. They illustrate this with the words "table" and "chair". Their example for paradigmatic similarity is that "chair" is more similar than "cat" as a response to the cue "table", although all three words could occupy the same place in a sentence. Regarding syntagmatic similarity, the verb "sit" is more informative or predictive of the cue word "chair" than "see".

Although this is intuitively obvious, it is difficult to apply in practice. To the best of my knowledge, no study exists that has attempted to code response as strong, moderate, and weak paradigmatic and syntagmatic associations. It is not clear whether it would be possible to apply such a coding system with adequate objectivity.

In sum, the paradigmatic-syntagmatic coding system, as it was originally proposed, was intended to be used to ascertain the degree of similarity between words, assuming that similarity would be indexed by the frequency of response across participants. The main motivation was the hypothesis that WA data could form a proxy for measuring the distribution of words in text (now possible through computer-assisted corpus analysis). The purpose of coding WA data was thus similar to that of producing norms for such data. This approach was influenced by information theory; similarity was related to the predictive strength of the cue word to elicit certain responses. In line with behaviourist theories, WA behaviour was assumed to reflect associations formed from experiencing the cue and response words together in texts. Of note, the two categories were not viewed as mutually exclusive; rather, responses that reflected both types of association (paradigmatic and syntagmatic) were assumed to be stronger than those that held only one type of association.

As described in Section 2.2, the purpose and implementation of the paradigmatic-syntagmatic coding system in subsequent WA research has differed somewhat from Osgood et al.´s (1954) proposal. It has been productively used to investigate the developing mental lexicon in both first (L1) and second (L2) language speakers (e.g., Burke & Peters, 1986; Ervin, 1961; Hirsh & Tree, 2001; Nissen & Henriksen, 2006; Sandgren et al., 2021; Wolter, 2001). However, some studies have relied solely on word class to distinguish between paradigmatic and syntagmatic associates (e.g., Ervin, 1961; Piranian, 1983) and almost all studies have operationalised the categories as mutually exclusive (see Fitzpatrick & Thwaites, 2020, for further discussion).

Building from the paradigmatic-syntagmatic coding system, more recently a new system has been extensively developed and implemented by Fitzpatrick (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015). This system is of note because the language(s) spoken by participants (monolingual and bilingual speakers and L2 learners), the cue words used, and the research questions varied across studies. This means that the utility of this coding system has been comprehensively tested. It has been adopted by other researchers (e.g., Kim, 2013).

Fitzpatrick (2006) redefined the paradigmatic-syntagmatic division with the more user-friendly terms of *meaning-based* and *position-based associations*. She also added form-based associations and erratic associations, the latter used to code responses whose connection to the cue word could not be identified and non-responses. This coding system was based partly on Nation's (2001) description of word knowledge, in which he highlighted the need to know the form and meaning of a word and also how to use it in text. One of the main potential advantages of this coding system compared with the paradigmatic-syntagmatic system is that the main categories were further divided into various subcategories, enabling a richer analysis of the data (see also Cremer et al., 2011, for a similar approach). It is this aspect that makes this coding system relevant to the new coding system developed in this thesis, as will be further explored in Chapters 7 and 8.

In studies using the meaning-, position-, and form-based system (described in Sections 2.1.2 and 2.2.2), the three main categories remained constant, although Fitzpatrick altered the subcategories (described in more detail in Chapter 6) and Fitzpatrick and Izura (2011) did not use subcategories. In the first study, Fitzpatrick (2006) used 17 subcategories. In the 2007 and 2009 studies, this was reduced to 10 subcategories to increase the data within each subcategory (the erratic association category was also divided into erratic and no association in the 2009 study). In Fitzpatrick et al. (2015), an additional four dual-link categories were added. The use of dual coding makes this system distinctive, however this issue is not investigated in this thesis, as it would have resulted in sparse data. As all eight subcategories used here would need to be combined with each other to investigate dual links, this problem would become intractable, leading to unreliable analyses.

To the best of my knowledge, Fitzpatrick's (2006, 2007, 2009) meaning-, position-, and form-based coding system has not been directly compared to the paradigmatic-syntagmatic coding system. Such an analysis would be useful, given that position-based responses are not identical to syntagmatic ones, nor are meaning-based and paradigmatic response the same. Position-based responses comprise solely collocations, therefore this category captures fewer associates than the syntagmatic category. All other meaningfully associated responses are coded as such; thus, this category includes all paradigmatic associates but also some syntagmatic ones. These two coding systems demonstrate that coding decisions can have a significant impact on the analysis and interpretation of WA data.

2.1.3 Assessing the Validity of Word Association Data

As reviewed above, there are two main approaches to analysing WA data: production of, or comparison to, norms or the application of a coding system. However, the usefulness of either approach cannot be assessed without first considering the validity and reliability of WA as an empirical paradigm. WA research related to these two issues of validity and reliability is described below.

Perhaps the easiest objection to raise regarding WA is whether people actually produce the first word that comes to mind. This issue was explored in a series of experiments by Baror and Bar (2016). They collected word associations from participants while also giving them another task to complete, in a high or low cognitive load condition. For example, participants gave one associate to ten cue words while trying to remember a sequence of two or six digits, in total responding to 100 cue words. Across each experiment, they found that under high cognitive load participants were more likely to give an associate produced by other participants. They suggested that this result indicated that participants adopted different strategies for producing word associations in the two conditions: Under low cognitive load participants preferred to retrieve less frequent associations, whereas under high load they tended to exploit more frequent associates (i.e., ones that were easier to produce). However, the number of participants in each condition was quite low (typically 10-12 people) and there was no control group who only gave associates without an additional task. Thus, these results do not offer clear evidence as to whether WA behaviour while performing another simultaneous task is comparable to WA under normal task conditions, in which it is the sole activity.

In addition to cognitive load leading to higher conformity in WA responses, one early WA study (Entwisle, 1966) found evidence that perceived power imbalances may lead participants to try to produce more desirable responses. In investigating whether the format (individual spoken or group written format) influenced responses, Entwisle found that power relations might be a more important, confounding factor. She found children aged 11 gave more responses of the same word class as the cue to verb and adjective cue words in the individual spoken format. In a detailed analysis of the adjective cues, she found this was due to children from a higher socio-economic status (ascertained by caregiver income) providing twice as many antonym responses in the individual spoken format. Entwisle speculated that this group of children changed some responses to ones they felt was more correct or desirable to please the experimenter. This led her to speculate that response differences between the two formats were due to role relations between these children and the adult experimenter, rather than the task format.

To confirm this hypothesis, Entwisle (1966) compared responses of female college students under both administrative formats of the WAT and when the interviewer was a peer or a male professor. The students gave almost identical proportions of antonyms to adjectives when interviewed by a peer (similar role relations) or performing the task in a group written format. However, they gave a significantly higher proportion of antonyms when responding to the male professor (an authority figure). This implies that perceived power relations can create pressure on some participants to provide a "correct" response rather than the first response that comes to mind, increasing conformity. Taken together, these findings from Entwisle and the later study by Baror and Bar (2016) suggest that participants tend to produce the first word that comes to mind provided they do not experience extraneous task pressure.

A related issue is the stability of responses over time. One early investigation found that response stability may be subject to individual differences. Brotsky and Linton (1967) collected WA responses to 158 cue words from 146 students on two separate occasions, 10 weeks apart. They found approximately one-third of responses were identical, although there was huge variation across participants in the number of identical associates produced (ranging from 8 to 99). They also found a similar variation in the number of dominant responses given by participants (the number of times they gave the associate that was given by the largest proportion of participants). These two response patterns were also very highly correlated, suggesting they might both reflect differences in creativity in WA behaviour.

In a more recent study, Rodd et al. (2016) investigated the influence of recent experiences on WA behaviour. They found that low frequency meanings of homophones (such as "bark") can be primed in WATs, but that such priming effects are small and quickly decay within about 20 minutes. Moreover, they found that word associations given by participants who had long-term experience within a particular field (rowing) were influenced by very recent experiences to a much greater extent than by their long-term experience. Specifically, rowers who had rowed on the day of completing the WAT showed a greater tendency to produce rowing-related associations to words which have a common unrelated meaning as well as a specific rowing-related meaning (such as “feather”) than those who had rowed the day before. These findings suggest that very recent prior exposure to any cue word included in a WAT may influence the associates given by some participants to that word, but that the effect seems to be quite limited.

The stability or reliability of WA responses has been investigated by Burke and Peters (1986), using a paradigmatic and syntagmatic coding system, and by Fitzpatrick et al. (2015), using a meaning-, position-, and form-based coding system with ten subcategories. Fitzpatrick et al. (2015) found evidence of age-related effects on WA behaviour. They collected associations from two different sets of L1 English-speaking twins, aged either 16 or over 65, using 100 cue words which were randomly selected from word frequency norms, from the 2000 and 3000 highest frequency word bands of the British National Corpus (BNC; Davies, 2004). They found the associates produced by any individual were more similar to the responses produced by other members of their age group than they were to those of the other age group. Burke and Peters (1986) administered a WAT consisting of 113 high frequency cue words from all four main word classes to 80 young (average age 21) and 80 old adults (average age 71). They found that both groups gave similar proportions of paradigmatic and syntagmatic associates, with a higher proportion of paradigmatic associates. However, they did find some differences in the dominant associates between both groups. These studies illustrate a potential problem when applying previously collected norms, such as the USF norms (Nelson et al., 1998), described in Section 2.1.1, to examine word knowledge or association behaviour.

In the same studies, Fitzpatrick et al. (2015) administered the same WAT on two separate occasions to a group of 36 16-year-old participants, coding the data from both tasks. They used ten subcategories of meaning-, position-, and form-based responses. For the main subcategories (collocations, synonyms, taxonomic associates and conceptual associates), individual participants tended to give associates of the same type of relation to the cue words at both times. These results showed that WA behaviour is reliable over time within individuals, even if it varies between respondents (at an individual and/or group level). They also confirm that response type is an informative unit of analysis, not just response word. Burke and Peters (1986) gave the same WAT to 27 participants from each age group, 2 to 3 months later, although in the retest they analysed the actual words produced, not response types. They found that both groups produced, on average, the same response to 40% of the cues, indicating consistency and variability in response behaviour.

More concerning for the use of WA norms, Fitzpatrick (2007) found substantial individual differences in WA behaviour. She gave two matched lists of 100 cue words to 30 L1 English speakers, one week apart. Each participant produced one associate to each cue word. The majority of responses were collocations, synonyms, and conceptual associates, in that order. However, there were large individual differences in the number of each type of association. This indicates that not only do responses vary as a function of the cue words used (see Section 2.1.1), they also vary across participants. One positive finding was that the proportion of each type of association produced by each participant to the second set of cue words was more similar to their own response types to the first set of cues than to the response types produced by any other participant to the same set of cues (calculated as Euclidean differences). This confirms the within-subject reliability of WA behaviour, in contrast to the between-subject variability.

Together, these findings suggest that WA can provide a reliable measure of the mental lexicon, provided associates are collected from a large sample of participants. As Deese (1965) stated: “the meaning of any [linguistic] form is not given by a single response or, indeed, by a collection of responses at some particular time, but by the *potential distribution of responses to that form*” (p. 41). Given that norming data shows that some cue-response pairs are more homogenous across people than others, the reliability of responses across individuals is a factor that should be considered when choosing the cue words to use in a WAT. Depending on the aims, cue words can be chosen that elicit one or a few dominant responses or that elicit a wide range of responses. Effects of participants’ socio-cultural context should also be taken into account in WATs. As Galton (1879) noted in the very first reported WA study, his associations were influenced by ideas that were “prevalent in that stratum of English society in which I was born and bred, and have subsequently lived” (p. 158). With these caveats in mind, the utility of coding WA data to understand the mental lexicon, particularly as it changes over time, is discussed in the next section.

2.2 Coding Word Associations to Study the Developing Mental Lexicon

In this section, the application of coding systems to WA data is reviewed. First, the use of the paradigmatic-syntagmatic coding system in studies investigating the developing lexicon of L1 and L2 speakers is evaluated. Finally, the use of the more detailed coding system developed by Fitzpatrick (2006, 2007, 2009) is described, which has been used in studies with monolingual, bilingual, and L2 speakers.

As established in the previous section, the paradigmatic-syntagmatic coding system was intended to be a linguistic coding system, to capture the transitional probabilities in language. Although Osgood et al. (1954) did not claim that WA behaviour was necessarily a product solely of linguistic knowledge (they predicted their coding system would need refining), this was implied by the explanation of their coding system. Therefore, it is not surprising that in early research this was an explicit assumption (Brown & Berko, 1960; Clark, 1970; Ervin, 1961). For example, Ervin (1961) interpreted the associations of young children as being "based on training by forward contiguity in speech" (p. 372) and Clark (1970, p. 272) stated that "language must play a central role" in WA. Although such claims are not made in more recent WA research, an implicit assumption that word knowledge resides in the language system(s) of an individual appears to have remained. Although links between words and concepts (the physical stimuli they refer to) are recognised within word association research, the emphasis is on how words relate based on grammatical class and structure. As Nissen and Henriksen (2006) observed:

> By classifying data according to response types such as paradigmatic, syntagmatic and phonological responses only, a purely linguistic understanding based in syntax and grammar is forced upon response associations without considering either the rich variety of relations between words or the various ways in which words may be known to a language user. (p. 405)

The research reported in this thesis is motivated by research in cognitive science that suggests the assumption that WA is solely or primarily a linguistic task should be questioned (see also McRae et al., 2012; Zareva & Wolter, 2012). Therefore, the key findings, controversies, and challenges of applying the paradigmatic-syntagmatic coding system are reviewed next (it is noted that challenges were anticipated by Osgood et al., 1954).

2.2.1 Coding the Word Associations of L1 Learners

The widespread adoption of the paradigmatic-syntagmatic coding system within WA research arose from claims that in normally developing children there is a syntagmatic-paradigmatic shift in response type (e.g., Ervin, 1961; Entwisle, 1966). In a more recent study, Wojcik and Kandhadai (2020) also found a syntagmatic-paradigmatic shift in children's WA responses, although there were large differences between participants and cue words in the proportion of paradigmatic associates produced.

Entwisle´s (1966) study is particularly notable, as she collected associations to 96 cue words of various classes and frequencies (including 30 words from the Kent-Rosanoff cues) from over 1000 children aged from 4 to 11 and conducted detailed analyses. One set of analyses, by word class, suggested that a shift in response behaviour may not be an accurate description. The youngest children tended to produce nouns in response to all cue words, whereas older children were increasingly likely to provide paradigmatic associates to verb and adjective cue words. She also found a slight decrease in the proportion of paradigmatic responses given to nouns and adjectives by college students compared to 11-year-olds, which she interpreted as suggesting that new meanings continue to develop in early adulthood. This idea has not been explored in depth in subsequent WA research. However, it should be noted that Entwisle retained the notion that WA behaviour reveals the development of the mental lexicon, speculating that "as associations mature, a word first elicits a noun, second a syntactic response, third a paradigmatic response, and fourth a different syntactic response." (p. 72).

Although children´s WAs are not investigated in this thesis, L2 studies employing the paradigmatic-syntagmatic coding system were motivated by L1 research. In view of this, it is extremely important to note that two explanations were proposed for the syntagmatic-paradigmatic shift. Although these explanations are not mutually exclusive, only one is transferable to L2 learners.

WA studies with children were interpreted as revealing the development of the mental lexicon, leading to an assumption that L2 learners' lexicons would develop in a similar way. They showed that very young children give many "clang" responses (words that are phonologically similar to the cue word but have no semantic relation, see Deese, 1965). This is followed, between 5 and 9 years of age, by a shift from primarily syntagmatic to more adult-like response behaviour, with a predominance of paradigmatic associates (Nelson, 1977). For example, Ervin (1961) found an age-related increase in synonym and antonym associates. She suggested two explanations for these findings, namely that paradigmatic associations between words were indicative of greater word knowledge, arising from increased exposure to language, or of educational experience.

The children being observed in WATs are both learning language and attending school, making it difficult to disentangle these two proposed causes. However, if educational experience is an explanatory factor contributing to the syntagmatic-paradigmatic shift, this shift should at least be attenuated in adult L2 learners. Support for the role of education is that the shift was found to occur during the first few years of formal education (between 5 and 9). As noted in a review by Nelson (1977), young children's responses "reflected how things were used, or their relations in the real world rather than in logic or language" (1977, p. 110). This is also in line with the results of studies conducted by Luria (1976) with illiterate and low-literacy adults. He found that illiterate adults not only grouped objects by thematic (real world) relations, they could not even conceive of any similarity between objects that had coordinate (paradigmatic) relations (such as "axe" and "hammer"). Similarly, Entwisle (1966) found that the proportion of paradigmatic associates given by children was associated with IQ as well as age, leading her to suggest that WA might not be as appropriate for studying the development of adult's language knowledge (p. 118).

The possible role of education is also hinted at by Petrey´s (1977) suggestion that a division between episodic and semantic associates might be more appropriate than the paradigmatic-syntagmatic distinction, due to the latter's dependency on word-class. She reanalysed some of Entwisle's (1966) data, claiming it shows that children give more episodic associates (based on autobiographical experience, either in a verbal or situational context) and adults give more abstract semantic associates, such as synonyms and antonyms. Petrey (1977) also noted that very young children's responses were also much more diverse, with only three cue words eliciting a dominant associate given by over 30% of kindergarten-aged children. She suggests that two of these ("salt-pepper" and "table-chair") "are unusually frequent because linguistic and situational contiguity establish the same link" (p. 67). This notion of episodic and semantic associates is similar to the theoretical basis of the alternative coding system developed in this thesis. However, many of the experiential associations given as examples by Petrey are completions of utterances (often multiword), similar in nature to a collocation, so clearly arising from linguistic experience, not situational experience. This complicates the interpretation of her findings.

In sum, it is of note that although similarities might be assumed between children learning an L1 language and L2 learners, the application of findings from L1 WA research to L2 research is problematic. It is highly plausible that the syntagmatic-paradigmatic shift reflects cognitive development in general, rather than being specific to language development (Cremer et al., 2011, reach a similar conclusion based on the results of a WAT conducted with bilingual children and adults, coded with their own categorisation system). If this is the case, adult learners of an L2 may show less of a shift or even no shift at all. Research related to this issue is reviewed next.

2.2.2 Coding the Word Associations of L2 Learners

The evidence for a syntagmatic-paradigmatic shift in children´s WA responses led to interest in investigating whether such a change can also be found in L2 learners and, if so, whether it is related to increasing proficiency (e.g., Nissen & Henriksen, 2006; Piranian, 1983; Zareva & Wolter, 2012). Piranian (1983) provides a specific rationale for such research, suggesting that a similar shift in response type would indicate that the shift is due to language development, whereas failure to detect a shift would indicate it is due to general cognitive development. Overall, the results have provided mixed support for a developmental syntagmatic-paradigmatic shift in L2 learners. It should be noted that most WA studies with L2 learners have not relied solely on word class to determine whether associates are paradigmatic or syntagmatic (excepting Piranian, 1983). Noun pairs that form compound words or collocations (such as “camp-fire” or “football-field”) have been coded as syntagmatic. However, decisions such as this, often varying slightly between researchers, have led to various issues regarding the generalisability of studies using this coding system (Fitzpatrick & Thwaites, 2020). In addition, some studies have taken cue words from the Kent-Rosanoff list (e.g., Norrby & Håkansson, 2007; Postman & Keppel, 1970). Meara (1983) suggested this may be a problematic source for investigating the L2 mental lexicon as it contains only high frequency cue words and many of these elicit one highly dominant associate. In addition, translation equivalents of these cue-response pairs exist in many languages. Some illustrative studies are discussed next to give an indication of the different approaches, findings, and interpretations within L2 WA research.

Some studies have found that L2 learners produce fewer paradigmatic associates than L1 speakers on WATs. For example, Norrby and Håkansson (2007) compared the WA behaviour of two groups of elementary level L2 learners of Swedish, one studying in Sweden and the other in Australia, with those of L1 speakers in Australia. They asked all participants to provide one response to each of the 100 Kent-Rosanoff cue words, translated into Swedish. They found the L2 learners in Australia produced more no responses, phonological responses, and translations compared to either other group. Both learner groups produced fewer paradigmatic associates than the L1 speakers, but whereas the learners in Sweden produced more syntagmatic associates, the group in Australia produced fewer syntagmatic and paradigmatic associates than L1 speakers.

Similarly, in a WAT with 45 cue words of different word frequencies, Wolter (2001) found that L1 speakers of English gave more paradigmatic associates than L2 learners (L1 Japanese). However, both groups gave similar numbers of syntagmatic associates; the L2 learners produced more clang associates (phonologically associated, idiosyncratic associates, and no response). Wolter followed the WAT with a depth of knowledge test for each of the cue words, which showed that L2 learners gave more clang and fewer paradigmatic responses to words that were moderately well-known compared with L1 speakers. Even for words that were very well-known (i.e., participants could produce a synonym or translation of the word and create a sentence illustrating its meaning), L2 learners produced fewer paradigmatic and more syntagmatic associates than L1 speakers.

Despite this finding, Wolter (2001) interpreted his findings as a shift from meaningless (clang) to meaningful associates with increasing lexical knowledge, rather than a syntagmatic-paradigmatic shift. Similarly, in a comparison of WAs to familiar words, given by L1 speakers and advanced and intermediate L2 learners, Zareva and Wolter (2012) found that the intermediate level learners produced a greater proportion of syntagmatic associates than advanced learners and L1 speakers and thus, fewer paradigmatic associates. However, they interpreted this as a "gradual transition" (p. 60), explicitly rejecting the idea of a response shift. The interpretations of their results given by Wolter (2001) and Zareva and Wolter (2012) would appear to reflect an opposition to the idea in L1 research that paradigmatic associates indicate superior organisation within the mental lexicon compared with syntagmatic associates (e.g., Burke & Peters, 1986; Ervin, 1961; Sandgren et al., 2021), a notion that does seem worth challenging.

Other studies have found partial support for the hypothesis that L1 speakers produce more paradigmatic associates than L2 learners. For example, Nissen and Henriksen (2006) reported a WAT in which participants acted as their own controls, by responding to the same cue words in their L1 (Danish) and L2 (English). In this study, all cue words were high frequency words. Paradigmatic and syntagmatic responses were coded similarly to Wolter (2001), a further two categories were phonological and other associates. The other response category included idiosyncratic, affix change, and chained responses. In contrast to other studies, they found that participants gave more syntagmatic associates overall. They also produced more syntagmatic and paradigmatic associates in the L1 than in the L2, and more other and non-responses in the L2. Thus, although paradigmatic associates were more frequently given in the L1, this was also true of syntagmatic associates, challenging the idea of a shift in response type.

Finally, some studies have failed to find evidence that L1 speakers produce more paradigmatic associates than L2 learners. For example, in a study with L2 learners of Russian, Piranian (1983) did not find significant differences between the proportion of paradigmatic and syntagmatic responses across proficiency levels. She interpreted this as suggesting the syntagmatic-paradigmatic shift may be due to cognitive rather than language development. It should be noted that she classified response words as paradigmatic or syntagmatic purely based on word class. In addition, she did not consider the quality of the connection beyond responses that were only connected phonologically to the cue words. These two analytical choices means that the number of syntagmatic associates may have been underestimated and it is unknown whether L2 learners produced more response words that held weak or idiosyncratic associations with the cue words or whether this varied with proficiency. In this line, Meara (1978) also found L2 learners of French produced predominantly paradigmatic associates, although WAs were classified as paradigmatic, syntagmatic, or phonological, regardless of whether a meaningful link could be determined, such as “man-snail”.

Not all studies using the paradigmatic-syntagmatic coding system with L2 learners have collected a single response to each cue; some have collected multiple responses (Nissen and Henriksen, 2006; Zareva, 2007). These studies varied in their treatment of first and subsequent responses. Nissen and Henriksen (2006), described above, based their analyses on the first response words, reporting only that second responses showed similar patterns. In contrast, Zareva (2007) only reported results based on all three responses collected to each cue word. She compared L1 speakers with advanced and intermediate L2 learners, finding that all groups produced mostly paradigmatic associates. The proportion of paradigmatic associates compared with syntagmatic ones tended to increase with proficiency, but the differences between groups were not significant. Neither of these studies analysed responses based on their position (first, second response, etc.), an aspect of WA behaviour that is explored in depth in this thesis.

In sum, the evidence suggests that increasing proficiency in L2 learners does not lead to a syntagmatic-paradigmatic shift in WA behaviour. However, most research has found that L2 learners produce more phonological or form-based associates than L1 speakers, as a function of proficiency level (depending on the study, this category includes words with similar sounds but different meanings, such as “lose: loose”; affix changes, such as “lose: loser”; and/or grammatical inflections, such as “lose: lost”). This suggests that, at initial stages of learning, word forms may be prominent within the mental lexicons of L2 learners, either in the creation of relations between words or accessing word knowledge. L1 speakers produce mostly meaning links, with few phonological associates, although this type of association does not disappear entirely.

In addition, research with both L2 learners and children suggests that the tendency to produce mostly paradigmatic responses, found in early studies with L1 speaking adults, may have been, in part, a consequence of the cue words chosen. As noted above, early research often used cue words from the Kent-Rosanoff word list, which include a lot of contrasting nouns (e.g., “king”, “girl”) and adjectives (e.g., “soft”, “black”), which tend to elicit a dominant paradigmatic response. There may be a syntagmatic-paradigmatic shift in the WA behaviour of children, but it does not seem to occur in L2 learners. This suggests that any shift may reflect cognitive development in general rather than being specific to language development. Finally, it is not yet known whether the proportion of paradigmatic and syntagmatic responses changes when multiple associates are produced to each cue word, a possibility that is explored in Chapters 6 and 7.

As noted in Section 2.1.2, an alternative coding system to the paradigmatic-syntagmatic system, which has also been applied in multiple studies, is the meaning-, position-, and form-based system developed by Fitzpatrick (2006, 2007, 2009). This coding system was used with L2 learners (2006) and bilingual speakers (2009). It should be noted that in these two studies Fitzpatrick avoided concrete nouns and selected all cue words from the academic word list (AWL; Coxhead, 2000), which contains high frequency words in academic texts that are not within the first 2000 highest frequency words. Some examples of cue words used in Fitzpatrick (2006) are “income, sequence, emerged, hierarchical, notwithstanding”. Such cues were chosen to avoid high frequency nouns, which may influence responses (as noted above, many of the Kent-Rosanoff cue words elicit one highly dominant paradigmatic associate). However, the impact of using specialised language on WA behaviour has not been explored. It remains possible that the proportion of each type of response given by participants was influenced by the cue words all being representative of academic language, especially as no general language words were added to the task.

In the first study using this coding system, Fitzpatrick (2006) investigated its ability to capture differences between L1 and L2 speakers. She collected one response per cue for 60 words from the AWL from 40 L1 and 40 L2 speakers of English studying at a Welsh university. She created four main categories and 17 subcategories for coding responses. Meaning-based responses were further divided into six subcategories, including synonyms, taxonomic relations, qualities and conceptual associations. Position-based responses were divided into five types of collocation. There were also four subcategories of form-based responses, covering affix changes and phonological associates, and two types of erratic association (false cognate and no identifiable connection).

Considering the main categories, she found that both L1 and L2 speakers gave mostly meaning-based responses. However, L1 speakers also gave more position-based responses than L2 speakers, whereas L2 speakers gave more form-based and erratic associates. Analysis of the subcategories showed that, within meaning-based responses, L1 speakers produced more defining synonyms whereas L2 speakers gave more associates with loose conceptual links. These findings suggest that some syntagmatic associates, namely those that are collocates, are indicative of a developed L1 lexicon, rather than a developing lexicon, as assumed in the notion of a paradigmatic-syntagmatic shift. They also suggest that responses based on meaning can capture differences between L1 speakers and L2 learners, if finer distinctions are made by using subcategories for coding associates.

Fitzpatrick (2009) gave two different lists of 100 cue words from the AWL to bilingual speakers of English (L1) and Welsh (L2) to test whether their response profiles were similar in both languages. In this study, she reduced the subcategories within the main meaning-, position-, and form-based categories to ten (note that this is similar to the system developed in this thesis; a detailed description of the categories used by Fitzpatrick is given in Chapter 6). She found that within-participant differences were significantly smaller than between-participant differences, showing that individual response types are stable across different languages, as well as time (as demonstrated with L1 speakers in Fitzpatrick, 2007). Finally, she found that response type similarity was positively correlated with L2 proficiency, indicating that participants´ L2 response pattern was more similar to that of their L1 as their L2 proficiency increased.

In two more recent studies, Fitzpatrick and Izura (2011) and Fitzpatrick et al. (2015) introduced “dual-link” responses, that were associated to the cue by form and meaning or by meaning and position (collocation), in line with Osgood et al.’s (1954) proposal for coding WA data (see Section 2.1.2). Fitzpatrick and Izura (2011) did not use subcategories to code responses, although they used two meaning-based categories (synonym and taxonomic associates and other conceptual associations). In that study, L1 Spanish speakers completed two WATs, in Spanish then English. They gave mostly conceptual responses in both languages. The largest difference was that participants produced a greater proportion of collocations on the WAT in their L1.

However, the total number of responses was not reported and many L2 cue words and responses were removed prior to analysis, so there may have been further absolute differences.

A further indication of the importance of Fitzpatrick´s coding system is that it has been adopted by other researchers for use in WA studies. For example, Kim (2013), used 12 specific categories based on Fitzpatrick's (2006) meaning-, position-, and form-based system (adapted to better capture Korean grammar). Kim compared the WA behaviour of heritage and non-heritage learners of Korean in the US, learners in Korea, and L1 speakers in Korea (40 in each group). Kim´s study is particularly relevant as she collected three associates to each cue (an approach adopted in the experiments reported in Chapters 4 to 7). She found that although all groups produced mostly meaning-based associates, non-heritage learners produced significantly fewer due to giving more erratic associates. Within the specific categories, she found L1 speakers produced more collocates and qualities than either learner group. Non-heritage learners produced fewer strong conceptual associates than the other two groups. The contrast in learner groups is interesting and suggests that greater exposure to a foreign language outside of a classroom context may lead to more native-like conceptual associates. However, Kim used only 10 nouns as cues, which she stated were carefully selected from Korean language textbooks. This makes it difficult to generalize from her results to WA behaviour in general. In addition, she reported the data for all three responses, but did not consider whether the results were similar if only first responses were analysed.

In sum, the meaning-, position-, and form-based system developed by Fitzpatrick (2006, 2007, 2009) has been shown to be a highly informative system for coding WA data. The three main categories enable direct comparison with the paradigmatic-syntagmatic coding system, although the meaning-based response category covers a larger range of response types than the position-based response category. The subcategories within each main category offer a systemic method for coding responses in greater detail, which has revealed interesting differences between L1 and L2 WA behaviour which were not captured by the paradigmatic-syntagmatic distinction. However, the finding that individuals differ in the types of association they access most frequently questions whether the WA behaviour of children or L2 learners can be reliably compared to the response types produced by L1 speaker adults. This raises doubts about whether coding WA data can provide a more satisfactory benchmark for comparison than response norms. Despite this caution, coding WA may be a useful way to understand the developing mental lexicon, without necessarily assuming that becoming more native-like should be the target of L2 WA (Fitzpatrick, 2006).

2.2.3 Coding Word Associations: Summary

The research reviewed above shows that coding WA data may reveal insights into the mental lexicon and its development in first and second language learners, despite the inconclusive findings (Fitzpatrick & Thwaites, 2020). The main coding system used, distinguishing between paradigmatic and syntagmatic responses, has proven difficult to apply, possibly in part because these categories were not originally intended to be mutually exclusive (Osgood et al., 1954). The most comprehensively developed alternative coding system is the meaning-, position-, and form-based system developed by Fitzpatrick (2006, 2007, 2009). One of the main strengths of this system may reside in its use of main categories and subcategories, enabling a broad and detailed analysis to be performed simultaneously. Both coding systems were developed based on linguistic theory (although Fitzpatrick explicitly recognised the need for psycholinguistic interpretation of WA data). This approach minimises consideration of conceptual influence on WA behaviour. Yet the conceptual association subcategory captured most associations when general English cue words were used (Fitzpatrick et al., 2015), as opposed to the academic cues on which the system was originally developed. Based on these observations, the aim of this thesis is to develop and test a WA coding system informed by current theory within cognitive science.

2.3 The Use of Word Association Tasks in Cognitive Science

This section reviews the impact of WA findings on psycholinguistic research and semantic network models. As described next, WA has been used in psycholinguistics as priming studies have shown that WA captures relevant information about the mental lexicon (Meyer & Schvaneveldt, 1971). WA data has also informed semantic network models created from distributional semantics based on corpus analysis. To explain this latter use, studies that compared WA and corpus data are also described here. In this review two issues emerged that form the central concerns of this thesis: (a) priming studies have provided robust support for the psychological validity of WATs, although this validity is not well understood, and (b) comparisons with corpus data question the extent to which WA should be considered a language task.

2.3.1 Word Association and Priming Studies

Within psychology, an early use of WA data was to create stimuli for priming studies (Lashley, 1951). In priming studies, participants are briefly shown an image, word, or sentence (the prime) and then have to respond to another image, word, or sentence, which is either related to the prime (target item) or not (a distractor or filler item). Priming occurs when, over multiple trials, reaction times across participants are statistically significantly shorter for primed compared to filler or non-primed items. Thus, it is assumed that the prime contains information that facilitates responding to the target.

In word priming experiments, a prime word is presented, subconsciously or consciously, followed by a target word, which either is or is not related to the prime. A robust finding from such studies is that when the prime is related to the target, it is responded to more quickly (for a review, see Hutchison, 2003). Common word priming tasks include lexical decision (deciding whether a word is a real word or a pseudoword), word naming, and semantic judgement tasks (such as whether the word refers to a living or non-living thing). An important commonality between WA and priming tasks is that both assume pre- or co-activation of multiple words, whether at a sub-conscious or conscious level (Lashley, 1951).

The first evidence that word pairs that were associated according to WA norms might prime each other was provided by Jenkins et al (1958). They gave participants one of four lists of 24 words to learn that were composed of word association pairs from Kent-Rosanoff norms (plus three filler words at the beginning and end). The strength of association (the proportion of respondents who gave the dominant response to each cue) was varied between the lists. The lists were randomised so that no cue-response pair were adjacent on the lists. Participants recalled significantly more words from the two lists that were composed of highly associated word pairs, and for all lists, participants showed a tendency to recall associated word pairs together. These results show that association according to WA norms facilitates recall.

However, the earliest priming experiment reported was a lexical decision task by Meyer and Schvaneveldt (1971). They simultaneously presented pairs of words and/or pseudowords. Participants were significantly quicker to judge whether a pair of words were both words when they were associated according to the Connecticut Free Associational Norms (Bousfield et al., 1961; cited in Meyer & Schvaneveldt, 1971), compared to non-associated word pairs. This suggested that associations between words facilitate processing and that meaningful relations are activated very quickly in the perception of words, even when they are unnecessary to perform the required task. These effects of word association knowledge highlighted its importance in cognition, at least at the level of individual word stimuli. They established the paradigm of WA-based priming studies, with a focus on the strength of the association between words or response stereotypy, as judged by norming studies.

Subsequent priming experiments have frequently compared priming effects of prime-target words related by semantic similarity (category or taxonomic relations) or association (i.e., primes are cue words and their targets are the dominant response on WA norms; see Hutchison, 2003 for a review). For example, in a series of priming experiments, Shelton and Martin (1992) found automatic priming for words from the same taxonomic category which were also highly associated according to WA norms (such as “bed-sheet”) but not for such pairs when they were not associated (such as “coat-dress”). They suggested this could be due to co-occurrence frequency, reflecting lexical rather than conceptual effects. However, the lack of priming for non-associated compound words (such as “fruit-fly”) would seem to contradict that interpretation. This issue was investigated by de Mornay Davies (1998) in a lexical decision task. He created prime-target pairs from associates which were also semantically related, such as “hand-glove”, and opaque compound words, such as “fox-glove”, which have a lexical relation but no semantic relation. He found that only the associates significantly primed responses, suggesting that the priming effect of associates is semantic or conceptual, not purely lexical.

Although many studies have operationalised semantic similarity as shared category membership, some studies have considered additional relations between words as indicative of semantic similarity. For example, Moss et al. (1995) reported three experiments aimed at elucidating priming effects of categorical coordinates (e.g., "airplane-train"), functional relations (e.g., "hammer-nail"), and script relations (e.g., "restaurant-wine"; Schank & Abelson, 1977) for noun pairs that were also dominant cue-response associates or not (non-associated) according to WA norms. In the first two experiments, in which participants heard the words, they found similar priming effects across all three semantic similarity relations, which were stronger for associated word pairs. When the prime and target were presented with a minimal interstimulus interval, natural kind category coordinates (e.g., "dog-horse") and functional relations were effective primes even when the targets were non-associated. In contrast, there was only a priming effect for artifacts that were category coordinates and for script relations if the two words were associated. In the final experiment, with visual presentation, there was only a priming effect for coordinates when they were associated, in line with Shelton and Martin (1992). However, there was a priming effect of functional relations regardless of association. They interpreted their results as suggesting differing activation rates. Specifically, they suggested that functional relations and word association are activated more quickly than taxonomic and script relations. These findings not only confirm the importance of association within the semantic and/or lexical network, but also challenge the centrality of category relations for conceptual knowledge.

Finally, the robustness of WA to predict priming effects was confirmed in a large-scale priming study, involving 768 participants, the Semantic Priming Project, described in Hutchison et al. (2013). This project used the USF norms (Nelson et al., 1998) to construct highly, weakly and non-related prime-target pairs. The authors identified 1661 dominant associates (highly related targets) that were also given as an associate by at least two people to a different cue word in the USF norms (weakly related targets). They used these words in a lexical decision and a speeded naming task to construct a large behavioural database of priming effects. Although the priming effects were smaller in this study than in previous studies, both highly and weakly associated primes exerted a priming effect in the lexical decision task.

In a detailed meta-analytic review of priming experiments, Hutchison (2003) analysed semantic and association priming effects observed within a variety of conditions. He concluded that the strongest and most reliable priming effects had been shown with prime-target pairs that shared featural properties or were dominant cue-response pairs on norms lists such as the USF norms (Nelson et al., 1998).

Hutchison demonstrated that these priming effects were more robust than those based on perceptual similarity, synonym, or taxonomic category. Of note, Hutchison also argued that neither association strength nor semantic similarity are “pure” measures of priming effects between two words, as it is almost impossible to separate the two effects. Together, these findings from priming studies suggest that WA accesses conceptual semantic knowledge to a greater extent than, or at least, simultaneously with, lexical knowledge. In other words, they provide strong evidence that WA is not purely a language task, challenging the appropriateness of applying the paradigmatic-syntagmatic coding system to analyse WA data, given that this was devised as a purely linguistic coding system (Osgood et al., 1954).

2.3.2 Word Association and Language Use

The evidence from priming studies calls into question Osgood et al.’s (1954) assumption that WA responses solely reflect transitional probabilities in language. As discussed in Section 2.1.2, this assumption was the motivation for coding WA data at a time when conducting and analysing WATs was easier than creating and analysing corpora. Now that the opposite is true, an obvious issue to explore is the extent to which WA data is similar to corpus data. As examined next, the evidence from corpora also challenges the notion that WA is primarily a language task.

The first comparison of corpus collocations and WA norms was conducted by Church and Hanks (1990). They highlighted the subjective nature of WA and proposed an alternative measure of word co-occurrence from corpora, which they called the “association ratio”, based on the words’ mutual information score. Mutual information is the probability of the co-occurrence of two words given the probability of their individual frequencies of occurrence. Although corpus data is more appropriate for Church and Hanks’ aims as lexicographers, the one comparison given, for the word “doctor”, indicated that there were large differences between the two types of data. For example, “nurse” is the dominant WA response, whereas “dentist” had the highest association ratio for words following doctor in their corpus (“honorary” had the highest score for words preceding doctor).

Some subsequent authors (Wettler & Rapp, 1993; Wettler et al., 2005) asserted that WA behaviour resembled information from corpus collocations. However, this claim is based on the performance of individual participants, not a comparison of corpus and WA data. Both studies compared corpus data to the dominant associate given to the 100 Kent-Rosanoff cue words. Wettler and Rapp used the Minnesota norms (MN; Russell & Jenkins, 1954), whereas in the later study, they took the dominant associates from the Edinburgh Associative Thesaurus (EAT; Kiss et al., 1973). Wettler and Rapp found that the dominant associate in the MN coincided with the highest frequency collocate (determined using a bidirectional 12-word span and a variety of corpora) for only 17 of the cue words. Although the quality of the corpora available at that time provides a partial explanation for this result, a later study did not show a dramatic improvement. When the same comparison was made using the BNC (Davies, 2004), the match increased to 29 cue words (Wettler et al., 2005). The average participant gave 28 dominant responses. These results demonstrate that corpus collocation and any one participant produce the dominant associate in the EAT with a similar frequency. In other words, the results indicate that corpora are equally reliable at producing the dominant associate to these particular cue words as any individual person (see also Rapp, 2002). Furthermore, 36 of the corpus-predicted dominant associates were not given by a single participant. The corpus data partially overlapped with the WA data, but it did not accurately predict responses.

A more recent, but very similar comparison, again found that corpus data matches individual WA behaviour but not WA data. Bel Enguix et al. (2014) claimed that word association behaviour can be simulated by graph-theoretical corpus-based measures. This assertion was based on a comparison between data obtained from the BNC and nearly 6000 cue words from the EAT, for which a single associate was collected from 100 people. However, averaged across all cue words, the highest frequency collocate was given as a response by just over 6% of people. This is similar to the number of times any one participant gives the same response as another participant. This shows that corpus data matches the commonality between individuals´ WA behaviour, but it largely fails to predict the dominant associates.

Likewise, Schulte im Walde et al. (2008) collected multiple WA responses to nouns and verbs then compared this data with a German newspaper corpus of 200 million words. Using a bidirectional 20-word span, they found that only about a quarter of all associates were collocated with their cue word at least 50 times in the corpus. However, most associates appeared once within this large span. Although they viewed this as evidence of similarity between WA and corpus data, it seems to point to greater differences. They also suggested that WA captures world knowledge not present in corpora.

Studies that have examined the ability of corpus data to predict collective WA behaviour have concluded that the two sources capture different information. For example, Mollin (2009) compared all responses to 30 random cue words in the EAT with all collocations to those cues that occurred at least five times in the BNC (using a bidirectional four word span to identify collocations). She found that nearly 50% of cue-response pairs in the EAT were not collocations in the BNC. Furthermore, when she compared only the cue-response pairs that were collocations in the BNC, there was no correlation between association strength and co-occurrence frequency (measured by mutual information score). She concluded that WA data is informative about the mental lexicon and semantic knowledge but not about language production.

Likewise, in a different approach to comparing WA in the EAT with collocations in the BNC, Kang (2018) also found limited similarities. He compared a much larger set of cue words (3177), composed of all nouns, verbs, adjectives and adverbs used as cues on the EAT which occurred at least 10 times per million words in the BNC. Given this large set of words, he limited analysis to the dominant associate to each cue word. He compared each cue-response pair to the 100 highest frequency collocations from within the same four word classes, setting the text range for determining co-occurrence as either the sentence or the paragraph. He found that only about half of primary WA responses were in the top 100 collocates. Furthermore, association strength was only weakly correlated with collocation strength. In addition, paragraph-level collocation provided a better approximation to WA data than sentence-level collocation. This suggests that the role of language in building links in the mental lexicon operates at the level of extended text or even at a discourse level, rather than at a local sentential or phrasal level.

Finally, in a recent, detailed comparison of WA and corpus data, De Deyne et al. (2018), found WA norms were a better predictor of psycholinguistic variables than word frequency. As described in Section 2.1.1., they have created the largest database of L1 speaker WA norms, the “Small World of Words” norms, available online in a variety of languages, including English (SWOW-EN; De Deyne et al., 2018). Importantly, in the paper presenting the SWOW-EN norms, they also compared these norms (for first responses only and all three responses) to the USF norms (Nelson et al., 1998), the EAT norms (Kiss et al., 1973), and a corpus-based measure of word frequency (from the SUBTLEX-US corpus of American film and TV subtitles; Brysbaert & New, 2009). Specifically, they calculated the correlation of each dataset with response latencies on lexical decision, naming, and semantic decision tasks.

De Deyne et al. (2018) found similar correlations between all four WA norms lists and response latencies on all three tasks, which accounted for additional variance after controlling for word frequency. Moreover, word frequency was a poor predictor of semantic decision response latencies after controlling for WA. This analysis not only confirms the importance of WA for predicting psycholinguistic effects, but also showed that WA is specifically capturing semantic processing. It also demonstrated the reliability of even relatively small WA datasets, consistency between WATs, and the equivalence of single or multiple associates (at least up to three). In addition, when they analysed all four groups of norms compared with semantic similarity judgements, they again found similar performance, with the best performing norms list being that composed of all three responses from the SWOW-EN. These results suggest WA yields greater insight into semantic knowledge than corpus collocation and that additional information may be generated by collecting multiple associates, as done in the experiments reported later in this thesis.

Finally, de Groot (1989) provided evidence that WA does not reflect statistical probabilities of language use. She collected WAs from 100 L1 Dutch speakers to 160 nouns that had either high or low imageability ratings and were high or low frequency. She found that high imageability nouns were responded to significantly more quickly and elicited a dominant WA response that was given by more participants. In contrast, cue word frequency had no effect on either response measure. In the same vein, Van Rensbergen et al. (2016) found that WA data outperformed much larger corpus datasets in predicting human judgements of the affective value of individual words.

The evidence from studies comparing WA and corpus data shows that Osgood et al.´s (1954) hypothesis that WA could be used as a proxy for language use has not been upheld. Although there are similarities between cue-response associations and corpus collocations, there are also significant differences. Moreover, the lack of correlation between association and collocation strength shows there is no frequency effect. The independence of WA behaviour from word frequency has also been shown by de Groot (1989), who further showed that, in contrast, the imageability of the cue words does influence responses. Corpus comparisons support the interpretation of WA based on priming studies: WA appears to reflect conceptual knowledge as well as distributional or lexical knowledge.

2.3.3 Modelling Semantic Networks: Word Association Vs. Corpus Data

A further connection between WA, word priming studies, semantic similarity (discussed in more detail in Section 2.4), and corpus studies of collocation is their use in creating semantic network models (for a recent review, see Kumar, 2021). A key issue in this area is the definition of semantics. In its original conception, the term *semantics* was proposed for the study of meaning in language, the "science of significations" (Bréal, 1897/1900), pointing to the referent, but distinct from it. However, semantic networks tend to be described as models of lexical and conceptual knowledge, conflating both types of semantic meaning. Moreover, as reviewed in Hutchison (2003; see Section 2.3.1), many studies contrast semantic similarity (categorical relations, properties, and sometimes synonyms) with associative similarity (strength of association in WA norms). Yet, two important assumptions underlying much WA research are that the mental lexicon can be modelled as a semantic network and that WA data reveals its structure (e.g., Fitzpatrick & Thwaites, 2020; Henriksen, 2008; Wilks & Meara, 2002). Key ideas in theories of semantic networks are reviewed below, before focusing on studies that have compared WA and corpus-derived semantic networks, to show how WA may contribute towards modelling the mental lexicon.

Both the term *semantic memory* and the idea that it can be modelled as a network were first proposed by Quillian (1967). His main aim was to develop a computer model of human semantic memory (e.g., Quillian, 1969). Quillian did not propose his semantic network as a comprehensive model of semantic memory, but rather a descriptive model, based on dictionary definitions of words and common or world knowledge (but lacking action schema and emotion). He suggested that words could be modelled as nodes (*types*) connected by different types of associative links to multiple other nodes (*tokens*), with each token being a property (attribute value) of the type. For example, the type "canary" is connected to the token "bird" by the relation "is a". In this model, semantic memory is a dense, broadly taxonomic network (*supersets* in Quillian's terminology) in which the information expressed by all the tokens related to a type constitutes the concept of that type. The more tokens that are shared by two types, the more similar those types are to each other. For example, the types "canary" and "robin" both belong to the superset "bird" and share multiple properties, such as "can fly", "has wings" (Collins & Quillian, 1969).

Although Quillian originated the idea of semantic memory, its standard definition was provided by Tulving (1972), in a paper contrasting semantic and episodic memory. He defined semantic memory as both "the memory necessary for the use of language" (p. 386) and "a rich multidimensional structure of concepts and their relations" (p. 391). In this way, lexical and conceptual semantic knowledge were conflated. In addition, although Tulving suggested perceptual features could inform semantic knowledge, both were described as essentially symbolic information. As discussed in Sections 2.5-2.7, these two assumptions about semantic knowledge have since been challenged and this issue is questioned through the WA experiments conducted in this thesis.

Quillian's (1967) model of semantic memory is also important due to its usage within one of the most influential theories of processing within semantic networks, spreading activation theory (Collins & Loftus, 1975). Quillian's original ideas are modified in this theory, but the central idea is retained: Concepts, represented by words, can be modelled as nodes, which are connected by links which represent categorical and featural relations, thus revealing the semantic similarity between words. Activation in the semantic network spreads through links between shared features (nodes), such that the intersection of activation reveals the shared properties between two concepts/words. Priming effects arise from such intersections as activation spreads across nodes within the network. Collins and Loftus are explicit that the importance (criteriality) and accessibility (strength and/or length) of the links between nodes are variable, so that these qualities of the shared properties determine the similarity of two concepts, not just the number of shared properties. Likewise, the absence of shared properties in determining that two concepts are dissimilar is also influenced by the importance and accessibility of the properties to each concept. In addition, spreading activation between nodes is considered to proceed in a parallel manner, such that multiple features can be simultaneously checked in semantic memory.

Spreading activation theory has often been used to explain priming effects (see Section 2.4.1), in the assumption that a semantically related word primes a target via spreading activation between shared features within a category-based hierarchical network (e.g., Meyer & Schvaneveldt, 1971). However, the general claim that conceptual knowledge is processed via activation spreading in parallel across a semantic network is independent of whether or not conceptual knowledge is formed exclusively of taxonomies and properties, as posited by Collins and Loftus (1975). In WA, the proportion of people who give a certain response to a cue could represent the strength of activation between those words (e.g., Fitzpatrick & Thwaites, 2020; but see Nelson et al., 2003, 2013).

As discussed below, WA data has been used as a benchmark to assess how well corpus-based semantic network models capture the structure of the mental lexicon, revealing similarities and differences between these two types of data. In WA research the strength and/or type of association between cue words and human-produced responses is analysed. In building semantic network models word co-occurrence data from large corpora are used to represent words as numerical vectors within a multidimensional space.

To provide an example of how semantic network models are constructed and can be applied to human data, one of the earliest and most influential models, Latent Semantic Analysis (LSA), is briefly described. In their LSA model, Landauer and Dumais (1997) used whole texts to determine word co-occurrence both within and across texts. Their model created direct associations between words that appear together in a text and indirect associations between words that occur in two different texts via words that were common to both. Subsequently, through the decompression of semantic space into fewer dimensions, higher-order representations of words were created, based on direct and indirect word associations relative to all the words in the model lexicon. Thus, the totality of associations in the network provides latent information about the position of any word in the semantic space.

Landauer and Dumais (1997) found that LSA approximated human word learning by updating relations between words encountered in text in relation to all word knowledge and the previous contexts (paragraphs) in which the words had and had not appeared. In addition, when their model “read” a paragraph of text, vocabulary gains were greater from indirect learning (the effect of the text on the model’s total word representations) than from direct learning of target words. This assumption in LSA that information about a word is derived from the occurrence and non-occurrence of that word in a particular context has received empirical support. Young children use both sources of information in learning novel word-referent associations, although adults use only positive (direct) information (Ramscar et al., 2013). These findings suggest that LSA models implicit vocabulary acquisition.

Semantic networks have been claimed to provide models of semantic memory that are superior to WA data. For example, Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996), was compared to empirical data from priming experiments by Lund et al. (1995). They used a bidirectional 10-word span to identify collocations, then converted the 70,000 highest frequency symbols (only half were correctly spelled words) into a matrix, from which they extracted the 200 most frequent collocations for each symbol. They then tested the semantic similarity of words in HAL (i.e., the distance between two word vectors in the multi-dimensional space) with the priming data reported in Shelton and Martin (1992; see Section 2.3.1) and with novel data they collected in a lexical decision task, using word pairs that were category coordinates (semantically related; such as “table-bed”), associated according to WA data (such as “mold-bread”; source not stated), or both semantically and associatively related (such as “boy-girl”). They found that semantic similarity as modelled in HAL predicted the priming effects observed by Shelton and Martin. From their own lexical decision task, Lund et al. reported priming for word pairs related by semantic similarity, which was increased by associativity, but not for word pairs which were related only by WA. They also found HAL produced a similar pattern of results, although the correlation between semantic distance in HAL and reaction time was low (.24).

However, the data reported in Lund et al. (1995) showed that average reaction times to associated word pairs were faster than to semantically related word pairs. In addition, the semantic distance in HAL was smallest for associated word pairs, suggesting that semantic distance in HAL was related as much to association as to categorisation. Thus, despite their claim that association is less important than semantic similarity for word processing, their findings do not appear to provide unequivocal support for this interpretation.

Other authors have adopted graph theoretic approaches to building semantic networks. The first such models were created by Steyvers and Tenenbaum (2005). Their work is of particular relevance, as they created and compared network models based on association, semantic categories, and semantic similarity. Association networks were created from the USF norms (Nelson et al., 1998), semantic category networks from Roget's Thesaurus (Roget, 1911), and semantic similarity networks from WordNet (Miller et al., 1990), a large lexical database in which words are grouped together in synsets, based on relations of synonymy, taxonomy (category), meronymy (properties), and antonymy (adjectives only). They found that all three sources produced properties associated with small-world and scale-free networks, such as sparse connections between words, the presence of a few highly connected (hub) words, and small path lengths between any two words (i.e., any two random words were connected via a few intermediate words). These findings demonstrated that semantic networks, formed by various means, are highly structured, supporting the notion that they capture features of the mental lexicon.

Steyvers and Tenenbaum (2005) further showed that this small-world structure is not shown by LSA networks, suggesting corpus-based distribution of words may not provide a good model of the mental lexicon. Moreover, in a comparison of each network with a model based on age-of-acquisition (AoA) norms (the average age at which children learn words), the WA model performed best. Although they do not provide an explanation for this, one hint is provided by the hub words, which showed little overlap between networks. Four of the five most highly connected hub words in the WA network were nouns ("food, water, money, car", plus "good"), whereas the hub words in the other two models were mostly verbs ("break, cut, run, make" and "clear" in the thesaurus model and "cut, hold, set, turn" and "light" in the WordNet model). There is robust evidence that very young children learn nouns more easily than verbs, especially in "noun-friendly" languages such as English (for a review, see Waxman et al., 2013). Moreover, AoA has been shown to produce independent and typically stronger effects on word processing speed than word frequency across a variety of tasks, including picture and word naming and lexical decision tasks (for a review, see Juhasz, 2005). Thus, the correspondence between WA and AoA suggests that WA may reveal hub words that play a vital role in the structure of the mental lexicon.

Similarly, Ma (2013) found that WA data can improve the connections between words in WordNet. As shown by Steyvers and Tenenbaum (2005), the corpus-based distributional relations on which WordNet was built create extremely sparse networks. Ma considered two ways of incorporating WA data with WordNet synsets. She used a calculation of association strength from the USF norms (Nelson et al., 1998) and collected multiple associations to a set of words within the WordNet synsets to disambiguate word senses. She showed that WA data improved the strength of the semantic connections between pairs of words, providing a better model of the mental lexicon.

Baayen et al. (2019) created a dynamic model of the mental lexicon by training graph networks on co-occurrences of words in texts. They were able to train networks to process and produce words accurately, without stable morphological units or representations. Instead, meanings and forms dynamically mapped onto each other such that in comprehension word forms reconstruct the semantics and in production the semantic memory reconstructs the required word form. Although they recognised that their approach limited their model to an approximation of the mental lexicon, the huge advance of this model is that it does not require lexical semantics to be stored in a dictionary-like manner. This makes the model much more realistic as it allows representations to continuously change over time.

In sum, despite the huge growth in computational power, the data from corpus-trained models of language does not coincide with that from WA. There is no reason why they should, given this is not their aim, but it further suggests that the associations people form between words are not solely a result of their exposure to texts. More importantly, such models have not been shown to interpret word meanings in any human-like manner. Their strength comes from the huge amounts of data they are trained on (the most recent model, GPT-3 [Brown et al, 2020], was trained on corpora totalling over 500 billion words). This would appear to be very different to the human capacity for language based not only on experience of language but also the socio-environment and conceptual knowledge. However, the processes underlying WA are still poorly understood (Kumar et al., 2022). A coding system for WA data informed by current cognitive psychological theory might be able to contribute towards this endeavour.

2.4 Theories of Conceptual Knowledge

The literature reviewed above on the use of WA in cognitive science indicates that WA is informative about the structure of the mental lexicon, yet the nature of the semantic knowledge accessed in WATs remains unclear (see also De Deyne et al., 2018). There is no clear understanding of why WA data produces effective primes. Frequently given cue-associate pairs often do not appear as collocates in corpora. Semantic networks created to model the mental lexicon have incorporated links based on taxonomic, associative, and linguistic knowledge, but have not illuminated why there appear to be differences and similarities between these three types of knowledge. Moreover, a common theme emerged that questions whether WA is purely a language task, an assumption which may have arisen, in part, due to Osgood et al.'s (1954) proposal to classify associations in purely linguistic terms, as paradigmatic and syntagmatic associations. It seems plausible that the meanings accessed through WA do not only reflect implicit knowledge of which words co-occur in language, but also non-linguistic, or conceptual, knowledge. The aim of the coding system developed and applied in this thesis is to allow a systematic test of this hypothesis.

The theory that was chosen to guide the research conducted here proposes that conceptual knowledge is embodied, that is, concepts are multimodal representations grounded in perception, action, and bodily experiences (e.g., Barsalou, 2008). Before this theory is presented (in Section 2.6), this section provides an overview of prior theories of conceptual knowledge which remain influential. Due to space limitations, it is not an exhaustive review, but such information can be found elsewhere (e.g., Barsalou, 2016; Gabora et al., 2008; Komatsu, 1992; Margolis & Laurence, 2015; Murphy, 2002). Some of these theories are consistent with embodied theories and have been integrated rather than replaced (e.g., Rosch, 1999). My own view is that "weak" theories of embodied cognition (e.g., Barsalou et al., 2008) offer the best available theories of conceptual knowledge and its interaction with language knowledge. However, important challenges to this view are recognised (e.g., Mahon & Caramazza, 2008), as is the ongoing debate regarding the nature of conceptual knowledge (Mareschal et al., 2010).

2.4.1 The Classical View of Cognition, Concepts, and Language

As the field of cognitive science emerged in the 1940s and 1950s it was inspired by the notion of the computational mind (e.g., Varela et al., 1991, chapter 3), and the computer became the main metaphor for the brain. This is slightly ironic, given that the computer was inspired by Pitts' mathematical modelling of neurons and synapses (McCulloch & Pitts, 1943), an ambition which he realised was refuted by empirical evidence that the brain was not a logical machine and could not be reduced to digital (binary) processes (e.g., Gefter, 2015; Humphries, 2021; Von Neumann, 1958). Applying the computer metaphor, the brain was viewed as an information processing system that uses rules to manipulate symbolic representations. Information about the environment is processed by syntactic rules acting on meaningless symbols, which must then be semantically interpreted to realise "aboutness" or cognitive processes such as intentions, beliefs, and goals (Pylyshyn, 1980).

These theories have been labelled the Classical View of concepts (e.g., Murphy, 2002). This view necessitates mental representations of symbols and rules for how to act on those symbols. For example, comprehending a sentence requires mental representations of the semantic meaning of each word-symbol as well as an explicit symbolic representation of the rules (grammar) governing how they are combined. The emphasis on rules and computations necessitates logical relations between concepts and necessary and sufficient conditions to define them. In addition, information related to perception, action, cognition, and language is held to be processed in separate modules (Fodor, 1983; Fodor & Pylyshyn, 2015). Representations underlying cognitive processes are fundamentally distinct from those underlying perception. In this way, perceptual stimuli are processed in a purely bottom-up fashion; cognitive processes play no role in interpreting the raw input.

Due to his commitment to a modular brain, Fodor (1983) claimed that language and other cognitive processes have no access to perceptual information; they are encapsulated or autonomous processes. It may be possible to guess a word from the preceding linguistic context, but, Fodor argued, this does not mean that such background knowledge interacts with the auditory perception of speech. He claimed that linguistic forms are processed in a language module that has no access to semantic information. Fodor also suggested that subconscious associations can form between words due to their co-occurrence, but that conceptual knowledge cannot activate words at a subconscious level. Cognitive processes are assumed to be amodal operations, carried out via "the language of thought" over propositional representations (Fodor, 1975). Pylyshyn (e.g., 2003) claimed that any mental imagery is epiphenomenal to cognition, having no causal role in thought processes (for a detailed argument against this view, see Kosslyn et al., 2006).

This theoretical approach is aligned with the Classical View of concepts, in which concepts were viewed in terms of (verbal) definitions based on critical features (Murphy, 2002). An approximate summary is that concepts are represented as encyclopaedic knowledge or feature lists in one area of the brain and language as a dictionary and grammatical rules in another area (e.g., Quillian, 1969). As reviewed in Section 2.3.3, Quillian proposed that such information was stored in a network, in which concepts and words can be represented by nodes with labelled relations among them. An alternative explanation of semantic memory was provided by set theories, according to which concepts are stored as sets of elements or features with each feature exhibiting graded importance to the concept or category (Rips et al., 1973; Smith et al., 1974). At one end of the cline are defining features and at the other end are characteristic features. For example, "has wings" is a defining feature of the concept "robin" because it is a defining feature of the superordinate category "bird", whereas "seen in winter" is a characteristic feature of "robin" and is not true of all instances of birds. In sum, categories and features are important in both network and set theories of conceptual knowledge and are viewed as symbolic, language-like definitions (Fodor, 1975; Tulving, 1972). One of the earliest and most influential challenges to theories within the Classical View came from Rosch's prototype theory, as described below.

It should be noted that some more recent theories incorporating the view that language is symbolic and modular nevertheless do not rely on the notion that thought is language-like. For example, the MOGUL framework (Sharwood Smith, 2009) or Acquisition by Processing theory (Truscott & Smith, 2004) are modular approaches to language and cognition, but do not claim that words are discrete items produced within one module. Instead, words are viewed as phonological and syntactical modules connected with conceptual structure, which interact to produce linguistic meaning. However, such theories are more concerned with syntax than explaining what makes words meaningful and lack neural implementation. Other authors differentiate between natural language and language of thought, yet hold that a symbolic, logical, and syntactic language of thought, in some form, is the basis of human cognition (e.g., Dehaene et al., 2022) or even cognition across all species (e.g., Mandelbaum et al., 2022). As discussed below, my view is that symbols must be grounded to be meaningful (Barsalou, 2008; Harnad, 1990), yielding two complementary semantic systems (e.g., Barsalou et al., 2008), both of which may be accessed in WATs.

2.4.2 Prototype, Exemplar, and Causal Theories

The key difference between the Classical View described above and prototype, exemplar, and causal theories is that these latter theories all propose similarity is based on perceptual features. Based on evidence from categorisation tasks and priming studies, Rosch (1973, 1975) developed prototype theory, which posited that people form concepts based on similarity rather than definitions. She claimed that each member of a category or instantiation of a concept is a good fit to the extent it shares attributes with other members of the category and does not have attributes typical of different categories, in other words, to the extent it matches the prototypical member of the category. However, she argued that no single attribute is necessary or sufficient to determine category membership, similar to the view proposed by Wittgenstein (1953/2009).

Prototypicality, or semantic structure, was further refined as being determined by overlapping attributes or family resemblances, based on features and functions of category members (Rosch & Mervis, 1975). For example, cars are more vehicle-like than bicycles or tractors because the prototypical vehicle has four wheels and is used to transport people. Overall, research by Rosch and others has consistently shown that people exploit prototypes to facilitate cognitive processing of instances within categories.

This line of research was one of the first systematic empirical approaches to understanding conceptual knowledge. It has been hugely influential not just as a theory of conceptual knowledge, but also applied to lexical knowledge (Nerlich & Clarke, 2000). However, conceptual knowledge is largely conflated with category knowledge based on features and functions, which it will be argued below may not capture the full richness of concepts and is difficult to apply to abstract concepts.

Exemplar theory (e.g., Medin & Schaffer, 1978) represented an important challenge to prototype theory. The main claim of exemplar theory is that categories are formed based on experiences with instances from that category. Each new member of the category or concept is compared to all previously encountered instances, or features of them that are pertinent to the current task, enabling identification by analogy.

One way to unite prototype and exemplar theories was suggested by Murphy (2002). Based on his comprehensive review of experiments in support of both theories, he proposed that prototype and exemplar theories are complementary. Individual examples are most important in learning a concept or category and prototypes are more important in their later usage in guiding behaviour. In other words, the formation of a concept or category consists of the abstraction from individual referents to idealised representations (Rosch, 1978). However, in later work, Murphy (2016) argued that learning exemplars cannot result in the hierarchical structure that is inherent in (taxonomic) categories. In addition, he claimed that abstraction from exemplars is impossible, so previously encountered instances are not informative of current occurrences. In contrast to these objections, Gärdenfors (2014) suggested that each time an exemplar is experienced it modifies the prototype slightly. Thus, exemplars do not get abstracted or stored, but nonetheless influence prototype representations.

Various theories have emphasised the role of causal beliefs in conceptual knowledge and how they guide theories about the world (e.g., Johnson-Laird, 1983; Fodor & Pylyshyn, 2015). For example, Heit and Rubinstein (1994) suggested that people's similarity judgements are constrained by their theories or causal beliefs, which determine the relevant properties for categorical decisions. Murphy and Medin (1985) argued that concepts are formed from our theories about the world, emphasising that conceptual knowledge could not be formed solely by the internal structure of concepts, but depends also on the relations between different concepts, which enable concepts to be flexibly combined (Lake & Murphy, 2023). Similarly, Ahn (1998) proposed the causal status hypothesis, that "people have a bias toward weighting features that serve as causes of other features more than their effects" (p. 138). She suggested causal features are given importance as they provide better support for inductive reasoning. This argument has since been expounded at length by other causal theorists (e.g., Pearl & Mackenzie, 2018; Sloman, 2005). The key idea is that causal concepts enable organisms to predict their environment and choose optimal behaviours for interacting with it (Murphy, 2002; Sloman, 2005).

One objection that has been raised against causal theories is that people's explanatory knowledge is frequently incomplete. For example, Rozenblit and Keil

(2002) showed, in a series of experiments, that not only do many people have incomplete theories of how complex objects function or natural phenomena occur, they are also unaware of their lack of understanding (see also Sloman & Fernbach, 2017). Rozenblit and Keil (2002) concluded that we may rely on our perceptions of objects and natural kinds and schematic knowledge, rather than on detailed internal representations. However, other authors have suggested that causal theories and mental models do not need to be complete or entirely accurate in order to successfully guide behaviour (e.g., Johnson-Laird, 1983; Sloman & Fernbach, 2017). Importantly, causal theories can incorporate prototypical knowledge of features and functions (e.g., Sloman, 2005), positing that causal beliefs about such features and functions are prioritised.

Prototype, exemplar, and causal theories have all made important contributions to understanding conceptual knowledge. However, this research has focused on categories of entities, such as animals and vehicles, rather than abstract concepts or events (Komatsu, 1992). Thus, categorical relations and features of concrete entities are emphasised, rather than contextual relations. Research demonstrating the importance of context and thematic relations in conceptual knowledge is reviewed next, before showing how this can be incorporated in embodied theories (Section 2.5).

2.4.3 Context Effects on Concepts

One aspect of all the theories of conceptual knowledge reviewed above is that concepts are typically described as having defining and characteristic features or core and peripheral attributes. One proposed reason for this is that the role of concepts and categorizations is to simplify representations of our environment. It is simpler to focus on core, central attributes rather than attributes closer to fluid and undetermined boundaries (Hampton, 2007). This idea was reframed by Barsalou (1982), who proposed that concepts have context independent and dependent properties. He suggested that properties that define or distinguish an entity are context independent and are automatically activated by the word expressing the concept. In contrast, context dependent properties are only consciously retrieved when activated by a supporting context (see also Tversky, 1977). Such characteristic or peripheral attributes are not inherently represented as part of conceptual knowledge and so vary both between individuals and within any individual over time and/or context (Barsalou, 1992). Carefully controlled tasks conducted to test theories of conceptual knowledge minimise contextual information (e.g., Rosch, 1975; Smith et al., 1974), which may have led to an emphasis on stable features of concepts rather than dynamic changes (Anderson et al., 1976).

One influential study that showed context effects on conceptual knowledge was reported by Barsalou (1983). He showed that, when provided a context, ad hoc categories or concepts, such as “things that could fall on your head”, have similar structure to taxonomic categories, such as “fruits”, and can be used by people to organise information and aid memory. He suggested that taxonomic categories are not privileged in our conceptual or semantic networks due to inherent structure but rather due to their entrenchment in memory. Likewise, Dieciuc and Folstein (2019) have argued that prototype or typicality effects in categorisation are based on structural typicality, related to accessing a core conceptual representation in long-term memory. In contrast, functional typicality, as found in ad hoc categories (Barsalou, 1983) is derived from processing the concept in working memory, where it is integrated with situational or contextual information (see also Barsalou, 2021; Rosch, 1999).

Of note, in later work, Rosch proposed that prototypes are context-dependent reference points in cognition formed from sensorimotor interaction with the world (Gabora et al., 2008; Rosch, 1999). For example, she argued that prototypical category members could change through categorical combinations (Rosch, 1999), such that, although a car is a prototypical vehicle, a tractor is a more prototypical farm vehicle. This is in line with Tversky’s (1977) argument that prototypicality is relative and multidimensional, such that category members with more features that are common across many members of the category and fewer distinctive features are more salient or more prototypical members. Importantly, the relative importance given to shared and distinctive features is task-dependent. Thus, prototype theory can be interpreted as an explanation for how people subjectively form flexible categories grounded in objective real-world structure. As such, prototype theory can be seen as a precursor to theories of embodied cognition, which are reviewed in Section 2.5 and provide the basic assumptions for the research presented later in this thesis.

The establishment of context effects on conceptual knowledge is considered relevant to this thesis due to the questioning of the assumption that WA is purely a language task (see Section 2.3). WATs, and the tasks used in priming studies that have indicated the psychological validity of WA, involve the presentation of individual words. One possible explanation of the differences between WA and corpus data is that WATs do not completely capture context effects related to real-world processing tasks, as discussed further in Chapter 8.

2.4.4 Thematic Relations in Conceptual Organisation

The theories and research reviewed in this section so far have focused on categories of entities, such as animals and vehicles, rather than abstract concepts or events (Komatsu, 1992; Murphy, 2002). However, thematic knowledge might be equally important in structuring concepts (for a review see Estes et al., 2011). Thematic knowledge is experience-based knowledge, derived from the situations in which instances of concepts are encountered. It includes knowledge such as instrumental (action) and situational (event) scripts (e.g., scripts for sweeping a floor or dining in a restaurant; Schank & Abelson, 1977). In this vein, Fillmore (1976) coined the term *semantic frame* to refer to word knowledge gained from meaningful contexts, created from experience of the world and from the co-text around novel words (Fillmore, 1982).

Thematic knowledge has tended to be viewed as less structured or abstracted and thus inferior to taxonomic knowledge (Murphy, 2002). In line with this assumption, research into thematic relations has tended to suggest that they are a feature of children's concepts that are later replaced by taxonomic relations (for a review, see Sloutsky & Deng, 2019). However, it has been shown that educated adults continue to use thematic relations extensively (Lin & Murphy, 2001) and that primes representing both types of relation can facilitate picture naming (Ovando-Tellez et al., 2022). Moreover, De Deyne et al. (2015) showed that thematic knowledge may be important in the organisation of lexical relations. They constructed a semantic network from WA data, then extracted hierarchical clusters from it based on connective density between words. They found words were clustered based on thematic structure at lower levels (rather than predominantly taxonomic structure).

The suggestion that the development of taxonomic relations may originate in event-related thematic relations was first made by Nelson (1988). She called children's first categories "slot-filler" categories, as the concepts in such categories fill the same slot within repeated scripts or events. For example, she suggested that the category "foods" is learned from stable experiences of eating breakfast, lunch, and dinner foods. She further suggested taxonomic categories develop from slot-filler categories with experience of language. In this way, taxonomic categories are grounded in linguistic semantics and slot-filler categories in event-based conceptual knowledge. In more recent support of these claims, Sheng & Lam (2015) found that the number of words produced by young children within taxonomic categories (animals, clothes, and food) was significantly related to maternal education level. In contrast, the number of responses given within slot-filler categories (zoo animals and farm animals) was not significantly related, suggesting children learn the latter type of categories more easily or earlier. As Sloutsky and Deng (2019) argued, if semantic networks are constructed based on similarities between entities combined with knowledge of different domains of experience, this could give rise to behaviour that suggests hierarchical categorisation without the latter being an underlying organisational principle of such networks.

Luria (1976) suggested that taxonomic relations have to be taught, and as such tend to be a product of formal education. He conducted various studies in the 1930s with adults in rural areas of what is now Russia, who had received no or minimal schooling. He found that illiterate adults grouped objects according to thematic or situational relations, refusing to accept taxonomic relations as valid pairings (for example, participants stated that an axe goes with wood, as it can be used to chop trees, but denied any suggestion that it could be paired with a hammer as the two objects have different functions). In contrast, adults with minimal formal education could be coaxed to view a connection between objects such as an axe and a hammer as possible; they could be taught to conceive of a similarity between the two tools. Adults with basic literacy skills could use both thematic and taxonomic relations. Thus, he concluded that thematic relations were primary and taxonomic relations were learned through formal education. Of note in this regard, despite arguing that concepts are categorised on the basis of properties into hierarchical (taxonomic) sets, Collins and Loftus (1975) state that concepts are categorised via explicit instruction, rather than by features they share with similar kinds (p. 426).

In a recent review, Mirman et al. (2017) evaluated the evidence that semantic knowledge of concepts and words is based on taxonomic and thematic relations and that the two types of information can be distinguished by analysing behavioural and neural responses to stimuli. From behavioural findings, they concluded that taxonomic relations may be more important for living things whereas thematic relations are more important for inanimate objects. They also suggested that there may be cognitive differences between the relations. Thematic relations are activated more quickly than taxonomic relations and are used more by young children, although adults may show individual differences in preference for taxonomic or thematic relations. In addition, they found studies that indicate that there are cognitive costs associated with switching between taxonomic and thematic relations.

However, Mirman et al. (2017) concluded that the neural evidence for a distinction between the two types of knowledge was more controversial. They found some support for a division, with taxonomic relations potentially involving greater processing in the anterior temporal lobes and thematic relations in the temporo-parietal cortex. They speculated that such topological specialisation for processing taxonomic and thematic relations could arise due to different information being more salient. They suggested features such as shape and colour may be more informative for taxonomic relations and action and location for thematic relations. In addition, in a pre-registered direct comparison of the processing of taxonomic or featural relations and thematic relations using functional magnetic resonance imaging (fMRI), Geng and Schnur (2016) found limited support for activation of the anterior temporal lobes in processing taxonomic and feature relations and the temporo-parietal cortex in processing thematic relations.

In sum, although there is some robust evidence for the contribution of thematic relations in structuring conceptual knowledge, more research is needed in this area. Given the findings reported in De Deyne et al. (2015), WA might be a useful behavioural paradigm in this regard. One potential advantage of WATs is that this task places no restrictions on the type of knowledge accessed, in contrast to many of the tasks used in the research on conceptual knowledge reviewed above, which assessed category membership and features. In addition, the contested notion of a developmental syntagmatic-paradigmatic shift in WA behaviour (see Section 2.2) is paralleled by that of a thematic-taxonomic shift in conceptual knowledge (Estes et al., 2011).

2.4.5 *Dual Coding Theory*

The embodied cognition theory that became pivotal to the research conducted in this thesis is Language and Situated Simulation (LASS) theory, a dual processing theory proposed by Barsalou et al. (2008; see Section 2.7). This theory explicitly builds on Paivio's (1986) dual coding theory (DCT) of semantic knowledge. In contrast to the other theories presented in this section, DCT is a theory of conceptual and linguistic knowledge. The central tenet of DCT is that human cognition consists of two separate symbolic and modality specific representational systems which process different types of information: a verbal system for processing language and a nonverbal imagery system for processing sensorimotor representations (Paivio, 1986).

DCT, developed by Paivio since the 1970s based on empirical evidence, can be viewed as a connection between all the theories of conceptual knowledge reviewed in this chapter. In line with the Classical View, DCT assumes symbolic linguistic representations play an important role in semantic knowledge. However, DCT assigns equal important to analogue, imagistic representations, learned through sensorimotor experience of exemplars, but with a probabilistic structure, allowing for prototypical representations. In line with causal theories, Paivio also assumed that representations were goal-directed, enabling the construction of models of the world that can guide behaviour. DCT can also be viewed as a predecessor to embodied cognition theories, aligning with weak views of embodiment, in which cognition involves multimodal representations and symbolic language representations (e.g., Barsalou, 1999; see Section 2.6). Thus, DCT is important as its ideas are foundational to the weak embodied cognition theory that guided the research reported in this thesis (Barsalou et al., 2008; see Section 2.6) and due to the connections it forms between the theory used here and previous theories of semantic knowledge.

Paivio (1986) posited that the verbal and nonverbal representational systems were structurally and functionally independent but interconnected. He used the terms *logogens* to refer to verbal representations and *imagens* to refer to nonverbal representations. DCT is clearly influenced by the structure of verbal and visual entities, which he assumed must be reflected in their mental representations. Language is experienced as a linear sequence of arbitrary individual words and logogens exhibit the same structure and processing. Visual perception is experienced as continuous and synchronous affordances (Gibson, 1979/1986) in the environment (we perceive entities, their features, and the wider context in which they are embedded), thus imagens consist of this structure and processing (Paivio does not discuss sign languages, which allow for simultaneous as well as sequential processing; for reviews, see Fenlon et al., 2017; Sandler, 2017)

Three aspects of DCT (Paivio, 1986) are particularly relevant for this thesis. First, a key feature of DCT (Paivio, 1986) is the mapping of concrete and abstract words onto imagens and logogens. He posited that concrete words were represented through nonverbal imagery and verbal associations whereas abstract words were represented mainly through verbal knowledge and associations. He developed this idea further in collaboration with Sadoski (e.g., Sadoski & Paivio, 2013). Although Paivio avoided the term *concept*, he stated that processing could involve either type of representation in isolation, suggesting that he conceived of both types of representation as being meaningful (semantic). As noted above, other theories presented in this section focused on knowledge of concrete concepts. Paivio's DCT is distinct in this regard. His suggestion that language plays a central role in the representation of abstract concepts has influenced later theories (e.g., Borghi et al., 2019; Dove et al., 2022; Wang et al., 2010). Abstract and concrete words have been suggested to differ in the degree to which they reflect distributional (linguistic) and similarity (taxonomic) knowledge, which in turn could be mapped onto syntagmatic and paradigmatic knowledge, respectively (Crutch & Jackson, 2011).

Second, in his discussion of how representations are learned, Paivio (1986) posited that nonverbal representations typically precede verbal ones, which is similar to claims made within some embodied cognition theories (see Section 2.5). However, in contrast to such theories, he questioned whether imagery was necessary for initial language learning. He also held that once the word-referent relation was learned, such knowledge could be represented and processed entirely within the verbal system.

Third, Paivio proposed three different kinds of processing: the activation of verbal or nonverbal representations by stimuli of the same type; cross-activation or referential processing whereby a verbal stimulus activated a nonverbal representation or a nonverbal stimulus activated a verbal representation; and associational processing whereby a stimulus activated a related representation. Thus, in DCT, associations between words and other words and words and their referents are central to making meaning. Paivio stated that active representations are a function of the current context and past experience, such that the images and words that are activated in an individual at a particular moment are based on "associative probabilities" (Paivio, 1986, p. 63). In contrast to other theories, in which concepts are structured by semantic similarity and categorical relations, in DCT any kind of association between image or word representations is relevant to meaning and processing, thus also making DCT a potentially relevant theory for WA.

2.4.6 Neuroscientific Studies of Conceptual Representations

This section reviews neural evidence regarding conceptual knowledge from brain damaged patients and healthy participants. Importantly, much research in this area supports the key claim of embodied cognition theories that conceptual knowledge involves multimodal representations.

In many of the theories reviewed above concepts and categories are treated as being essentially equivalent. Likewise, in the review of priming studies above (Section 2.3.1), it was highlighted that such studies distinguish between words that are related by association and those related by semantic similarity, where semantic similarity is operationalised as category membership and shared features. As illustrated below, studies of patients with neurological impairments have tended to be interpreted as suggesting that features may play a more important role in organising semantic knowledge than categorical or taxonomic relations (e.g., Warrington and Shallice, 1984; but see Caramazza & Shelton, 1998).

In the earliest such study, Warrington and Shallice (1984) found that patients with severe bilateral damage to the temporal lobes showed more severe deficits in identifying living things than inanimate objects. They interpreted their findings as evidence for category-specific representations arising from the salient information that differentiates members within categories. They suggested that inanimate objects are processed according to their function or use, whereas living things and foods are processed based on their sensory attributes. Hence, the differential importance of functional and sensory attributes gives rise to distinctions that resemble a categorical distinction between inanimate and animate entities (see also Warrington & McCarthy (1987).

More recently, in an object picture naming study with participants with aphasia, Blackett et al. (2022) demonstrated that access to taxonomic and thematic relations for verbalisation can be differentiated in the left hemisphere (right hemisphere areas and connections were not assessed). They confirmed that taxonomic errors were related with damage to the anterior temporal lobe (e.g., Bozeat et al., 2000). However, they found thematic errors were associated with damage over larger areas of the cortex, including areas in the temporal and parietal lobe. They suggested that thematic knowledge might involve access to multimodal representations which are integrated with taxonomic (transmodal) knowledge in the anterior temporal lobes (Rice et al., 2015). This interpretation is in line with the notion that thematic relations reflect direct experience of the world whereas taxonomic or semantic relations represent more abstracted knowledge.

Other researchers have investigated how and where concepts are represented in the brain in healthy individuals. One of the main findings of such research is that concepts are represented across large areas of the brain, rather than localised to one specific area (Huth et al., 2012). In one of the earliest studies, using positron emission tomography (PET), Martin et al. (1995) showed participants black and white drawings of common objects. In three separate scans, participants either named the object, stated a colour associated with it, or stated an action associated with it (for example, when shown a pencil they might respond “pencil,” “yellow,” “write.” They found that different areas of the temporal lobe were activated in the colour and action naming tasks, both immediately anterior to areas that had previously been identified as processing colour and motion perception, respectively. This activation pattern was replicated using words as stimuli rather than pictures. In a subsequent PET study, Martin et al. (1996) compared naming of drawings of animals and tools. They found that images of tools activated areas involved in processing action, whereas animals elicited activation in visual processing areas. These findings further suggested that category-related impairments in naming objects may be a product of the importance of different attributes in constructing different kinds of concepts.

Based on these two studies, Martin (1998) suggested semantic information about an entity combines abstracted features of that kind of entity (typical shape, size, colour, etc.) as well as adaptive behavioural and affective responses towards it. Such information is spread over the brain close to the areas that process the corresponding perceptual and motor attributes and affective values. In this way, dispersed representations of entities enable a finite number of properties to be combined into an unlimited number of concepts.

Similarly, in an fMRI study, Bauer and Just (2017) found evidence that categorical concepts are differentially structured. They located differences in activation when participants thought about animate and inanimate objects (presented as word cues) that were either basic level concepts (such as bird) or typical or atypical subordinates. They found increased activation across sensorimotor areas of the cortex for all words. In addition, basic level concepts and typical subordinates activated anterior temporal areas (thought to represent abstract semantic information). Basic level concepts also elicited higher activation in language areas of the brain. They suggested that basic level concepts activate both abstract and concrete representations, whereas subordinate concepts activate primarily concrete (sensorimotor) representations.

Also using fMRI, Fernandino et al. (2022) found evidence that conceptual knowledge is formed through experience of the world. They recorded activity patterns in a large set of brain regions that have been identified as involved in concept representations while people made familiarity judgements about 300 lexical concepts (animate and inanimate objects, events, and abstract nouns). They then compared this data with six models based on experiential, taxonomic, and distributional (corpus-derived) conceptual knowledge (two models for each type). The two experiential models dramatically outperformed all other models in capturing the neural representations of the set of concepts. In addition, only the experiential models predicted unique variance in the representational structure of the word-concepts in the semantic network. The best experiential model was created from participant ratings of various sensorimotor, spatial, emotional, and mental attributes (48 attributes in total) provided for 535 word-concepts (Binder et al., 2016). Fernandino et al. concluded that taxonomic information may not be central to conceptual knowledge, but only appears to contribute due to overlap with experiential information. In other words, categories emerge from experientially learned attribute clustering (see also Binder et al., 2016).

There is also neural evidence that the meaning of abstract and concrete words is processed differently. For example, West and Holcomb (2000) conducted an event related potential (ERP) study to explore processing differences between concrete and abstract nouns, which were presented as the final word in short sentences. The important finding for this thesis was the comparison between a semantic processing task requiring general knowledge about the final word and an imagery task, in which participants indicated the ease of forming a mental image. Participants responded significantly faster to concrete words than abstract words in both tasks, and to concrete words in the imagery task compared to the semantic task. The time course of the ERP data showed that the N400 effect[3] produced by concrete and abstract nouns was similar in both tasks, whereas a later, N700 component, showed greatest negativity for concrete words in the imagery task, suggesting it indexed strength of imagery. They interpreted their results as indicating early linguistic-based processing and later imagery-based processing, in line with dual coding theories (Barsalou et al., 2008; Paivio, 1986).

In sum, considerable evidence from neural studies suggests that sensorimotor information gained through direct experience and abstracted representations combine to structure conceptual knowledge (although see Handjaras et al., 2016, for a different view). This has led to various theories of conceptual knowledge that are compatible with (weak) theories of embodied cognition, which is the framework used in this thesis. Such theories have proposed that the anterior temporal lobes are a convergence zone where input from different modal areas of the brain (visual, auditory, olfactory, etc.) is integrated, giving rise to complete concepts (Damasio, 1989; Fernandino et al., 2016; Martin, 2016; Patterson et al., 2007; Tranel et al., 1997), possibly including verbal information for that concept (Rice et al., 2015).

2.4.7 Conceptual Representations and Word Association

To end this section, connections between theories of conceptual knowledge and word association (WA) research are highlighted. One commonality between all the theories of concepts reviewed above is that they highlight the importance of relations between concepts. In the Classical View, the relations between concepts are akin to the relations between words in language (Fodor, 1983); in prototype and exemplar theories they arise from shared features (e.g., Rosch, 1975); whereas causal

---

[3] In electroencephalogram (EEG) studies measuring ERPs the N400 component is a negativity that occurs about 250 to 500 ms after a stimulus is presented and which is widely considered to indicate semantic activation (Kutas & Hillyard, 1983; for a review, see Kutas & Federmeier, 2011). The amplitude of the N400 correlates strongly with the unexpectedness of the stimulus (surprisal) at the semantic level but not the syntactic level, suggesting it indexes the cognitive effort required to comprehend a word.

theories give prominence to causal relations (e.g., Sloman, 2005). As noted in Section 2.3.1, one of the main reasons for the interest in WA within psychology and cognitive science is the evidence for priming between associated word pairs, as taken from norms lists (e.g., Hutchison, 2003; Meyer & Schvaneveldt, 1971, Shelton & Martin, 1992), suggesting that WA reveals the relations that structure the mental lexicon.

Thus, to some extent, the debates about conceptual knowledge can be encapsulated in WA. As noted in Chapter 1, WATs are used to elicit the "first word that comes to mind" (Clark, 1970, p. 272), meaning they access a range of relations which can be defined in multiple ways, such as synonymy, meronymy, coordinate category members, and collocates. Ascertaining why different cue words elicit different response words and different types of association is one of the puzzles of WA (Fitzpatrick, 2007, 2009). Yet, whatever the associations are that are accessed, they represent just those associations that are most salient, given no supporting context (Moss et al., 1995). Moreover, the WA norms that have been used in priming studies owe their very existence to the consistency of these associations. This observation is the main motivation of the research reported in this thesis. What are the salient relations accessed by WATs and what kinds of (shared) meaning do they reveal?

The Classical View, prototype, and exemplar theories of conceptual knowledge assume that concepts are grouped into categories based on shared properties or features (e.g., Murphy, 2002). However, further evidence supports the importance of thematic or experiential relations in structuring concepts, in addition to features and categories (e.g., Estes et al., 2011). As suggested by Couchman et al. (2010), concepts may incorporate all these types of information to the extent that they provide useful structure for creating mental representations of natural kinds (categories) in an external world that the brain has no direct access to. The best approach to solving this inference problem (Friston, 2010) may be to use whatever is available. As such, there is no necessity for concepts, or mental representations, to take only one form or to result from only one kind of learning. Prototypes, exemplars, theories, thematic relations, socio-linguistic experience, and any other aspects of experience, may all contribute to conceptual knowledge.

This importance of meaningful associations of any type has received some empirical support. For example, in three different perceptual discrimination tasks

Afiki and Bar (2020) showed that association affects performance. Prior to each judgement task, they presented participants with images of associated or non-associated objects, as determined by WA norms, or, in the control condition, no images (task only). Across all three tasks, they found that the presentation of images of unrelated objects led to worse performance, compared to no images or associated images. They interpreted these results as indicating increased cognitive load in the no-association condition, caused by participants continuing (subconsciously) to seek a connection. They claim that associations, based on contextual co-occurrence rather than semantic category, are pervasive throughout cognition, as they "reflect the need to find coherence in our environment" (p. 1). Similarly, Trapp et al. (2015) found that associated images were preferred to non-associated images, both when the images were meaningless shapes or known objects. Although these studies used images not words as stimuli, they highlight the importance of understanding the role of associations in cognitive processing.

2.5 Embodied Cognition Theories

In this section theories of embodied cognition and their main claims are reviewed. First, a brief overview of the development of embodied cognition theories is provided. Then, the basic premises of embodied cognition theories are outlined. Next, behavioural and neural evidence is presented that supports embodied theories of concepts and language. Finally, evidence against embodied cognition theories is considered, which is argued supports the adoption of a weak version. Weak embodied cognition theories (such as the one that informs the research conducted in this thesis) posit that semantic knowledge involves multimodal conceptual representations and (symbolic) linguistic meaning. Such theories are discussed in Section 2.6.

2.5.1 The Development of Embodied Cognition Theories

The expression *embodied cognition* was first used by Varela et al. (1991) in their seminal work on this theory. They described the embodiment of cognition as referring not just to its embeddedness in biological experience (i.e., the body and its perceptions and actions), but also to social and cultural history and language experience (p. 149/173). Their ideas drew directly on Merleau-Ponty's (1945/2005) phenomenological account of experience, in which he emphasised the embodied subject whose thought is shaped by lived-in experience of the body and the world (see also Gibson, 1979/1986; Maturana, 1970). Varela et al. (1991; see Thompson, 2007, for an updated version of this work) claimed that organisms enact the world around them, rather than representing it symbolically, such that organism and world co-create each other (cognition emerges from this enactment). Embodied cognition theories can be viewed as a return to the idea of perception informing the brain, in which inform meant to shape or mould (Capurro & Hjørland, 2003). In this thesis, embodied cognition is interpreted as the theory that an individual's knowledge is constructed via their sensorimotor and internal (bodily) experiences which occur in a socio-environment created by the interactions of other humans with the world and with each other. Some authors, notably Barsalou (e.g., 2008, 2016), prefer the term *grounded cognition*, to emphasize that cognitive processes are coupled with the body and the socio-environment. In this thesis, both words, *grounded* and *embodied*, are used interchangeably to describe this theory of cognition.

Theories of embodied cognition can be seen as an extension of theories that posited that mental imagery underlies much of cognition, most notably the work of Kosslyn and Paivio, along with their collaborators (e.g., Kosslyn et al., 2006; Paivio, 1986; see Palmiero et al., 2019, for a review of this evolution). As documented in Kosslyn et al. (2006), he and his colleagues conducted various experiments over many years in an attempt to dispute Fodor's (1975) claim that cognition was based on the language of thought. Although Kosslyn et al. suggested that most humans rely largely on visual imagery, they highlighted that imagery can be in any sensory modality or involve motor imagery. However, they did not explicitly link their ideas to embodied cognition theories. As reviewed in Section 2.4.5 above, Paivio's (1986) DCT remains a highly influential theory in which conceptual knowledge is viewed as a combination of imagery and verbal representations.

Embodied cognition theories are tightly linked to language comprehension (e.g., Barsalou, 2008), making them an appropriate framework for analysing WA data. This may be partly due to two factors. First, one of the earliest arguments for embodied cognition was based on the role of concrete concepts in structuring abstract ones. In their theory of linguistic metaphor, Lakoff and Johnson (1980) claimed that metaphors are grounded in concrete concepts, often related to our bodies and how they interact with the environment, such as spatial orientation (Johnson, 1990). According to this theory, at least some of the content of abstract concepts is an extension of direct experience.

Second, a key motivation for embodied theories of cognition has been to resolve the "symbol grounding problem" (Harnad, 1990). Harnad argued that words or symbols cannot gain meaning solely from other symbols; they must be grounded in non-symbolic representations. A famous example of this problem is Searle's (1980) Chinese room thought experiment, in which he argues that even if an artificial intelligence (AI) was programmed to give correct answers in Chinese about a story written in Chinese, it would not understand the *intentionality* of the words it produced (what the words were about). According to Searle, this is sufficient to prove an AI does not have cognitive states and also that cognitive states require more than the capacity to process formal symbolic information (syntax), they also require semantic meaning. The solution that Harnad (1990) proposed to this problem is that meaning emerges from the juxtaposition of combinations of symbols (words) with iconic and categorical representations of real-world referents.

Moreover, theories of embodied cognition are arguably becoming the dominant theories within cognitive science (Spivey, 2023). Based on evidence reviewed below regarding such theories, a weak embodied cognition theory will be used in this thesis to develop a new coding system for analysing word association data. Research presented in Sections 2.3 and 2.4 was used to demonstrate that WA can provide insights into the mental lexicon but that this task would seem to access conceptual as well as lexical semantics. Therefore, in reviewing the evidence for embodied cognition theories, work related to the embodiment of language and conceptual knowledge is most relevant and is focused on in this chapter.

2.5.2 Embodied Theories of Concepts and Language

Within the perspective of embodied cognition, concepts are (subconscious) situated simulations of previous or imagined interactions with the socio-environment. Thus, the mental retrieval of a concept activates multimodal sensorimotor processing areas (e.g., Barsalou, 1999, 2008; Gallese & Lakoff, 2005). There are various theoretical frameworks which make different claims about the nature of embodiment (for reviews see Barsalou, 2008; Meteyard et al., 2012; Wilson, 2002; see Ellis, 2019, for a review related to foreign language learning). In strong views of embodiment, there are no neural representations, only perceptual-motor couplings through which the organism enacts its environment (e.g., Chemero, 2011; Thompson, 2007). In more moderate views, concepts are representations, but they are distributed throughout the sensorimotor areas, thus such areas are necessary (and sufficient) for concepts (e.g., Gallese & Lakoff, 2005; Zwaan, 2004). In weak views of embodied cognition multimodal information is integrated with lexico-semantic knowledge: Concepts are partially grounded by experience but language may contribute abstract knowledge and so aid in structuring concepts (e.g., Barsalou, 1999; Clark, 2006; Vigliocco et al., 2004).

In this thesis, a weak embodied cognition theory is adopted, in which conceptual knowledge is grounded in multimodal simulations which ground language, but it is also posited that language may contribute (symbolic) structure to conceptual representations (following Barsalou, 1999, 2008). In relation to the previously presented theories of concepts, this means thematic or event relations are hypothesised to structure concepts and semantic knowledge as well as features and taxonomic relations. This follows Rosch (1999), who suggested that (prototype-based) categorical relations are multimodal, dynamic, situated, and function-orientated representations, placing situational contexts at the centre of conceptual knowledge. It is also noted that one important entailment of an embodied approach is that concepts are dynamic (re)constructions. Each time we simulate a concept the underlying activation pattern in the brain will be slightly different, depending on the focus or function of the concept at that particular instance (Barsalou, 2009). In addition to effects due to the task, goals, and the individual's current psychophysical state on each instantiation of a concept, the structure of the concept itself will constantly evolve over the individual's lifetime (Connell & Lynott, 2014). Next, empirical evidence is presented that supports embodied cognition theories of concepts and how such multimodal representations may ground language.

2.5.3 Behavioural Evidence for the Embodiment of Language

Various priming studies have shown that multimodal properties that are typical of concepts can facilitate the semantic processing of language. For example, Stanfield and Zwaan (2001) asked participants to read sentences that described objects such as they implied a horizontal or vertical orientation, followed by an image of an object that either was or was not in the sentence and was or was not in the implied orientation. They timed the speed with which participants stated that the object was or was not in the preceding sentence. They found that average response latencies to objects that were present were significantly shorter when objects were in the orientation implied by the sentence than when they were in the alternative orientation. They concluded that participants formed (subconscious) perceptual images in comprehending the sentences. Similarly, Zwaan et al. (2002), using a sentence-picture priming task, found further evidence that people form mental images when reading. Participants read sentences involving objects that implied they were in one of two different states. For example, they read that an eagle was in the nest or in the sky, or an egg was in the fridge or the frying pan. They found that picture naming was faster when the depiction of the object agreed with the previous sentence. Both these findings have since been replicated as online experiments with large numbers of participants (Zwaan & Pecher, 2012). Such priming effects suggest people form subconscious images when reading and, more importantly, that such images have a measurable effect on language comprehension.

In addition to perceptual effects, mental processing of motor actions has also been shown to influence language processing. For example, Glenberg and Kaschak (2002) conducted a task in which participants judged whether a series of sentences were sensible or not. On the critical trials, participants read one of a sentence pair such as “Bill gave me the apple” or “I gave Bill the apple”. They had to respond by pushing one of three buttons arranged on a box from front to back, as the participant held the box. The middle button was pressed to make the next sentence appear, so that responding yes or no to judge whether the sentence made sense meant either moving their hand towards or away from their body. For half of the participants the farther button was yes, for the other half it was the nearer button. Decision times were significantly shorter when the direction of transfer in the sentence matched that of the response action than when it was in the opposing direction. This suggests the sentence primed the motor action, an effect they called the Action-sentence Compatibility Effect (ACE).

Recent research using a semantic judgement task supports the claim that sensorimotor simulations play a functional role in comprehending word meanings (Davis et al., 2020). Participants listened to words referring to either animals or to non-living things which varied in ratings of visual experience. In the control condition, participants just decided if each word was an animal. In the experimental condition, there was also a visual interference task in which participants saw an array of four nonsense shapes before and during the judgement task. To ensure they attended to these shapes, afterwards they were shown one shape and asked to decide whether it had featured in the array. Davis et al. found that higher visual experience ratings of the non-living things were associated with longer response times on the semantic judgement task. In a second experiment, participants made the same judgement to the same words, but this time they responded with their feet and participants in the experimental condition repeated a cycle of three hand movements during the judgement task. Response times were correlated with degree of manual experience ratings for the non-living things. Subsequent analyses showed that these effects were condition-dependent (i.e., the degree of visual experience ratings of words was not correlated with response times in the motor interference task). As both the visual shapes presented to participants and the hand movements they made were meaningless patterns, this interference was not semantic or conceptual, but due to sensorimotor processing demands, suggesting that the sensorimotor representation of words contributes to processing their meaning.

Research has also shown that abstract concepts may be partly structured by multimodal representations, similarly to concrete ones (for a recent review, see Kiefer & Harpaintner, 2020). For example, in a study using decision-making tasks it was shown that abstract words can be primed by pictures showing an event or situation that depicts the concept and that such pictures can also be primed by abstract words (McRae et al., 2018). Abstract conceptual knowledge may exhibit a complex structure incorporating multiple ways of experiencing the self and the socio-environment, including language experience (e.g., Damasio, 1989; Gabora et al., 2008).

Various rating studies support the idea of multimodal abstract concepts. In the first such study, Vigliocco et al. (2014) found that abstract words are rated as more highly valenced than concrete words and showed that emotion may play a role in grounding abstract concepts. Villani et al. (2019) used participant ratings of 425 abstract concepts (nouns) on 15 pre-determined dimensions to construct a multidimensional space representing each abstract concept. They were able to extract three components from these dimensions, which they labelled as concreteness /

abstractness, sensorimotor and internal /social. Within these components they further identified four clusters, which they characterised as different types of abstract concepts, which can be briefly summarised as physical, social, philosophical, and emotional. Likewise, Harpaintner et al. (2018) asked participants to generate features, situations, and associations for 296 abstract nouns, on which they then performed hierarchical cluster analysis. They found abstract concepts were highly associated with sensorimotor properties (including interoception) and emotions and internal states. In contrast to Villani et al., social features accounted for less than 10 per cent of all responses.

In sum, behavioural tasks have shown support for sensorimotor processing of concrete concepts and action verbs. Abstract concepts may also be grounded in emotion, interoception, and social events, in addition to language (Kiefer & Harpaintner, 2020).

2.5.4 Neural Evidence for the Embodiment of Language

There is also a substantial body of neuroimaging research that supports the claim that conceptual knowledge is represented via activation patterns across large areas of the brain, depending on the perceptual-motor content of the concept being retrieved (Bauer & Just, 2019). In a seminal fMRI study, Hauk et al. (2004) measured activation in motor cortical areas of the brain while participants read matched sentences with action or stative verbs, such as “Tom kicked the ball” or “Tom liked the ball”. Activation was greater when the sentences contained action verbs. In addition, areas representing arms and legs fired more when reading sentences with verbs such as “kick”, whereas areas representing the face fired for verbs such as “bite”. Similar findings were reported in a study that compared processing of action versus non-action verbs in fronto-parietal areas of the brain implicated in processing the observation of actions (Tettamanti et al., 2005). Likewise, González et al. (2006) found that when participants read words that denote a referent with a strong positive or negative odour, such as “rose” or “stink”, activation in olfactory cortical areas was greater than when they read words referring to a neutral odour, such as “guitar” or “kick”.

There is also evidence from populations with neural impairments that supports the involvement of sensorimotor representations in language processing.

For example, studies involving patients with Parkinson’s disease suggest this condition is associated with impaired language processing, particularly for (action) verbs, as reviewed by Cardona et al. (2013). They interpreted the results as evidence for coupling between motor and language systems, although they posit that such coupling involves the basal ganglia as well as cortical areas, suggesting that the premotor cortex may not play a causal role in deficits in processing action verbs.

Just et al. (2010) demonstrated that perceptual features of concepts can be decoded from resting state neural activity (i.e., with no behavioural task). They conducted factor analysis on fMRI imaging data of participants as they thought about 60 concrete nouns (five words from each of 12 taxonomic categories). They were able to elucidate four factors that were consistent across participants, which they interpreted as eating, manipulation, shelter, and word length. The first three of these factors were associated with cortical brain areas related to the activity, for example the eating factor was associated with areas active when chewing or biting. This indicates representations of words referring to concrete objects are multimodal.

The ability of words to elicit multimodal simulations has also been shown in a learning paradigm. James and Gauthier (2003) trained participants to associate novel 3-D objects with three sound- or movement-related actions (such as squeals or hops). They then asked participants to complete a matching task, which did not require the verbal labels, during fMRI imaging. The results showed that the superior temporal gyrus, which is activated by auditory stimuli, showed significantly increased activation when the matching task involved objects that had been associated with sound words. Objects associated with movement caused higher activation of the posterior superior temporal sulcus, which is active when processing motion. Both these effects were present in the right hemisphere only, further suggesting the words activated conceptual representations, rather than purely linguistic ones.

Importantly, these theories do not claim that multimodal representations of word meanings reach conscious awareness (e.g., Barsalou, 1999). There is some evidence that multimodal representations can facilitate processing of word meanings without behavioural effects. In a word priming study, Kellenbach et al. (2000) found that words for objects that had a visual-perceptual relationship but no conceptual relationship, such as “button-coin” did not elicit behavioural priming effects on a lexical decision task compared with word pairs with no relationship. However, such word pairs did reduce the N400 component, suggesting a processing advantage for primed words. They interpreted this as evidence that concrete words activate verbal and non-verbal processing systems and that overlapping activation in either system between a prime and its target facilitates processing.

One key debate within embodied cognition theories is whether simulations are integral to cognitive processing or are activated in a task-dependent manner (e.g., Gallese & Cuccio, 2018; Meteyard et al., 2012). For example, Simmons et al. (2005) found evidence that images elicit multimodal representations, not just visual imagery, suggesting automatic, task-independent imagery. They investigated the representation of foods by showing participants pictures of processed foods, places, or scrambled images while undergoing fMRI. Participants were asked to complete a simple picture-matching activity, by indicating if the current picture was the same or different from the previous one. They found that, compared to the other two types of images, food images activated areas of the brain that process the taste of foods and their reward value. Moreover, this activation was related to whether participants were in a state of hunger or satiety. This shows that visual stimuli can activate representations across other relevant modalities to construct the underlying concept.

However, other studies suggest that simulation only occurs when it is relevant to the task or goal. For example, Kuhnke et al. (2020) found evidence for unimodal and multimodal processing of word-concepts, dependent on the properties involved in a judgement task. They collected fMRI data as participants read words for concrete objects that varied in their association with sound and action. Participants were presented with 192 words that represented concepts with a high or low association with sound and action, such that words were either highly associated with both, one, or neither modality (e.g., guitar is highly associated with both sound and action). Participants performed a lexical decision task (involving 192 pseudowords) and judgement tasks on each word-concept's association with sound or action. They found that during the sound judgement task increased activation was recorded in areas related with processing auditory stimuli for words highly associated with sound. In contrast, during the action judgement task motor areas were selectively activated by action words. This indicates that perceptual-motor areas are activated in word processing in a task-relevant manner.

Importantly, in a connectivity analysis of the fMRI data collected in Kuhnke et al. (2020), Kuhnke et al. (2021) found evidence that simulation is a constitutive element of conceptual knowledge, not epiphenomenal. They showed that the auditory and motor processing areas were selectively functionally coupled with a multimodal processing area identified in the left posterior parietal cortex in a task-dependent manner. They further found that the strength of this bi-directional functional coupling was significantly correlated with each participant's rating of each word-concept's association with sound or action. This suggests that activation circuits involving unimodal and multimodal processing areas are causally relevant for behaviour, including language processing.

In a meta-analysis of studies investigating semantic representations through fMRI and PET imaging during tasks involving verbal input, Binder et al. (2009) provided further evidence that semantic information is processed in multiple areas of the brain across both hemispheres. They found reliable activation in areas of the brain such as the angular gyrus and temporal lobe that have been suggested to associate inputs from different modalities, further positing that the angular gyrus is involved in concept retrieval and combination. At that time, they suggested these association areas could reflect heteromodal (multimodal) or supramodal (amodal) integration. However, they also note that at the time of conducting the meta-analysis there were insufficient studies to assess the possible involvement of sensorimotor cortical areas in semantic processing of words. Nonetheless, the widespread activation of the N400 component across temporal and frontal areas of the brain was sufficient for Kutas and Federmeier (2011) to state that "various neuroimaging data all point to a multimodal semantic system" (p. 629).

More recent studies support the hypothesis that integration or association areas of the brain process multimodal representations. Fernandino et al. (2016) found evidence that neural activity in integration areas could be predicted by sensorimotor ratings of words. They used fMRI to record activation across the brain while participants read 900 words and judged if each one could be experienced through the senses (600 concrete and 300 abstract words). These stimulus words had previously been rated as to the importance of colour, shape, visual motion, manipulation, and sound to their meaning. Neural areas that have been shown to be involved in perceptual and motor processing were significantly activated in processing words rated highly on each dimension. For example, words rated highly for the attribute of shape or colour were associated with activation in occipital cortical areas responsible for visual processing. Shape was also associated with areas involved in haptic processing. The activated areas were mostly secondary rather than primary areas, which are involved in higher level integration of features. In a further study, this group found the activation patterns in association areas of the brain involved in semantic representation match experiential models (Fernandino et al., 2022; see Section 2.4.6). These results support the claim that multimodal representations play a constitutive role in conceptual knowledge, rather than merely contributing to higher-level, amodal representations.

In sum, neuroimaging studies including various approaches and tasks have been able to localise processing of conceptual knowledge in relevant sensorimotor areas of the brain and suggest that association areas process multimodal rather than amodal representations. Some studies also indicate that simulation is task-dependent but plays a causal role in making meaning. Such evidence has contributed to the growing acceptance within cognitive science that cognition is embodied (Spivey, 2023).

2.5.5 Evidence Against Strong Embodied Theories

As reviewed above, there is an impressive and growing body of evidence in support of embodied cognition theories. As a result, some authors have argued that purely symbolic accounts of cognition can be rejected; cognition must incorporate sensorimotor representations to some extent, even if such representations are neither necessary nor sufficient (e.g., Meteyard et al., 2012). However, this is not a universally accepted claim. For example, Pylyshyn (2003) argued that imagery plays no role in cognition and Mahon and Caramazza (2008) suggested that perceptual simulations are a by-product of amodal cognitive processing, rather than a causal factor in cognition. More recently, Goldinger et al. (2016) argued that embodied cognition theories are vague and unable to explain multiple lab-based research findings within cognitive science. However, their definition of embodiment does not map well onto that of previous research; they equate it with action, and do not consider the key claim that cognition is entwined with action and perception (e.g., Hurley, 2005; Pessoa, 2022; McNorgan et al., 2011). For example, the importance of familiarity in face perception (e.g., Jenkins et al., 2011), which Goldinger et al. suggested refutes embodied accounts of cognition, indicates the importance of situated experience. Jenkins et al. found that people do not recognise faces accurately across time and position unless they have interacted with the individual, directly or through media, a finding which counters purely symbolic processing. Notwithstanding, there are empirical studies which offer evidence that challenges strong accounts of embodied cognition and language (e.g., Chemero, 2011; Gallese & Lakoff, 2005; Thompson, 2007). Three particularly problematic lines of evidence are summarised below.

The first, most direct empirical evidence against embodied cognition theories comes from failures to replicate previous experiments that supported such theories. An important study in this regard is the pre-registered multi-lab replication of the Action-sentence Compatibility Effect (ACE) first reported by Glenberg and Kaschak (2002; see Section 2.5.3), which was conducted by Morey, Kaschak, et al. (2022). None of the labs involved in this effort found a priming effect of the movement direction of sentences on response direction. Differences in response times as an effect of sentence and response movement direction (away or towards the participant) were negligible. These results suggest further replication research is needed to decide whether simulation plays a role in normal language comprehension at the sentence level. Similarly, Montero-Melis et al. (2022) failed to replicate an effect of making rhythmic hand and foot movements on the maintenance of action verbs in working memory, as reported by Shebani and Pulvermüller (2013).

In addition, the results from some lesion studies support weak rather than strong embodied cognition theories, as they question whether simulation is necessary for language comprehension and production. For example, despite the evidence that patients with Parkinson’s disease have selective deficits in processing verbs (for a review, see Cardona et al., 2013), some studies have failed to find this effect (e.g., Aiello et al., 2022). Likewise, Argiris et al. (2020) found a dissociation between imaging and language processing. They conducted motor imagery and motor verb processing tasks with 20 participants with focal gliomas in somatotopic regions of the sensorimotor cortex in the left or right hemisphere and contralateral leg or arm motor impairments (and control participants). They found significant impairments in patients’ motor imagery but not in verb processing, which they interpreted as evidence against an essential role of simulation in language processing. However, patients showed enhanced activation in the intact hemisphere, which could indicate compensatory neural processes (Buzsáki, 2019).

The final challenge to strong embodied cognition theories comes from research with individuals who do not have access to all sensory modalities. Research comparing language processing in congenitally blind participants with those with normal vision has shown mixed results for embodied cognition theories. In a lexical decision task, Bottini et al. (2022) found that, for both groups, response times were shorter for words representing concrete than abstract concepts. Importantly, concrete words (nouns and adjectives) were divided into two groups according to ratings of perceptual experience: one group referred to concepts experienced primarily through visual processing, such as “blue” and “sparkling”, and the other group referred to multimodal concepts experienced through other senses, such as “delicious” and “chill”. There were no significant different in response times to visual and multimodal word-concepts within either group. Their findings suggest that the processing advantage for concrete nouns over abstract ones is not a result of simulation. However, overall response times were significantly shorter for normally sighted participants, such that their average response time for abstract words was highly similar to that of the congenitally blind participants for concrete words. Moreover, lexical decision is a shallow processing task, which may not require access to meaning and therefore not involve simulation (e.g., Barsalou et al., 2008; Ostarek & Huettig, 2017). Nevertheless, other authors have highlighted the commonality in conceptual processing between sighted and blind individuals, suggesting conceptual knowledge is abstracted away from perceptual referents (Bedny & Saxe, 2012).

The evidence reviewed above questions the tenability of strong embodied cognition theories which suggest simulation is necessary and/or sufficient for processing semantic meaning (conceptual and/or linguistic). Of note in this regard, authors such as Gallese, who was associated with strong views (Gallese & Lakoff, 2005) has more recently explicitly adopted a weaker stance (Gallese & Cuccio, 2018). This is partly due to the evidence that simulation is task-dependent (e.g., Ostarek & Huettig, 2017). In particular, many theories suggest that cognitive processes involve simulation and (symbolic) language processing (e.g., Barsalou, 1999; Barsalou et al., 2008; Binder & Desai, 2011; Borghi et al., 2019; Connell & Lynott, 2013; Vigliocco et al., 2004), akin to dual-coding theory (DCT; Paivio, 1986). Such theories are discussed next.

2.6 Language and Simulation Theories

As indicated above, weak theories of embodied cognition posit that conceptual knowledge is grounded in multimodal simulation and also structured by more abstract, linguistic knowledge (e.g., Andrews et al., 2009; Meteyard et al., 2012; Vigliocco et al., 2009; for a review see Davis & Yee, 2021). Such theories suggest that the contribution of each source of knowledge depends on the task and possibly other factors. These theories have their origins in DCT (Paivio, 1986; see Section 2.4.5), although rather than contrasting concrete and abstract knowledge they reframe the distinction as one between deep (multimodal) and shallow (linguistic) processing (e.g., Barsalou et al., 2008; Connell, 2019). Task-dependent, dual processing theories are becoming increasingly accepted (Gainotti, 2017) due to substantial empirical support. For this reason, the WA coding system developed in this thesis is informed by a dual processing theory, LASS theory (Barsalou et al., 2008). Before describing this theory, evidence in favour of dual processing is reviewed below.

2.6.1 Evidence for Language and Simulation Theories

The main reason to prefer a weak embodied cognition theory is the growing body of evidence that simulation in language processing is task dependent (Papeo & Hochmann, 2012). For example, in judgement tasks, Louwerse and Jeuniaux (2010) found evidence that simulation was activated by word stimuli when the task required an iconicity judgement but not when the task required a semantic similarity judgement based on categorisation. Tomasino et al. (2008) found that primary motor cortex is involved in tasks requiring the explicit mental imagery of action verbs, but not linguistic decisions related to them. Likewise, Ostarek and Huettig (2017) found that visual noise slowed reaction times to concrete but not abstract nouns in a concreteness judgement task that required access to word meaning, but not in a lexical decision task. Finally, as reviewed in Section 2.5.4, Kuhnke et al. (2020) found task-dependent activation in multimodal areas in a semantic judgement task involving words that were differentially associated with sound and action. However, they also found increased activation in the anterior temporal lobes for words compared to pseudowords, which was not affected by the judgement task, supporting the involvement of this area in abstract processing of linguistic-conceptual information. They suggested that word-concept representations combine unimodal perceptual-motor, multimodal, and amodal information.

Wurm and Caramazza (2019) found evidence that action representations in the brain based on language and imagery are similar, but not identical. They conducted an fMRI study in which participants watched short videos and read sentences about different actions. The actions varied according to whether they were directed towards another person, an object, a person and an object (e.g., give), or neither (e.g., scratch). They found significant activation in left lateral posterior temporal cortex (LPTC) and frontoparietal cortex that could discriminate between different types of actions presented through videos and sentences. However, they also found that only the LPTC encoded cross-modal representations. In this region, activation patterns elicited by watching videos could be used to predict activation patterns elicited by reading sentences, and vice versa. In contrast, activation in the frontoparietal cortex was modality specific. This suggests that imagery and text activate overlapping but distinct conceptual representations. Of note, decoding accuracy was greater for video than language stimuli, which they attributed to richer, more detailed representations. This could be interpreted as support for the claim of LASS theory (Barsalou et al., 2008) that language representations are shallower compared with multimodal simulations, a claim investigated in this thesis.

Research comparing neural processing of words by sighted and early blind participants further suggests that conceptual representations can be structured by simulation or language. In an fMRI study by Bottini et al. (2020), sighted and early blind participants judged whether action or colour word pairs were similar or different. They found greater differences in activation patterns for colour word judgements than for action word judgements, which can be grounded in multimodal conceptual representations. In sighted participants, colour word judgements elicited activation in occipital cortical areas that process colour information, whereas in early blind participants an increase in activation was observed in areas of the temporal lobes. This suggests the colour concepts of early blind individuals do not include visual features that are accessed by sighted individuals but are based on linguistic information. A similar finding was observed in a comparison of action verb processing (silent reading) in patients with Parkinson’s disease and healthy controls using fMRI (Abrevaya et al., 2017). Although the premotor cortex showed increased activity in both groups, the functional connectivity patterns were different, suggesting simulation in healthy controls, but lexical processing in patients. Importantly, the strength of this alternative activation pattern in participants with Parkinson's disease was associated with damage to the basal ganglia. These findings are in line with the known degeneracy and plasticity of the brain (Buzsáki, 2019): typical processing may involve multimodal representations, but alternative strategies are available.

2.6.2 The Interaction of Linguistic and Conceptual Knowledge

Some dual theories posit that the two processing systems are interconnected, such that semantic knowledge is co-constructed by multimodal concepts and language (e.g., Borghi, 2019; Evans, 2009). As noted earlier, research and theories regarding semantic knowledge tend to conflate conceptual and linguistic (word) knowledge. However, studies of aphasias demonstrate that the two types of knowledge can be distinguished (e.g., Dronkers, 2004). In embodied cognition theories multimodal simulations (conceptual representations) ground word meanings. In weaker views, it is posited that the interaction of conceptual and language knowledge is critical to human cognitive abilities. This suggests that language can also play a role in structuring concepts and highlights the continuum of abstraction from simulation.

The idea that linguistic and conceptual representations exert a reciprocal influence can be traced back at least as far as Humboldt (1836/1999). He claimed that the synthesis of the sound-forms of language and concepts "creates something that does not lie, *per se*, in any of the conjoined parts" (p. 88). More recently, many authors have proposed theoretical accounts that human semantic knowledge is a combination of conceptual and language representations. A few such theories are summarised here to illustrate the popularity of this idea.

In the featural and unitary semantic space (FUSS) hypothesis (Vigliocco et al., 2004), conceptual structure is composed of multimodal featural representations, which are bound into lexico-semantic representations, which then interface with other linguistic features (such as phonology and grammatical patterns). Clark (2006) suggested language is a "material symbol", a cognitive tool that enables us to think at a higher level of abstraction about real world problems, thus mental representations are a hybrid of multimodal and linguistic (symbolic) representations. He speculated that multimodal thought processes and linguistic thought processes might be a loosely coupled system, with the contribution of each varying according to task demands, but with their differences meaning they complement, rather than impede, each other. Evans (2009; 2016) posited that a word-concept forms a reciprocally self-forming unit of meaning, such that conceptual and linguistic representations contribute complementary semantic information. Dove (2011, 2020) suggested that language is a symbolic system grounded in simulation, which enhances and scaffolds conceptual representations, particularly abstract concepts. Finally, Lupyan and Lewis (2019) argued that words help construct concepts or meanings, rather than mapping onto them. They suggested that words assist the acquisition of semantic knowledge from perceptual-motor experiences by promoting the abstraction of defining features of the concept, forming categorical representations. Rosch's (1975) priming studies could be interpreted as support for this claim. She found same category judgements for picture and word pairs were facilitated when the superordinate label was spoken two seconds prior to the stimuli and simultaneously.

In all these views of language, words do not get their meaning from the concept they refer to, but co-create meaning simultaneously with the sensory-motor representation of the stimulus. The regularities of individual instances, as delineated by a word, form a concept. Thus, language is not just a product of cognitive processes but is also a constitutive element of human cognition (see also Deacon, 1997). As McGregor et al. (2015) put it, "language ... can no more be the substrate of thought than it can sit on top of thought" (p. 60). Various studies support the idea that language influences, or is part of, conceptual knowledge, as reviewed next.

Research with infants suggests that word knowledge may be incorporated within concepts knowledge from an early age, forming word-concepts. Mani and Plunkett (2010) presented 18-month-old infants with an image prime followed by two simultaneous images of familiar objects, ensuring that all images were unrelated according to word association norms. When the word for the prime and one of the subsequent images started with the same sound (e.g., a cat followed by a cup), the infants fixated significantly longer on the image whose name started with the same initial sound than the distractor image. This suggests that infants may subconsciously label their world as soon as they start acquiring language. Language has been shown to assist in the understanding of large numbers (e.g., Carey, 2009) and may even structure numerical knowledge (Pitt et al., 2022).

This notion of a word-concept has been further supported by research with adults. Meyer et al. (2007) conducted a visual search task in which participants were shown four images simultaneously and had to respond whether or not a target object was present. In some trials none of the distractors were related to the target in any way, but in some trials either the name of one of the distractors was a homophone of the target object (such as boy and buoy) or there was a semantically related distractor. They found that the presence of a homophonic or a semantically related distractor in the image set caused a similar and significant delay in response time, both when the target image was or was not present. They also found that in trials when the target was absent but a distractor was present, the distractor was fixated on for significantly longer than any of the control objects, for both types of distractor. This suggests adults implicitly generate the labels for visual input and that such subconscious naming influences behaviour, even in tasks in which participants do not explicitly verbalise the names for the objects.

Despite the evidence that language can structure conceptual representations, in weak embodied theories it is maintained that language must be grounded to be meaningful. In particular, language must be grounded evolutionarily (e.g., Rizzolatti & Arbib, 1998) and in the initial acquisition of words. For example, Landauer and Dumais (1997) interpret the success of their Latent Semantic Analysis (LSA) model (described in Section 2.3.3) as conceptual evidence that most word meanings could be learned from linguistic input alone, without direct experience of their referents. However, extracting semantic relations from texts does not imply any interpretation of word meaning (e.g., Capurro & Hjørland, 2003; Lake & Murphy, 2023; Searle, 1980). Even if many word meanings can be learned from other words, initial words need to be grounded by their referents (Vincent-Lamarre et al., 2016). Louwerse (2018) also suggested that to the extent that words regularly co-occur with their referent they may develop an indexical relationship, rather than purely symbolic. Words point to their referent even in its absence. He claimed that in this way, symbolic language encodes multimodal conceptual knowledge, such that once some words are grounded by their referent, words with similar referents can be understood purely through linguistic association, a form of linguistic bootstrapping. Of note, he also stated that such language-derived word meanings can only provide a shallow understanding.

In this line, in a comparison of corpus- and image- derived models of semantic networks, Anderson et al. (2013) found that both types of information structure human cognition. They created models of the semantic space for concrete nouns based on the distribution of words in corpora and features extracted from images as well as a model which combined both sources. They then compared the models with fMRI data collected from participants as they thought about the properties of 51 concrete objects in 11 categories (such as animals, clothes, vehicles) that were presented as line drawings with a word label. They found significant correlations with the neural data that were strongest for a model that combined word and image-based distributions. These model results suggest that conceptual representations are constructed from both words and perceptual properties.

Most natural concepts and their labels are learned through multiple sources of sensory information and (inter)action, typically within social event contexts and accompanied by affective experience. This is especially true of the earliest learned words, which are supported by repeated exposure to their visual referents (Clerkin & Smith, 2022). As Sheya and Smith (2019) argue, from a developmental perspective, symbolic representations or cognitive processes are a product of perception-action loops that occur within a changing environment and actively change or structure the environment. This is analogous to the claim by Buzsáki (2019) that even if under some circumstances, perception may not require action, without initial action calibration there can be no perception. There is also evidence that visual information may support adult word learning to a greater extent that naturalistic language exposure. In a series of experiments aimed at understanding how adults learn new words, Leach and Samuel (2007) trained participants to learn nonsense words composed of three to four English syllables (such as "galasod" and "penivasher"). They found that participants showed greater learning of the new words when they were paired with images of unusual objects than when they were embedded multiple times in short narratives. These findings suggest that even if lexical semantics may not always require multimodal representations, this grounding is essential to the development of language. Simulations are a crucial foundation for more abstract knowledge representations such as words.

2.7 An Embodied Language Framework for a New WA Coding System

The weak embodied cognition theory that informed the research conducted in this thesis, Language and Situated Simulation (LASS) theory, is a dual processing theory proposed by Barsalou et al. (2008). As described in Section 2.4.5, the first detailed theory of this kind was DCT (Paivio, 1986; see also Sadoski & Paivio, 2013). LASS theory builds directly on DCT and also Glaser's (1992) lexical hypothesis, in positing that language knowledge is shallower than conceptual knowledge (see also Louwerse & Connell, 2011).

In LASS theory, concepts are multimodal representations grounded in sensorimotor simulations of situated experiences of the concept (e.g., Barsalou, 1999; 2008). Thus, contextual information contributes to the simulation; knowledge is not represented internally in a vacuum. Conceptual representations are distinct from representations of language forms, but the two systems are co-dependent and co-construct meaning. In contrast to other theories that distinguish between concrete and abstract concepts (e.g., Paivio, 1986), in LASS theory it is assumed that all concepts are constructed through situated simulation. Of note, both systems comprise analogue representations rather than amodal codes. This idea is similar to Paivio's (1986) distinction between imagens and logogens as units of representation in imagery and verbal systems, or to early descriptions of words in the mind as "word-images" (Kussmaul, 1877).

An important hypothesis of LASS theory for this thesis is that word stimuli initially focus attentional processes on representations within the language system. Only as the ability of this system to complete the task reduces does attention start shifting towards deeper processing of situated simulations in the conceptual system. This claim is supported by word and picture naming experiments, which show that word naming is reliably faster than picture naming, despite pictures being recognised faster than words (e.g., Glaser, 1992). Moreover, word stimuli are also predicted to activate associated words in the lexicon. Thus, associations between words are central to individual word processing. These two ideas are evaluated in more detail below.

Barsalou et al. (2008) suggested that word stimuli simultaneously activate linguistic forms and situated simulations, but that activity in the linguistic system initially dominates conscious (executive) processes. They also suggested that linguistic form is sufficient for superficial tasks, such as producing word associations, but deeper processing necessitates situated simulations. They posited that situated simulations may be rapidly activated after word presentation. However, such information remains subconscious in language processing tasks that do not require deep conceptual processing, such as word association tasks. This means that attention is directed almost entirely towards the linguistic processing system (p. 250). It is important to note that Barsalou et al. (2008) view the two systems as constantly interacting in making meaning. However, they argued that the relative contribution of each system is task dependent. In tasks such as WATs the linguistic system is more active, especially in producing early responses, such that the contribution of the simulation system is not attended and its contribution is therefore reduced. As is apparent from the above description, LASS is a theory of processing, not semantic memory per se. Moreover, its predictions pertain mainly to the type of representations that are consciously accessed in response to stimuli, not the total activation of semantic knowledge.

From the presentation of LASS theory, it would appear the authors view the linguistic system as a purely phonemic-orthographic system rather than one that contributes towards meaning-making. They suggest that word recognition depends on linguistic forms activating situated simulations which provide conceptual meaning, describing linguistic representations as “superficial” (Barsalou et al., 2008, p. 251). Yet they also suggest that word forms control simulations, enabling humans to manipulate the content of their thoughts. It is unclear how word forms that do not contain any semantic information could manipulate conceptual representations due to interaction based solely on association. Barsalou et al. (2008) even state that “in other contexts, the linguistic system would include the representation of meaning, thereby including the simulation system” (p. 253). However, they provide no indication of what such other contexts might be or how the interaction between the two systems may differ in such contexts.

In their presentation of LASS theory, supporting evidence was provided by Barsalou et al. (2008) from a behavioural (Santos et al., 2011) and an fMRI (Simmons et al., 2008) study conducted specifically to test the theory (in addition to other sources). In both studies, participants performed a WA and a property generation task for the same set of cue words. The behavioural WAT is described in detail in the next chapter, in which a replication of that task is reported. Two relevant points are highlighted here. First, the choice of these two tasks reflects the explicit assumption that “word association focuses attention on the linguistic system” (Barsalou et al., 2008, p. 256), whereas property generation is “more conceptual in nature” (p. 258). As noted above (Section 2.3.2), comparisons of WA and corpus data question the former assumption. Second, the authors assumed *a priori* that the production of collocations, synonyms, and antonyms reflects mainly linguistic processing, whereas conceptual processing (simulation) results in the production of properties and situational descriptors.

Across the three papers (Barsalou et al., 2008; Santos et al., 2011; Simmons et al., 2008), no explanation is given regarding why linguistic knowledge should include collocations, synonyms, and antonyms as well as form-based relations, but these three types of relation would seem to require access to meaning. Based on the description of taxonomic associates, it would seem that these types of associates are assumed to be linguistic as they are learned as “memorized lexical phrases” (Barsalou et al., 2008, p. 258) and cannot be simulated in situational contexts (their definition of conceptual knowledge). However, testing LASS theory before testing this assumption seems problematic. This issue is revisited in Chapters 5 to 7.

Despite this caveat, both the behavioural and neural results supported LASS theory. In both tasks (Santos et al., 2011), participants tended to produce linguistic responses earlier than conceptual ones. The property generation task showed that participants produced more conceptual than linguistic properties, although this might be anticipated as they were asked to produce typical characteristics of each cue. More importantly, they produced more linguistic properties in the first 7.5 seconds than in the second half of the response period. The results of the behavioural WAT reported in Santos et al. are discussed in more detail in the next chapter, in which this task is replicated. In the fMRI study (Simmons et al., 2008), areas of the brain implicated in multimodal processing were more active during property generation than WA. Within the property generation task, these areas were also more active during the second compared with the first half of the response period. This finding supports the hypothesis that conceptual responses are generated later than linguistic ones, suggesting that conceptual processing is deeper than linguistic processing because it relies on multimodal representations.

2.8 Concluding Remarks

Given the substantial evidence for weak embodied cognition theories and the increasing acceptance that such theories may currently constitute the most informative and useful framework for understanding conceptual and linguistic knowledge (e.g., Gainotti, 2017; Meteyard et al., 2012), this theory is appropriate for developing a new coding system for WA data that is informed by cognitive science. Within this field, LASS theory (Barsalou et al., 2008) was selected because it brings together ideas from weak embodied cognition and dual processing theories, making it an important theory to test. In addition, its emphasis on the role of language in processing word stimuli makes it suitable for research within the field of applied linguistics. Moreover, due to this emphasis, the word association paradigm was used by Santos et al. (2011) to test LASS theory. Therefore, it presents an immediate departure point for the research to be conducted in this thesis, using WA to examine the mental lexicon.

As well as the theoretical grounds for choosing LASS theory as a framework in which to attempt to develop a new coding system for WA data, there are also methodological motivations. The chief motivation is that in the studies conducted by Santos et al. (2011) and Simmons et al. (2008), they used a coding system that was composed of categories and subcategories. Importantly, the three general categories, linguistic, taxonomic, and object or situation descriptor (conceptual) are very different from the paradigmatic-syntagmatic or meaning-, position-, and form-based categories that have most frequently been used in L2 WA research (see Section 2.2). In addition, the existence of 10 subcategories within these general ones suggest that it might be suitable to compare this coding system with the subcategories used by Fitzpatrick (2006, 2007, 2009). The potential for comparing the new coding system with established ones provides an essential benchmark for evaluating the informativeness of the new system and establishing whether it may offer any further understanding of WA and the insight this task can yield into the (developing) mental lexicon. This is important because WA accesses "the *right kind* of meaning" (De Deyne et al., 2018, p. 1000), yet, as noted by Nelson et al. (2000) "we know little about the representations and processes involved" (p. 888).

In addition, given the importance of LASS theory within embodied cognition theories, further empirical evidence regarding this theory is vital. As Borghi et al.

(2019) stated, “it was the first influential theory considering both sensorimotor and linguistic experience” (p. 137). Despite this, to this author’s knowledge, LASS theory has been empirically tested in only two studies (Santos et al., 2011; Simmons et al., 2008), both of were produced by the same research group. Although these studies supported the theory, independent assessment is needed. This makes the WAT reported in Santos et al. an important experiment for replication.

Thus, the main aim of the first experiment reported in this thesis is to replicate the WAT conducted by Santos et al. (2011), further testing their two main hypotheses. The first hypothesis is that participants will produce mostly linguistic associates (collocation, synonym, and antonym responses), as WA is a language task which primarily accesses linguistic knowledge. Of note, Santos et al.’s results did not support this hypothesis, as participants produced a similar proportion of linguistic and conceptual associates. The second hypothesis is that linguistic associates will be produced earlier than conceptual ones, as linguistic processing is shallower and so is activated more quickly than multimodal conceptual processing.

If the replication produces similar results to those reported by Santos et al., confirmation of the second hypothesis would support the validity of the distinction between linguistic and conceptual associates, as it would indicate a processing advantage for linguistic responses in terms of speed. However, if the first hypothesis of Santos et al. is further disconfirmed in the replication, this would suggest that WA is not purely a language task (as suggested by comparison with corpus data in Section 2.3.2). Such findings would indicate that a new coding system, based on linguistic, taxonomic, and conceptual associations may be psychologically valid. Moreover, this coding system may yield novel insights into the nature of WA compared with the paradigmatic-syntagmatic coding system proposed by Osgood et al. (1954), given that this latter system was intended to provide a purely linguistic understanding of WA.

Chapter 3. Replication and Reanalysis of a Word Association Experiment

This chapter reports the replication of a WA experiment conducted by Santos et al. (2011), which was briefly described in Chapter 2 (Section 2.7). In doing so, the WA reported by Santos et al. is referred to as the original experiment in contrast to the replication experiment reported in this chapter. The main novelty of the original experiment was the coding system introduced by Santos et al., which divides responses into linguistic, taxonomic, and conceptual associates (described in Section 3.1.2). This coding system is derived from language and situated simulation (LASS) theory (Barsalou et al., 2008; see Chapter 2, Section 2.7) and the main aim of the original experiment was to test two key predictions of this theory. The main aim in this chapter is to replicate Santos et al.'s findings and thus, to test their hypotheses. These hypotheses were (a) participants will produce a greater proportion of responses that have a linguistic association to the cue words than responses that have a conceptual association; and (b) linguistic associates will be produced earlier than conceptual ones.

This chapter starts with a description of the original research conducted by Santos et al. (2011), comprising the theoretical assumptions and aims of their study, the methods they employed, the new coding system they introduced, and their main findings (Section 3.1). Section 3.2 describes the methods followed in the replication experiment, showing how the replication adhered to the information provided in Santos et al. as closely as possible. Section 3.3 presents the results of the replication experiment, produced by following the two analyses reported in the original paper as far as possible, based on the information provided. Due to substantive issues that emerged in conducting the replication experiment and analysing the results, further analyses are reported in Section 3.4. These additional analyses are similar to those conducted by Santos et al. and provide a more accurate representation of the data collected in the replication experiment. Finally, this chapter ends with a discussion of the results of the replication in relation to the predictions of the original experiment and to LASS theory (Barsalou et al., 2008).

3.1 The Original WA Experiment

The WA experiment replicated here was reported in Santos et al. (2011), together with a property generation task using the same stimulus words. A companion experiment was reported in Simmons et al. (2008), in which participants silently performed similar tasks while undergoing fMRI scanning. Results from both studies were discussed in Barsalou et al. (2008) to provide support for LASS theory. This research introduced a novel coding system for word associations, based on a distinction between linguistic, taxonomic, and object or situation descriptor (conceptual) associates.

The coding system introduced by Santos et al. (2011) has potential to yield new insight into the types of association people make in WATs for two reasons. First, it derives from a specific psychological theory of semantic knowledge, LASS theory (Barsalou et al., 2008). This contrasts with both the traditional paradigmatic-syntagmatic coding system, which was intended to be purely linguistic (Osgood et al., 1954), and the meaning-, position-, and form-based system (Fitzpatrick, 2006, 2007), which was influenced by Osgood et al. and Nation´s (2001) description of word knowledge. Second, Santos et al. distinguished 10 specific categories within the general response categories, which makes the system more informative (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Thwaites, 2020). These points suggest that this coding system could provide a fruitful new approach to analysing WAs to assess how such data might contribute towards understanding the mental lexicon.

However, to the best of my knowledge, Santos et al.'s coding system has not been employed in any subsequent word association study. Therefore, it is important to ensure the original findings reported by Santos et al. can be replicated, particularly given the authors who proposed LASS theory were involved in the empirical investigations of the theory. If the study can be replicated, then the new coding system can be further tested and developed and compared to existing coding systems for analysing word association data (see Chapter 2, Sections 2.2 and 2.3).

3.1.1 Theoretical Basis of the Original Experiment

In the first of two experiments reported in Santos et al. (2011), the authors used a multiple-response WAT to test two key predictions of LASS theory (described in Chapter 2, Section 2.7). According to LASS theory, it should be possible to produce word associations via shallow, linguistic processing. Moreover, word stimuli should activate linguistic (form-based and lexical) processing prior to situated simulation (conceptual) processing. These two predictions map onto two hypotheses regarding the WAT: (a) participants should generate more responses via linguistic processing than via conceptual processing; and (b) linguistic associates should be produced earlier and more quickly than conceptual ones. They further conjectured that this tendency could be heightened by requiring participants to provide associates as quickly as possible, without hesitation (pausing), as this would increase the tendency to rely on shallow (linguistic) processing. It should be highlighted that both these predictions refer to processing of word stimuli, not to the storage of words within the mental lexicon. In addition, Barsalou et al. (2008) were explicit that images or other non-verbal stimuli could activate conceptual processing as quickly, or even more quickly, than word or other language stimuli. Indeed, they hypothesized that in the property generation task (the second task reported in Santos et al.), participants should produce more responses through conceptual processing, especially as they were given 15 seconds to produce properties.

As the coding system introduced by Santos et al. (2011) was motivated by LASS theory (Barsalou et al., 2008), it is informed by the view of semantic knowledge provided by that theory, instantiating its assumptions. As explained in Chapter 2, Section 2.7, LASS theory is a dual coding theory, positing two sources of semantic knowledge: lexical knowledge acquired through experience of language and conceptual knowledge acquired through direct experience of the world (socio-environment). Linguistic responses are those that reflect knowledge of the distributional properties of words, or how they are related to other words. Thus, Santos et al. classified those responses related to cue words via collocation, synonymy, antonymy, and form (same root or sound) as linguistic associates. Conceptual associates are words that invoke multimodal representations of experienced situations (situated simulations). These responses are either properties (e.g., “green” is a property of the concept “grass”) or situated simulations (e.g., “cows” are situationally associated with the concept “grass”).

In addition, Santos et al. (2011) recognised taxonomic relations between words: superordinates, coordinates, and subordinates. Although they posited that these should be frequent within responses, they also stated that it was unclear whether this type of association reflects linguistic or conceptual knowledge. They related this uncertainty to how taxonomic relations or categorical knowledge is learned (see also Barsalou et al., 2008). They argued that on the one hand, taxonomic knowledge could be learned through abstracting shared features from experienced entities, making such knowledge conceptual. On the other hand, such knowledge tends to be taught and learned explicitly through language, making it linguistic. They also suggested that taxonomic relations may reflect both types of knowledge, as coordinate and subordinate relations are learned through experience, forming conceptual representations, whereas superordinate relations may reflect linguistic learning and knowledge. Therefore, they did not make any hypotheses regarding how early taxonomic associates would be produced, although if they reflected linguistic knowledge they should be produced early and if they reflected conceptual knowledge they should be produced later.

It should be noted that Santos et al. (2011) recognised a somewhat similar WA coding system was introduced by Chaffin (1997). Chaffin classified associates as defining and event-based relations, in which defining relations were typically taxonomic, synonyms, or meronyms (and thus also paradigmatic) and event-based relations were typically semantically related nouns, verbs, or adjectives that would be syntagmatic in speech. Using this classification system, he found unfamiliar words elicited mostly category responses (and a higher proportion of clang responses), whereas familiar words elicited event-based responses (all differences were statistically significant), which he suggested reflected sentential context effects.

The new coding system for WA data proposed by Santos et al. (2011) has three main, related differences from the traditional paradigmatic-syntagmatic coding system. First, as mentioned above, its theoretical basis comes from within cognitive science rather than applied linguistics. Second, within this new framework WA is not viewed as a purely language-based task (e.g., Clark, 1970), questioning a key assumption of such tasks that has been implicitly or explicitly made by many researchers, as noted by Nissen and Henriksen (2006). Third, the distinctions made between the main or general types of associates show little overlap. Linguistic associates in the new system comprise canonical syntagmatic (collocation) and paradigmatic (synonym and antonym) associates. The categorisation of properties and situated descriptors makes no reference to this distinction, so responses coded as conceptual will also include both syntagmatic and paradigmatic relations. Only the taxonomic response category coincides with paradigmatic responses. Together, these points demonstrate profound differences between the linguistic-taxonomic-conceptual and the paradigmatic-syntagmatic coding systems. It should be noted that the new system is similarly distinct from Fitzpatrick's (2006, 2007, 2009) meaning-, position-, and form-based coding system, given that meaning-based responses include linguistic, taxonomic, and conceptual associates. The coding system introduced by Santos et al. is described next, as well as how it was used to select the cue words for the WAT (and the property generation task).

3.1.2 The Coding System for Cue and Response Words

Santos et al. (2011) selected 64 cue words for the word association (WA) task, divided equally into eight categories, with the aim of selecting cues that would elicit a variety of linguistic, taxonomic, and conceptual responses. Categorisation (predicted possibility of eliciting the desired response type) was based on dominant responses given in the University of South Florida (USF) norms (Nelson et al., 1998; 2004; see Chapter 2, Section 2.1.1). Thus, coding of response type matched the coding of the cue word type. Four different types of linguistic association were identified (word relations assumed to be learned through exposure to language), so eight cue words were selected for each of these four categories of cue words. These were collocation, both forward (the response follows the cue word to make a compound phrase, such as "golf: ball"), and backward (the response precedes the cue word, such as "station: train"), synonym (e.g., "excellent: great"), and antonym (e.g., "heavy: light"). The other four cue word categories were taxonomic (e.g., "lobster: crab"), semantic field (e.g., "winter: summer"), brand names (e.g., "Tylenol: aspirin"), and cues that were stereotypical object properties of the dominant associate (e.g., "wings: bird").

As noted above, in Santos et al. (2011) response words were coded based on the categories determined for the cue words. First, responses were coded into the three main response type categories of linguistic, taxonomic, and situated simulation (conceptual) associates. These three main or general response type categories were further divided into 10 subcategories (referred to as concept types in Santos et al.). However, these subcategories did not match exactly to the eight cue word categories. The general linguistic response category was divided into six specific categories: the four cue word categories (forward and backward collocations, synonyms, and antonyms), plus responses with a similar root or sound to the cue word. The general taxonomic category was broken down into three specific response categories, separating superordinate, coordinate, and subordinate responses. In contrast, the general conceptual category was described as "object or situation descriptor" and was not divided any further.

The 10 subcategories of response types differed from the eight cue category types in the following ways. First, linguistic associates included two types of response that were related to the cue by sharing a similar form, either through the addition or subtraction of an affix or with no meaningful relation. Second, the cue word categories of semantic field and brand were not used as response type categories. Judging by the examples given for both types of cues and the results reported, associations that were words from the same semantic field as the cue or were brand or product names in response to a brand cue were coded as taxonomic associations. Finally, responses of the predicted type to object property cue words were coded as object or situation descriptor responses. In addition to these 10 categories, there was a "none" category for any associates that could not be coded within these 10 types, which corresponds to the clang, erratic, or other category that is included in most, if not all, WA studies. The cue word categories and the subcategories used for coding responses are evaluated in Section 3.1.5.

In coding responses, Santos et al. (2011) based their decisions on the predictions of LASS theory (Barsalou et al., 2008). They first determined if there was a possible linguistic connection between the cue and response words. If so, the response was coded as linguistic, even if it could have been produced by simulation (i.e., conceptual processing). An example they provided was that the responses "hive" and "honey" to the cue "bee" were coded as collocations (forward and backward). If the response did not enter into a linguistic association with the cue, they then considered whether it was a taxonomic association. A response was only coded as an object or situational descriptor (conceptual associate) if it didn't fit into any of the nine specific categories of linguistic and taxonomic associates. Santos et al. defined this as a probabilistic coding scheme. They argued that responses coded as linguistic were "*statistically more likely*" (p. 91) to have been produced via language processing, as were conceptual associates to have been produced by situated simulation. This assumption is discussed in Chapter 8.

3.1.3 Methods of the Original Experiment

As described above, the original WA experiment consisted of 64 cue words, categorised into eight concept types. The cues were divided into four lists, which were also presented in reverse order to create eight different lists. According to Santos et al. (2011), the cue words were quasi-randomly ordered, such that semantic relatedness between the words on each list was minimised and no two words from the same category type were adjacent on a list.

Santos et al. (2011) recruited 160 participants, who each produced associations to 18 cue words, two of each of the eight concept types plus two practice words, which were not analysed. Thus, 40 people produced associates to each cue word. The whole procedure (obtaining informed consent, instructions, and WAT) took about 10 minutes to complete, and participants were paid $2 for their time.

Santos et al. (2011) described the data collection procedure in detail. First, the experimenters approached potential participants individually in quiet areas on campus (Emory) and asked if they would be willing to take part in a short language experiment. If they agreed, they were asked to give informed written consent before proceeding. Each participant was instructed to say as many responses as they could to each cue word, stressing that there were no correct answers, they should just say the first words that came to mind, spontaneously and naturally. Participants were told that after they had said the first words that came to mind immediately, they did not need to say any more words. Before each cue word, the experimenter asked the same question, “For the following word, what other words come to mind immediately?” As soon as a participant paused, they moved on to the next cue, in order to minimise conceptual responses. The entire task was recorded on a hand-held digital audio recorder, and all response words were transcribed afterwards, so that they could then be coded.

3.1.4 Main Findings and Interpretation of the Original Experiment

Santos et al. (2011) reported the proportion of responses of each specific and general type as a function of cue word concept (category) type. The greatest proportion of responses of any specific type tended to be given to the cue words predicted to elicit that type of response. For example, 44% of responses given to forward collocation cues were forward collocations, whereas the cue word type eliciting the second highest proportion of forward collocation responses was brand cues, which elicited only 14% of such responses. The only exception was found with taxonomic associates: Semantic field cues elicited the highest proportion of coordinate and subordinate responses and brand cues elicited the highest proportion of superordinate responses. However, the total proportion of response given across the three general response category types were quite evenly balanced: Participants produced a total of 34% linguistic associates, 27% taxonomic associates, and 38% object or situation descriptors (conceptual associates). This shows that, contrary to their second hypothesis, conceptual associates were produced slightly more frequently than linguistic ones. Conceptual associates accounted for the highest proportion of responses given to four of the eight cue word categories, namely synonym, taxonomic, object property, and brand cues.

The first hypothesis was that, in line with LASS theory, linguistic associates would be produced earlier and more quickly than conceptual ones. To test this hypothesis, they calculated the average output position (response one, two, three, etc.) for each unique response word as a function of the response type. The average output position for unique linguistic associates was 1.61, for taxonomic associates it was 2.03, and for conceptual, 2.47. These average output positions were entered into a cue word concept type x general response code ANOVA. The results confirmed the prediction of LASS theory: Linguistic associates were produced significantly earlier than conceptual ones. In addition, taxonomic associates were intermediate between the two, and all pairwise comparisons were statistically significant.

It should be noted that this second analysis did not include all responses given, but the 1227 unique response words provided to each cue. Although they did not state the total number of associates produced (or the number that were coded as "none"), they report that 40 participants provided an average of 1.74 responses to each cue word, which yields approximately 4454 responses. From this information, it can be calculated that the same response word was given by an average of 3.63 people. An alternative calculation is that they condensed responses down to 27.5% of the total responses collected prior to the statistical analysis of average output position. This can be viewed as an unweighted calculation of output position, as each unique response word to each cue contributes equally to the analysis. Santos et al. (2011) also calculated the mean time to produce the first response. This analysis was not reported in detail but was stated to show a similar pattern to that observed for average output position.

In interpreting their results, Santos et al. (2011) referred to the WAT, the property generation task, and the neuroimaging data. However, only the conclusions relevant to the WA task are highlighted here. They stated that their results indicate that WA may not be purely a language task, but may also access conceptual processing, in other words, situated simulations. They also suggested that variance in the dominant type or types of association elicited by different cue words means the cue or stimulus words used in psycholinguistic tasks might affect behavioural outcomes. Any such effects will vary according to whether the task accesses conceptual or linguistic knowledge. However, they were also clear that the distinction between whether a response to a stimulus word is produced by accessing linguistic or conceptual knowledge should be viewed as a cline, rather than an absolute divide. In other words, a conceptual association may involve lexical semantics and a linguistic association may involve conceptual semantics. Likewise, they assumed that these two systems interact, such that a linguistic association can lead to the generation of a conceptual one and vice versa. Finally, they suggested that this assumption requires further investigation.

3.1.5 Original and Replication Experiment Hypotheses

To recapitulate, the main aim of the experiment reported in this chapter is to replicate Santos et al.'s (2011) findings. Accordingly, the first hypothesis is that linguistic responses will be produced earlier and more quickly than conceptual ones, in line with LASS theory (Barsalou et al., 2008). No firm prediction is made regarding the relative output position or timing of taxonomic associates (as no prediction was made in the original experiment), although it is noted that such associates were produced later than linguistic responses and earlier than conceptual ones in the original experiment.

The results of the original experiment did not support the second hypothesis, that participants should produce more linguistic associates than conceptual ones, as WA is a language-based task. In the replication it is predicted that, if WA behaviour is stable, as in Santos et al. (2011), participants will produce a similar proportion of linguistic and conceptual associates. In line with the original experiment, and considering the removal of brand cues, participants in the replication are expected to produce fewer taxonomic associates.

3.2 Replication Experiment Method

The experiment was constructed to replicate that of the original study by Santos et al. (2011) as closely as possible, based on the information they provided. Their cue word choice and methods were described in detail above (see Sections 3.1.2 and 3.1.3). Thus, this section focuses on the few changes that were made to the original experiment and decisions based on incomplete information. All materials and the procedure were approved by Cardiff University Ethics Committee.

It should be noted that in referring to the type of association (e.g., collocation, synonym) Santos et al. (2011) used the term *concept* to describe the cue words and *category* to describe response words. In this chapter (and all subsequent ones) the word *category* is used to refer to both cue and response types. This change was made for clarity. First, it seems more homogenous to use one word to describe the type of association the cues are predicted to elicit and the type of association between response and cue words. Second, object and situational descriptors are conceptual associations, so it is potentially confusing to use the term *concept* to refer to all cue words, including those that are predicted to elicit linguistic and taxonomic associates.

3.2.1 Materials and Cue Words

The 64 cue words used by Santos et al. (2011) were listed by category type in the appendices. The same cue words were used in the replication, coded identically, except that, as explained below, the eight brand names were omitted (see Appendix 2). Four lists of 14 cue words were created, with two words from each concept type on each list (all lists were also reverse ordered to create eight lists). As Santos et al. did not specify the practice words used, I chose two different words to add to each list. To limit any influence of the practice words on the cue words, the main grammatical word classes (noun, verb, adjective, adverb) were avoided. They also did not specify which two words from each category were assigned to each list, or the order in which the cue words were presented. Therefore, cue words were randomly assigned and ordered on each list, using an online randomiser (see

Appendix 3). In accordance with Santos et al., I ensured that no two words from the same category were adjacent and inspected the word lists for semantic relatedness.

As noted above, brand names were omitted in the replication due to theoretical and methodological concerns, reducing the set of cue words to 56. The theoretical motivations were: (a) no rationale was provided for the inclusion of brand names and this concept or category type was not used to code responses; (b) responses within this category were not discussed beyond noting that they elicited a high proportion of taxonomic associates (they elicited a similar proportion of conceptual associates and few linguistic ones, but this was not mentioned); and (c) no consideration was given of the potential effect of including proper names within a lexical experiment. These observations make it difficult to discern the relevance of the brand cues to the hypotheses being tested, concerning linguistic and conceptual processing. At the methodological level, three of the brand names would not be familiar to British participants, so would have required substitution (Crest, Tylenol, and Speedstick). More importantly, only three of the eight brand names used (Crest, Tylenol, and McDonald's) appear in the USF norms (Nelson et al., 1998), so their assignment to a cue category type would appear to have been made on the basis that these cue words were brands. All other cue words were chosen based on the type of association between a dominant response and the cue in the USF norms. It seems rather unsystematic to change the selection criteria for a subset of cue words. For these five reasons, I decided that the omission of the brand cues was unlikely to have a detrimental effect on the reliability or validity of the results or the ability to replicate the original findings.

3.2.2 Participants

All 160 participants were L1 English-speaking students at Cardiff University (predominantly aged 18 to 22, although some were mature students). This is a similar demographic to that in the original study, with the one major difference that participants were studying at a British, rather than an American university. It is possible that the participants in the original study included staff members. Santos et al. (2011) did not explicitly state that all participants were L1 English speakers, but this was assumed here. In total, there were 103 female and 57 male participants, with 40 participants providing responses to each cue word. Unlike in the original study, no participant received any financial compensation for taking part in the replication reported here.

3.2.3 Procedure

As in the original study, I approached potential participants individually in various campus areas and asked if they were willing to take part in a short language experiment. If they agreed, they were asked to give informed written consent before proceeding, as approved by the Ethics Committee at Cardiff University. The instructions given to each participant were as close as possible to the original: Participants were asked to say as many responses as they could to each cue word, and it was stressed that there were no correct answers, they should just say the first thing that came to mind, spontaneously. Participants were told that when they couldn't think of any more words they would be stopped, to move on to the next cue word. As in the original experiment, participants were unaware the first two words were for practice only. I repeated the same words as in the original study before each cue word, "For the following word, what other words come to mind immediately?" As soon as a participant paused, I moved on to the next cue prompt. The WAT typically took less than 2.5 minutes to complete and was recorded on a digital recorder. Responses to the 14 experimental cues were later entered anonymously into a spreadsheet for coding and analysis.

One issue that arose during the experiment was determining how long to wait until participants gave a single (first) response to a cue word. Santos et al. (2011) specified that they moved onto the next cue prompt if participants paused, but they did not give any information about the average length of time until the first response was given. Nor did they state whether this was restricted to prevent participants pausing before the first response. In conducting the experiment, I judged that interrupting participants before they provided a single response would inhibit responses to subsequent cues, reducing the total number of responses for analysis. Therefore, I waited for 5 seconds before moving onto a new cue prompt unless the participant indicated that they could not think of a response.

3.2.4 Data Coding

In coding the data, several issues arose, which are described next, before reporting the results. Firstly, as noted above, some responses were produced after a considerable delay. Santos et al. (2011) explained that participants were stopped as soon as they paused, in order to prevent participants having time to access the conceptual system, and therefore maximise linguistic responses. Although they stated that participants produced responses over an average period of 2.98 seconds, they did not give any information about the time until the first response. It would seem to confound the aim of the study if participants were allowed to pause before giving the first response as, if the assumptions of LASS theory are correct, this would increase the probability of the first response being produced via conceptual processing.

For this reason, I decided to limit analysis of response type frequency and output position (first, second, third response) to those responses with an onset within three seconds of the cue word presentation offset. This time frame was chosen partly based on the average response period in the original study, but also because in a recent timed WA study, Playfoot et al. (2016), stated that 70 per cent of L1 English speakers produced word association responses within 2.7 seconds. However, it means that first responses to cue words given after three seconds were excluded from position-based analyses (they were included in the analysis of time to first response, reported in Section 3.4.3). It also means that subsequent responses which were given without pausing, but after the three second time period, were omitted from the analyses reported below. After these exclusions it was calculated that participants produced an average of 1.74 responses to each cue word, within three seconds. This is exactly the same number of responses as in the original study, but produced within a fixed time frame across participants.

In addition, Santos et al. (2011) stated that very few responses did not fit into one of their valid response categories (p. 91), (coded as none), suggesting that coding responses was a very straightforward process. However, in the replication, I encountered occasional issues with some of the one-syllable cue words (20 words). Sometimes a word was miscommunicated, for example, one participant responded to the word “dull” instead of the cue word “doll”. Such responses were easily identified and were coded as wrong cue. In other instances, a response was given to a homonym of the intended cue word. This was more problematic, as it is less clear that responding to the letter “B” instead of the animal “bee” indicates that the cue word was miscommunicated. One participant even responded to both, producing the responses “C” and “buzz”. The policy adopted was to code associates to homonyms with orthographic differences as wrong cue, such as the response “John” to the cue “cane” (the politician John Cain). Homonym associates to identical cue words were coded according to the relation between the words, so the associate “pillow” to the cue “throw” was coded as a coordinate taxonomic response, both being soft furnishings.

In addition, the data from one participant was discarded and replaced during the study. The exclusion was made as half the responses given by the participant were multi-word phrases, rather than individual words.

More importantly, two cue words were excluded from the analyses as over half the associates produced in response were problematic (described below). One word was from a linguistic response category (synonym) and the other from a conceptual response category (object property).

The cue word in the synonym category that was removed was “guess”. Unfortunately, one of the two practice items added to the lists including this cue word was “who”, which may have primed the forward collocation response “who”, as ten people gave this response (although “who” was the fourth dominant response word in the USF norms). In addition, a further three people failed to produce any response, three people took longer than three seconds to produce a response, four responded to the cue word “guest”, two gave a response with no discernible relation to the cue word, and one person responded with a multi-word phrase. This meant that only 17 of the 40 participants presented with this cue word gave a valid response, which it was decided was too small a sample size. As this cue word was predicted to produce synonym associates as the dominant response, yet only two people actually did so, removing this item can only improve the likelihood of the analyses replicating the results of the original study.

The other cue word that was removed was “stubborn”, in the stereotypical object property category. This item was removed as nine people failed to produce any response, three people took longer than three seconds to produce a response and three gave responses with no discernible relation to the cue word. In addition to these 15 participants, six people gave 12 responses that had only a weak semantic connection to the cue word, such as ‘angry’ or ‘grumpy’. The remaining 19 participants produced 10 words that were people or animals for which stubborn was a stereotypical object property. However, as the other cue words in this category elicited mainly object property responses, exclusion of this word increased the overall proportion of predicted responses within this cue category.

3.2.5 Data Analysis

Analyses were conducted in RStudio (RStudio Team, 2020), run on R (R Core Team, 2021), using the tidyverse (Wickham et al. (2019) and vcd packages (Meyer et al., 2006; 2022). Coding was based on various sources, notably Field et al. (2012) and Levshina (2015). Figures were created using ggplot2 (Wickham, 2016) and the wesanderson package (Ram & Wickham, 2018), following Healy (2019).

3.3 Replication Experiment Results

The analyses conducted by Santos et al. (2011) were performed on the data collected in the replication experiment. They reported two main analyses in two tables: (a) the proportion of responses of each specific and general association type as a function of the cue word concept or category type, and (b) the mean output position of each specific and general association type as a function of the cue word concept or category type, based on the median output position of each unique response word. In addition, they performed ANOVA analyses on the median output positions for each general association type. Although I intended to attempt an exact replication of Santos et al.'s analyses, some changes were made and additional analyses are also reported, as explained later in this chapter.

3.3.1 Analysis One: Distribution of Response Types

First, the proportion of each type of response was calculated as a function of the cue word category types, as shown in Table 3.1. As in Santos et al. (2011), the proportions of responses of each category type produced to each cue word type are shown, such that the total for each of the eight types of cue word sum to one. I decided to report the number of responses coded as none (no or clang responses), to provide complete information about the data collected. Such associations were removed by Santos et al. prior to analysis and not reported. Although they are reported here, these responses (6% of the total) were not entered into any subsequent analyses.

As noted above, in the replication, participants produced exactly the same average number of responses per cue (1.74) as in the original study. The response time period for the number of responses and median output position analyses was restricted to 3 seconds, in line with the stated average production period of 2.98 seconds reported by Santos et al. (2011). For the general response categories, they reported participants produced "34% linguistic responses, 27% taxonomic responses, and 38% object-situation responses" (p. 95). The proportions obtained in the replication, with the inclusion of 6% clang responses, were similar, at 31%, 22%, and 41%, respectively. Fewer taxonomic associates were anticipated, given that brand name cues were removed, which were reported to elicit taxonomic and conceptual responses in the original study.

Table 3.1

Proportion of Responses in Each Cue Category by Type

Response Category	Cue word category							Overall Mean
	FC	BC	SN	AN	TX	OP	SF	
FC	**0.26**	0.13	0.05	0.15	0.04	0.02	0.05	0.10
BC	0.02	**0.24**	0.00	0.02	0.02	0.01	0.00	0.04
SN	0.06	0.09	**0.30**	0.12	0.00	0.06	0.02	0.09
AN	0.02	0.00	0.06	**0.32**	0.00	0.12	0.00	0.07
RS	0.02	0.00	0.04	0.01	0.00	0.01	0.01	0.01
SS	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
DH	0.05	0.08	0.01	0.01	**0.26**	0.02	**0.13**	0.08
DS	0.08	0.02	0.04	0.01	**0.30**	0.08	**0.36**	0.13
DL	0.00	0.01	0.00	0.00	**0.00**	0.00	**0.04**	0.01
OS	0.45	0.38	0.45	0.32	0.34	**0.64**	0.26	0.41
NO	0.04	0.05	0.05	0.04	0.03	0.05	0.14	0.06
Linguistic	0.37	0.46	0.45	0.63	0.06	0.21	0.07	0.31
Taxonomic	0.14	0.11	0.05	0.01	0.57	0.10	0.53	0.22
Conceptual	0.45	0.38	0.45	0.32	0.34	0.64	0.26	0.41

Notes: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, TX = taxonomic category, OP = object property, SF = semantic field, RS = root similarity, SS = sound similarity, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OS = object or situation descriptor, NO = no or clang response. The proportions of responses in the predicted categories for each cue type are indicated in bold font.

Within the specific categories, the participants in the replication produced substantially fewer collocations to such cues than in Santos et al. (2011), instead producing more object and situational descriptors. In contrast, participants produced slightly higher proportions of synonym and antonym responses to those cue types. The high proportion of domain higher and same responses to taxonomic cues and domain same responses to semantic field cues was similar here to in the original experiment, as were the proportion of object and situational descriptors produced to object property cues. Overall, it would seem that the WA behaviour of the participants in the replication was similar to that of the participants in the original study and that the coding system was applied reasonably consistently here, in relation to the original authors. This is important, as it suggests that both the WAT and the system for coding responses are reliable.

However, the absolute number of responses produced to each cue word type varied considerably in the replication, from 475 responses to the synonym cue words to 634 responses to the semantic field cues. Reporting proportions conceals this difference, as the proportion of responses to each cue word type is made to sum to one. Therefore, the absolute number of each type of response as a function of the cue word category types is shown in Table 3.2.

Table 3.2

Number of Responses in Each Cue Category by Type

Response Category	**Cue word category**							**Total**
	FC	BC	SN	AN	TX	OP	SF	
FC	**137**	71	26	80	21	12	29	376
BC	8	**133**	0	9	11	3	2	166
SN	32	48	**143**	64	2	27	10	326
AN	8	0	27	**165**	0	56	0	256
RS	10	1	20	6	1	3	4	45
SS	0	1	0	3	1	1	0	6
DH	28	47	6	3	**153**	8	**85**	330
DS	44	11	17	3	**179**	39	**225**	518
DL	0	3	0	0	**0**	1	**25**	29
OS	241	212	213	163	199	**308**	166	1502
NO	23	28	23	21	20	24	88	227
Total	530	555	475	517	587	482	634	3780
Linguistic	**195**	**254**	**216**	**327**	36	102	45	1175
Taxonomic	72	61	23	6	**332**	48	**335**	877
Conceptual	241	212	213	163	199	**308**	166	1502

Notes: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, TX = taxonomic category, OP = object property, SF = semantic field, RS = root similarity, SS = sound similarity, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OS = object or situation descriptor, NO = no or clang response. The number of responses in the predicted categories for each cue type are indicated in bold font.

It should be noted that the comparatively low number of responses to synonym and object property cue words is due to the removal of responses to one cue word from each of these categories prior to analysis, as noted above. In addition, the high number of responses to semantic field cues is partly due to the increased number of responses coded as none or clang responses (88). This large number is due to coding responses as chained responses when participants listed items in order from lexical sets (responding "Tuesday, Wednesday, Thursday" to the cue "Monday" or "September, October, November" to the cue "December"). It is not known how Santos et al. (2011) coded such responses (assuming they occurred in their data set).

As in Santos et al. (2011), both the proportions and numbers of each response type show that participants produced a large number of conceptual associates, rather than relying on linguistic responses, as hypothesized in the original study. In this replication, conceptual associates were the most frequent type of response given to all cue word types except semantic field cues, which elicited mostly coordinate associates (same category members), and antonym cues, which elicited almost identical numbers of conceptual and antonym associates. However, as in Santos et al., more responses of the predicted type were given within each category of cue words than were given to any of the other cue word categories (e.g., most forward collocations were produced to cues predicted to elicit that type of association).

3.3.2 Analysis Two: Median Output Position by Response Type

As in Santos et al. (2011) the next analysis involved the calculation of the median output position of each unique response word provided to each cue word. A total of 1331 unique responses were given, which is greater than that reported in the original study (1227). This is despite not using the brands category of cue words and the removal of two of the cue words prior to analysis. This means that the participants in this study showed less conformity or more creativity in their associations than those in the original study.

From these median values, I next calculated the mean output position for specific and general response categories as a function of cue word type, as shown in Table 3.3. It is of note that the mean output positions are much earlier values than those reported by Santos et al. (2011). The overall mean output positions for each general response type in the original study was 1.61 for linguistic associates, 2.03 for taxonomic associates, and 2.47 for object and situation descriptor (conceptual) associates. Thus, the mean output positions for all three types of associates were earlier in the replication. Despite this difference, the order across output positions was the same, with linguistic associates being produced earliest and conceptual associates produced slightly later than taxonomic ones.

Table 3.3

Mean Output Position by Cue and Response Category Type

Response Category	**Cue word category**							**Mean Position**
	FC	BC	SN	AN	TX	OP	SF	
FC	1.48	1.32	1.44	1.26	1.33	1.41	1.69	1.43
BC	2.08	1.56		1.83	2.25	1.50	2.33	1.77
SN	1.44	1.33	1.49	1.51	1.50	1.42	1.42	1.47
AN	1.00		1.50	1.24			1.61	1.34
RS	1.20	1.00	1.38	1.00	3.00	1.00	1.33	1.32
SS		1.00		1.00	1.00		1.00	1.00
DH	1.81	2.13	1.13	2.00	1.64	1.44	2.00	1.65
DS	1.68	1.60	1.33	2.00	1.72	1.84	2.06	1.76
DL		1.00				1.56	2.00	1.55
OS	1.70	1.88	1.78	1.63	1.98	1.89	1.60	1.77
Linguistic	1.44	1.24	1.45	1.31	1.82	1.33	1.56	1.46
Taxonomic	1.75	1.58	1.23	2.00	1.68	1.61	2.02	1.71
Conceptual	1.70	1.88	1.78	1.63	1.98	1.89	1.60	1.77

Note. Empty cells indicate response types that were not provided by any participant for a particular cue word category. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, TX = taxonomic category, OP = object property, SF = semantic field, RS = root similarity, SS = sound similarity, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OS = object or situation descriptor.

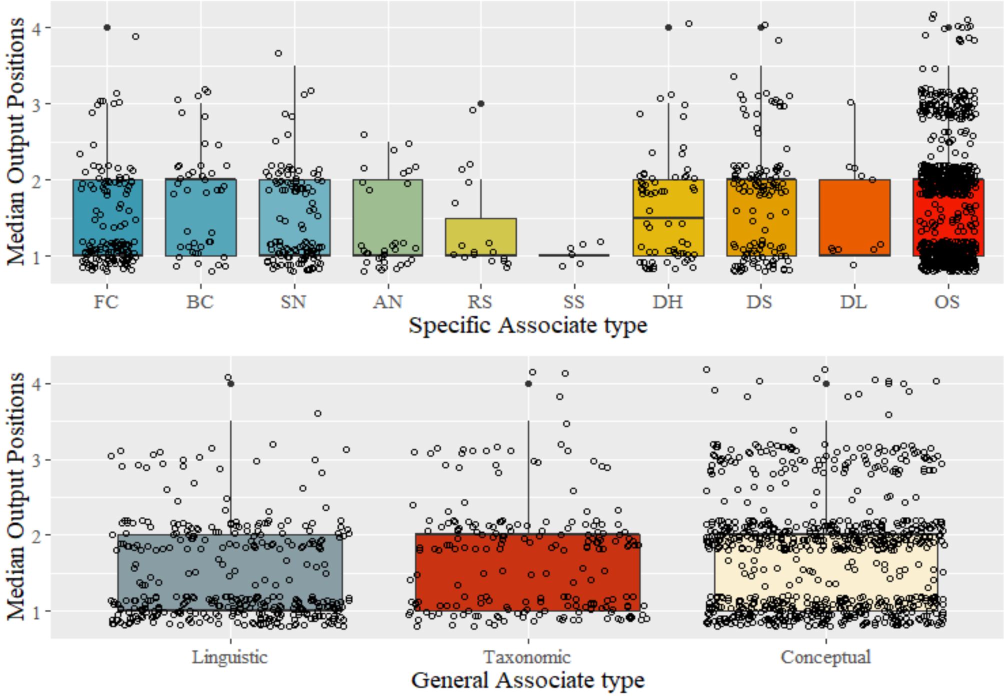
The differences in values for mean output positions found here compared with those reported in Santos et al. (2011) should be interpreted in light of the identical average number of responses provided per cue word, 1.74. This means that in both studies, on many trials only one response word was provided. Two, three, or four responses were provided less frequently. After removing all clang responses (including 33 instances of no response), in the replication there were 2064 first responses. A valid second response was produced on only 53% of trials (1086), a third response on 18% of trials (366), and a fourth response on less than 2% of trials (37). It may be that in the original experiment, participants tended to produce either one or three responses, with few instances in which two responses were produced. Otherwise, it is not clear how conceptual associates could account for 38% of all valid responses in the original experiment, yet unique conceptual associates could have an average output position of 2.47, as reported by Santos et al.

Although Santos et al. (2011) reported an ANOVA on the mean output positions for linguistic, taxonomic, and conceptual responses, a non-parametric version was conducted here, as the data violated assumptions of parametric tests. The mean output positions reported in Santos et al. were very different to those obtained here, so this could explain this discrepancy. The results from a Shapiro–Wilk test showed that responses were far from normally distributed (linguistic responses: W = 0.73, taxonomic responses: W = 0.82, conceptual responses: W = 0.82, all *p.* < .000001) and a Levene's test showed that the variance was also extremely unequal ($F(2, 1328) = 9.247, p = .0001$). This can be clearly seen in the jittered boxplots shown in Figure 3.1. There were many more unique responses with a median output position of one, compared with all other output positions (i.e.., the data was negative skewed). In addition, there were many more conceptual responses than linguistic or taxonomic responses, especially at a median output position of two or three. Unfortunately, in conducting non-parametric tests it was impossible to enter cue word category type as a random factor, as in Santos et al., so the effect of this difference on the results is unknown.

The results of a Kruskal-Wallis test showed that the difference between mean output position for the three general category types was highly significant (H(2) = 44.833, *p.* < .000001). A similar result was found when comparing the specific category types. Post hoc pairwise comparisons were performed using the pgirmess package (Giraudoux, 2022). These tests showed that there was a significant difference in the mean output position of linguistic compared with taxonomic and conceptual associates, but the difference between taxonomic and conceptual associates was not significant (linguistic-taxonomic: critical difference = 80.44; observed difference = 126.86; linguistic-conceptual: critical difference = 58.57; observed difference = 150.82; taxonomic-conceptual: critical difference = 72.54; observed difference = 23.96).

Figure 3.1

Median Output Position for Specific and General Associate Types



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, RS = root similarity, SS = sound similarity, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OS = object or situation descriptor.

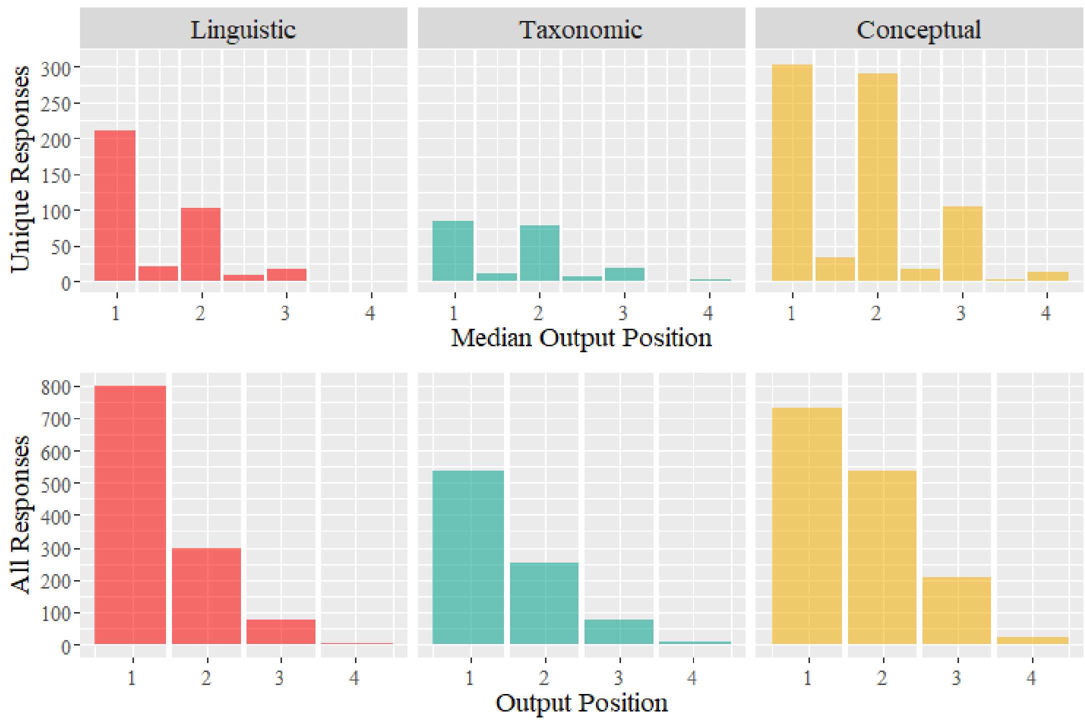
Comparisons between the specific category types suggested that the differences were driven by a significant difference between the output positions of forward collocations and coordinate category members (domain same taxonomic associates) and between both forward collocations and synonyms in comparison with object and situational descriptors. The difference between antonyms and object and situational descriptors also approached significance. However, due to the vast differences between response category sizes, especially between object and situational descriptors and all other specific category types, these results should be interpreted with caution.

In line with the original experiment, the median output position of unique linguistic associates was earlier than that of taxonomic and conceptual (object and situational descriptors) associates. However, backward collocations had the same overall mean output position as object and situational descriptors, so linguistic responses were not uniformly earliest. In addition, there were more unique conceptual associates given as first responses than linguistic ones, so the explanation may be less that linguistic responses were given earlier and more that they were not produced later.

Furthermore, in this analysis, the median output position for each unique response was first calculated, and these values were then used to calculate mean output position for each type of associate and to assess the effect of associate type on output position. In doing so, the amount of information lost per unique response word varied dramatically. For many associates no information was lost, in other words, only one person produced that response (it was truly unique). On the other end of the scale, some cue words elicited a highly dominant response, which was then treated equivalently to a unique response. At the highest extreme, 35 out of 40 people produced the response "nut" to the cue "cashew". The impact of this data loss is an issue that the original authors failed to raise. The extent of this data compression in the replication experiment, as a function of general association type, is shown in Figure 3.2.

Figure 3.2

Output Positions of Unique and All Responses by Associate Type



Note. The difference in the width of the bars reflect that median output position included responses in tied positions (1.5, 2.5, and 3.5), but has no significance.

Figure 3.2 shows that there were proportionately fewer unique linguistic and taxonomic associates than conceptual ones. In particular, a relatively low number of unique taxonomic associates were produced, indicating greater conformity between participants for this type of response. This means that although there was no statistical difference between the mean output position of taxonomic and conceptual associates, there was a large difference in the number of unique responses of each type. Considering all responses, both linguistic and taxonomic associates are mostly produced as a first response, although this trend is more pronounced for linguistic associates. In contrast, many more conceptual responses are produced as a second response.

3.3.3 Summary of Replication Analyses

The analyses reported above partially replicated the results of the original experiment, both in terms of the proportion of linguistic, taxonomic, and conceptual associates produced by participants, and the difference in the median output positions of such responses, with conceptual responses being produced later. However, the production of more conceptual than linguistic (or taxonomic) associates further refutes Santos et al.'s (2011) hypothesis that WA would predominantly elicit linguistic associates. There were also substantial differences in the replication experiment in the mean output positions of general and specific types of response in comparison with the original. The original ANOVA analysis could not be replicated due to the violations of the normal distribution and homogeneity of variance of the data. Visual inspection of the data also showed that this analysis resulted in unequal data compression, as there were more unique conceptual associates than linguistic or taxonomic ones. Moreover, in replicating the experiment a potential issue was discovered regarding the selection of cue words for each category type. Therefore, further analyses were conducted, as reported next.

3.4 Additional Analyses of the Data

Three additional analyses were conducted on the data collected in the replication experiment. These analyses are identical or similar to those reported by Santos et al. (2011), yet also introduce important changes. First, the distribution of response types was recalculated, based on the reclassification of many of the cue words according to the dominant associate in the USF norms (Nelson et al., 1998). In addition, the output position of each response type was analysed, without reducing the data to median output positions (i.e., the analysis was performed on the complete dataset). The time of the onset of the first response is also reported, an analysis which was conducted by Santos et al. but was only reported in a footnote. No reason is given as to why this analysis was not reported in more detail, so space constraints are assumed. However, given that calculating the median output position led to unequal data compression across the three main response categories, I felt that it was important to discover whether linguistic associates were produced faster than conceptual ones, as in the original experiment.

3.4.1 Recategorisation of Cue Word Types

In the explanation given of the cue word selection, Santos et al. (2011, pp. 93-94) referred to “a dominant word associate” not *the* dominant associate. In addition, I noticed that “slow” was classified as an object property cue word, yet the dominant response in the replication was the antonym “fast” (44% of all responses). As described above, Santos et al. used cue type as a factor in both the analyses they reported, indicating they expected the cue word type would influence response types. Therefore, it would seem most appropriate to define the association type of each cue word by the dominant associate given in the USF norms (Nelson et al., 1998). I decided to examine the USF norms to confirm whether this was the case.

The dominant response for each cue did not consistently match the coding system used by Santos et al. (2011). In the USF norms (Nelson et al., 1998), 52.7% of responses to the cue word “slow” are the antonym “fast” (compared with 11.5% of responses “turtle”). Similarly, “bee” was classified as a taxonomic cue, although the dominant response is “sting” (36.2%) and the most frequently given taxonomic response is “insect” (5.5%). Appendix 4 shows all cue words, as categorised in Santos et al., together with the dominant response and the percentage of respondents who gave it. When the dominant response is not of the type that the cue was coded, the most frequently given response of the cue category type is also shown, with the percentage of respondents who gave it.

One possible explanation for these classification decisions is that the original study may have at times used dominant elicitors of a response word, rather than dominant responses to a cue word. For example, the response word “snail” is elicited by the cue word “slow” more frequently than by any other cue word. However, such asymmetry between the dominant response to a given cue and the dominant cue for a given response is quite typical in WA norms. As Tversky (1977) noted, “A common pattern observed in such studies is that the more salient object occurs more often as a response to the less salient object than vice versa” (p. 337).

In view of this issue, the cue word categories were recoded, according to the dominant or primary response in the USF norms (Nelson et al., 1998), see Appendix 5. One cue word, “September”, was not found in the norms, so it was kept in the original category (semantic field). Two object property cues, “horns” and “smelly” appear in a different form in the USF norms (“horn” and “smell”), so the category for these cues was not changed. In the replication, all three words elicited responses that were mostly of the predicted type.

The changes made to the classification of cue types[4] are summarised in Table 3.4, which shows the number of cue words of each specific and general type in the original study and after reclassification of some cues based on the dominant associate in the USF norms. A new category was created for associates that held a conceptual association to the cue word, but were not object properties, such as the response “money” to the cue “financial”, or the response “green” to the cue “parsley”, where the response is a property of the cue, rather than the cue being a property of the response word. Although the number of cues in each general response category remained similar, only two of the original synonym cue words were retained in that category. Overall, only four cue words were classified as synonyms and four as backward collocations. However, as in the original experiment, the cue words were predicted to elicit more linguistic associates than taxonomic or conceptual ones, a point which is discussed later in this chapter.

[4] The cue word “throw”, originally classified as a synonym, elicited the responses “ball” (forward collocation) and “catch” (antonym) with equal frequency (23.4%) in the USF norms. It was recoded as an antonym cue type as this was the dominant response type in the present study.

Table 3.4

Reclassification of Cue Types Based on the USF Norms

	Specific cue type								General cue type		
	FC	BC	SN	AN	TX	SF	OP	CO	Ling	Tax	Conc
Original cue classification	8	8	7	8	8	8	7	NA	31	16	7
Retained cue words	4	4	2	6	6	6	5	NA	16	12	5
Reclassified cue words	6	0	2	3	3	0	2	5	11	3	7
Total cue words	10	4	4	9	9	6	7	5	27	15	12

Notes: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, TX = taxonomic category, SF = semantic field, OP = object property, CO = conceptual non-property, Ling = linguistic, Tax = taxonomic, Conc = conceptual.

Table 3.5 shows the number of responses of each type as a function of the original classification of the cue words and when they were recoded according to the dominant associate in the USF norms (Nelson et al., 1998). To improve clarity, the few root and sound similarity responses and no or other non-target responses have been removed and all three types of taxonomic response have been combined.

Table 3.5

Number of Responses in Each Cue Category by Type

Response Category	**Original cue word category**								**Total**
	FC	BC	SN	AN	TX	SF	OP		
FC	**137**	71	26	80	21	29	12		376
BC	8	**133**	0	9	11	2	3		166
SN	32	48	**143**	64	2	10	27		326
AN	8	0	27	**165**	0	0	56		256
TX	72	61	23	6	**332**	**335**	48		877
OS	241	212	213	163	199	166	**308**		1502
	Recoded cue word category								**Total**
	FC	BC	SN	AN	TX	SF	OP	CO	
FC	**188**	18	17	85	16	7	0	45	376
BC	20	**103**	1	9	14	2	17	0	166
SN	19	48	**123**	57	12	6	38	23	326
AN	1	0	25	**226**	0	0	4	0	256
TX	154	7	3	8	**337**	**271**	47	50	877
OS	240	73	63	190	270	131	**355**	**180**	1502

Notes: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, TX = taxonomic category, SF = semantic field, OP = object property, CO = conceptual non-property, OS = object or situation descriptor. The proportions of responses in the predicted categories for each cue type are indicated in bold font.

Taking into account that the number of cue words in each category varied from the original system, it can be seen that reclassifying the cue word types increased the number of responses in the predicted dominant response category. In particular it can be seen that halving the number of backward collocation and synonym cues led to a substantial reduction in the number of object and situational descriptor responses to these cue words, without greatly reducing the numbers of predicted response types. In addition, the new category of conceptual cue words elicited mostly object and situational descriptor responses. However, after recoding, forward collocation cues elicited many more taxonomic associates. However, the same two main trends can be seen using both the original and the revised cue word categories: All cues elicited many responses of the predicted type as well as many object and situational descriptor (conceptual) associates.

3.4.2 Output Position by Response Type

In the original experiment, Santos et al. (2011) calculated the median output position of unique responses, enabling statistical analysis of the data. However, as noted earlier, it also compressed the data considerably and unequally (a higher proportion of conceptual associates were unique). Therefore, I decided to complement this analysis with descriptive analysis of all the WA data collected. Table 3.6 shows the observed frequencies of specific and general response types in each output position, and the difference from the expected frequencies, if all eight specific and three general response types were evenly distributed across each output position. Sound and root similarity responses (51 responses) and no or other responses (220 responses) were removed, as these were not predicted (target) responses to any of the cue word types. Note that because the total number of responses in each output position varied greatly, expected frequencies across all response types and all output positions cannot be calculated. As there were only 37 instances in which a participant gave four responses (i.e., 1% of all associations), only responses across the first three output positions are shown.

Considering the general response categories, participants produced fewer taxonomic associates and more conceptual associates than expected under an equal distribution of response types, across all three output positions. Of note, the over-representation of conceptual associates is much greater for the second and third responses. This is because the number of linguistic associates produced was greater than expected as the first response but fewer as second and third responses. For the eight specific categories, the impact of not dividing conceptual associates into sub-types dominates the results, given the far greater number of object and situational descriptor responses compared with all other specific response types. Notwithstanding, domain same or coordinate responses also elicited a slightly higher number of responses across all three output positions than would be expected under an equal distribution.

Table 3.6

Observed Compared with Expected Frequencies of Response Types

Response type	Response output position					
	One		Two		Three	
	Obs	Dif	Obs	Dif	Obs	Dif
FC	265	11	85	-50	26	-19
BC	96	-158	51	-84	18	-27
SN	210	-44	96	-39	19	-26
AN	191	-63	55	-80	10	-35
DH	211	-43	91	-44	25	-20
DS	310	56	151	17	49	4
DL	13	-241	10	-125	5	-40
OS	733	479	537	403	209	164
Linguistic	797	109	297	-65	78	-44
Taxonomic	535	-153	252	-110	79	-43
Conceptual	733	45	537	175	209	87

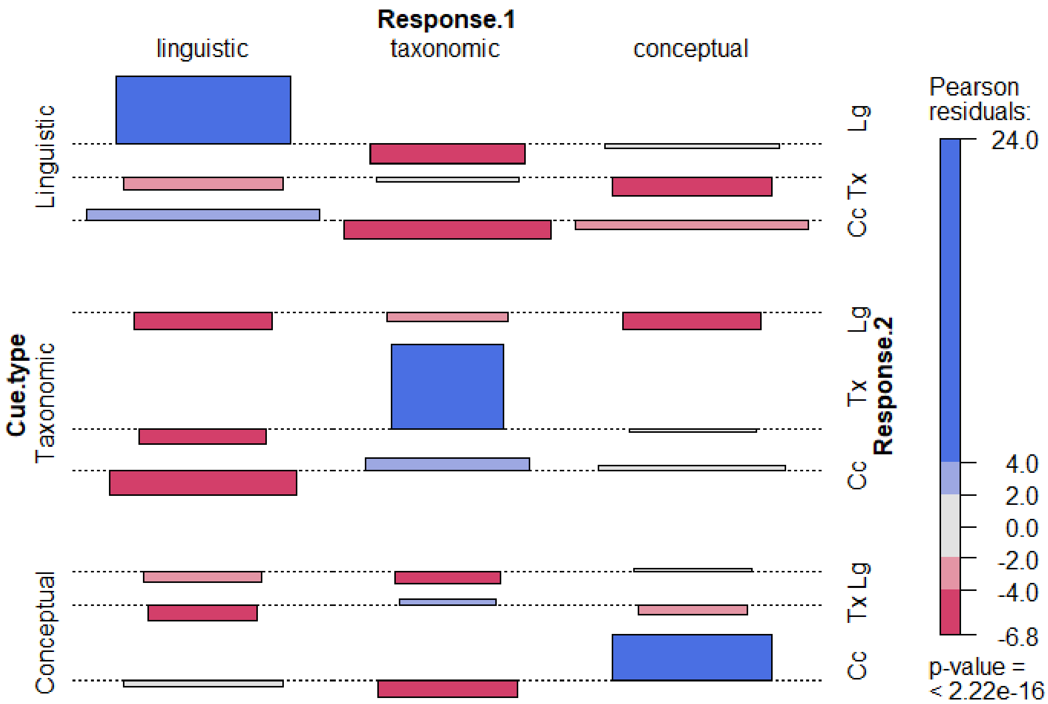
Notes: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OP = object property, OS = object or situation descriptor, Obs = observed number, Dif = difference from equal distribution.

This reanalysis partially supports the claim of LASS theory (Barsalou et al., 2008) and the prediction of Santos et al. (2011), that linguistic processing is accessed earlier. However, half the cue words elicited a dominant response in the USF norms (Nelson et al., 1998) that entered into one of the four types of association classified as linguistic. Thus, if, word association behaviour is relatively stable across different populations of L1 English speakers, at least the first response should be linguistic more often than taxonomic or conceptual due to the cue words used.

For this reason, further analysis was performed on a subset of the data (2148 responses) comprising those instances in which a participant produced two or more responses within three seconds, both of which held a linguistic, taxonomic, or conceptual relationship to the cue word (i.e., any pair of responses in which one word held no discernible relation to the cue were removed). The general type of relation of both responses was calculated as a function of the cue word type, to assess whether the observed number of response types differed from the expected number under an equal (chance) distribution. A Cochran-Mantel-Haenszel test with Bonferroni correction showed that response one type was significantly associated with response two type after accounting for the cue word type ($M^2 = 222.04$, df = 4, *p*. < 0.00001). These associations are visualized in an association plot in Figure 3.3, which shows the residuals (the differences) for each combination of response types.

Figure 3.3

Association of Cue Type with First and Second Response Types



Notes: Blue bars indicate higher numbers than expected and pink bars indicate lower numbers; deep hues indicate a significant difference. Lg = linguistic, Tx = taxonomic, Cc = conceptual.

The width of the bars represents the difference in the number of responses and the height indicates the significance, such that the area of each bar represents the difference of the observed number of response types from that expected under an equal (chance) distribution (deep blue upward bars indicate significantly more responses, deep pink downward bars indicate significantly fewer responses). This shows that for each cue word type, participants showed a robust tendency to give two associates of the general type predicted for that cue type. In addition, there was a weaker trend for participants to produce a conceptual associate as the second response to linguistic and taxonomic cues, after providing a response of the predicted type. However, this result should be interpreted with caution, given the greater number of linguistic cues. Conceptual cues exerted greater influence on response types than linguistic ones: 37 percent of linguistic cues elicited two linguistic associates, compared with 57 percent of conceptual cues eliciting two conceptual associates.

3.4.3 Time of First Response by Association Type

In the original study, Santos et al. (2011) referred to faster response times, although they only reported the analysis of the median output position of unique responses in detail, not response onset latencies. In a footnote (p.96) they stated that "Analyses performed on the mean time to produce a response showed virtually the same results." As they reported the results of two experiments (the WAT replicated in this chapter and a property generation task using the same words), space constraints may have prohibited presentation of this analysis. Here, I decided it might be informative to test whether the response time analysis could be replicated.

As in the original study, I recorded each WAT on a handheld digital recorder. This is not ideal equipment for the precise measurement of speech offset and onset, so the following analyses should be interpreted with caution. However, any errors in the timing of cue word offset and response onset should be equally distributed across all cue and response types and therefore should not introduce any systematic error or bias into the results. I used WavePad audio editing software (NCH Software, n.d.) to ascertain the voice offset time of each cue word (spoken by me) and the voice onset time of the first response to each cue, for each participant. Thus, response latency was determined as the delay between the cue offset and response onset.

A further caution regarding the response time analysis is that the cue words varied in length, frequency, and phoneme neighbourhood size. Moreover, the deviation point (DP), the point at which a word can be uniquely recognised from all other words in the lexicon, varied greatly. At one extreme were monosyllabic cue words such as “bee,” which have no unique DP. At the other extreme was the cue word “chimpanzee,” for which the DP occurs at chimp~, meaning that participants could initiate response production before hearing the final two syllables. Despite these caveats, if linguistic response times are significantly shorter than conceptual ones, this would provide important evidence in support of LASS theory (Barsalou et al., 2008).

Response latency was calculated for all linguistic, taxonomic, and conceptual responses produced, although due to the small number of root and sound similarity associates these two specific types were combined into form similarity (following Fitzpatrick, 2006, 2007, 2009). First responses that were given after a delay of more than 3 seconds were included in this analysis, as timing is the variable under investigation, so this does not introduce a conflict with the method of the original study (in which responses after a pause were assumed to reflect conceptual processing). In this analysis, the inclusion of delayed first responses cannot confound the assessment of whether the fastest responses are linguistic.

The mean response times and standard deviations for each specific and general response type can be seen in Table 3.7. Mean response times ranged from 0.84 seconds (domain same associates) to 1.56 seconds (form similarity associates). Although very few form similarity associates were produced, making it difficult to interpret this category, the relatively long mean production time suggests that such responses may reflect difficulties in producing a semantically related response. Overall, taxonomic responses were fastest and conceptual responses were slowest. However, it should be noted that within each response type (except for the small number of domain lower associates), the standard deviation is at least half as large as the mean response time, indicating that variation (between participants and/or words) was high across all response types.

Table 3.7

Mean Response Times by Specific and General Response Types

	Specific response type									General type		
	FC	BC	SN	AN	FS	DH	DS	DL	OS	Ling	Tax	Conc
Mean	1.07	1.09	1.16	0.92	1.56	1.07	0.84	1.19	1.20	1.08	0.94	1.20
SD	0.65	0.62	0.78	0.58	0.98	0.75	0.57	0.32	0.68	0.70	0.65	0.68

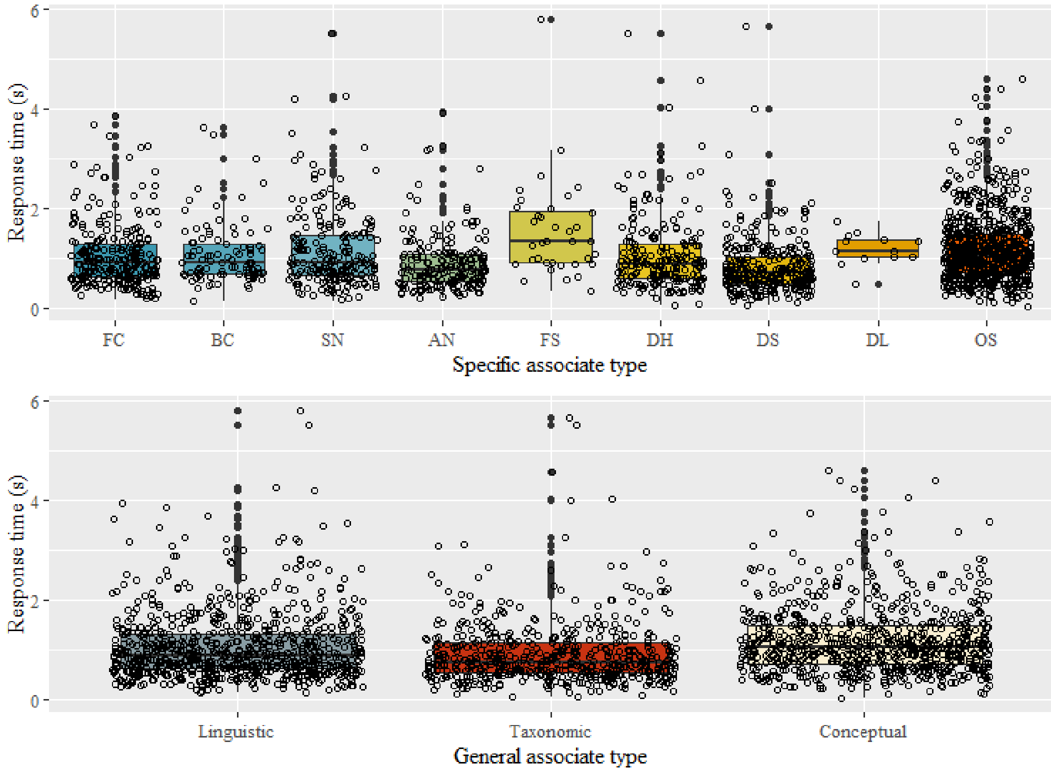
Notes: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, FS = form (root & sound) similarity, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OP = object property, OS = object or situation descriptor, Ling = linguistic, Tax = taxonomic, Conc = conceptual.

As the response times of the general response types were not normally distributed, a non-parametric Kruskal-Wallis test was performed. This showed that the differences in latencies were highly significant (H(2) = 77.349, *p.* < .000001). Post hoc pairwise comparisons were performed using the pgirmess package (Giraudoux, 2022). These tests showed that there was a significant difference in the response latencies of all three general associate types (linguistic-taxonomic: critical difference = 79.48; observed difference = 150.22; linguistic-conceptual: critical difference = 72.92; observed difference = 146.15; taxonomic-conceptual: critical difference = 80.98; observed difference = 296.38).

This analysis shows that, as predicted by LASS theory (Barsalou et al., 2008), conceptual associates were produced more slowly than linguistic ones. However, taxonomic associates were produced fastest, due to the higher response speed in producing category coordinates, which were the most frequent specific type of taxonomic associate. This data is shown in boxplots in Figure 3.4.

Figure 3.4

Response Times for Specific and General Associate Types



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, FS = form (root and sound) similarity, DH = domain higher (superordinate), DS = domain same (coordinate), DL = domain lower (subordinate), OS = object or situation descriptor.

3.5 Discussion

The WAT reported in this chapter partially replicated the results reported by Santos et al. (2011). As in their original experiment, similar proportions of linguistic and conceptual associates were produced, with somewhat fewer taxonomic associates. However, this further refutes the prediction that WA activates predominantly linguistic processing. The finding that linguistic associates tended to be produced earlier than conceptual ones, as predicted, was also replicated. The difference in median output position between linguistic and conceptual associates supports LASS theory (Barsalou et al., 2008). It indicates that this classification may capture two distinct types of association and that processing linguistic relations may be quicker than processing conceptual ones. Yet, in contrast to the original experiment, which found that taxonomic associates were produced earlier than conceptual ones (but later than linguistic associates) no significant difference was found between the median output position of taxonomic and conceptual associates in the replication. Interpreted according to LASS theory, this suggests that taxonomic associations may reflect conceptual processing.

These results were extended by further analyses. First, the cue words were recoded according to the dominant associate given to each cue in the USF norms (Nelson et al., 1998) and the relationship it was judged to represent. This showed that the dominant or primary response to half the cue words in the USF norms was a linguistic associate, making the large number of conceptual responses produced in the replication even more unexpected. If WA behaviour is relatively stable, responses should have been mostly linguistic due to the cue words used, regardless of the prediction of LASS theory that "word association focuses attention on the linguistic system" (Barsalou et al., 2008, p. 256). However, the high number of conceptual associates was due to such associates being more frequently given as second and third responses, which is in line with the prediction of LASS theory that conceptual processing is activated later than linguistic processing. This finding cannot be directly compared to the USF norms, which are composed of one associate per cue.

Although in the experiment reported in this chapter the median output position of linguistic associations was significantly earlier than conceptual ones, replicating Santos et al.'s (2011) finding, the reanalysis suggested a slightly different interpretation. When all responses were analysed, rather than unique response words, participants produced nearly as many conceptual associates as the first response as linguistic ones. The main difference was that conceptual associates were more frequently given as second and third responses than either linguistic or taxonomic associates. First and second responses tended to be of the type predicted by the cue word, indicating a strong influence of the cues on responses. However, linguistic and taxonomic cues also showed a weak tendency to elicit a conceptual associate as the second response. This suggests that rather than linguistic associations being produced earlier than conceptual ones, conceptual responses were more likely to be produced later, compared with linguistic ones. In other words, linguistic processing may not be initiated more quickly than conceptual processing, but it may fade earlier. This is in line with De Deyne et al. (2013), who also found that second and third WA responses tended to be conceptual (properties and thematic associations) more often than predicted. This is quite a different interpretation of later responses on WATs than that of authors such as Nelson et al. (2000), who suggested that second responses are more weakly connected to cue words than first responses.

Analysis of response latency showed that participants were quicker to produce linguistic and taxonomic associates as the first response than they were to produce conceptual ones, in line with LASS theory (Barsalou et al., 2008). However, against prediction, taxonomic associates were produced more quickly than linguistic ones. This result contradicts the analysis of median output position, suggesting that taxonomic associations may be produced via linguistic processing rather than conceptual processing. A possible explanation is that the dominant associates to four of the semantic field cue words, which were classified as taxonomic, could also be viewed as linguistic. The cue words “September” and “Monday” elicited “October” and “Tuesday” as the dominant response, respectively. If these responses were produced based purely on category membership, then all months of the year and days of the week should have been produced in similar proportions. This suggests that these were processed more akin to forward collocation, based on the memorisation of months and days as a list (in 13 of the 15 instances that “October” was followed by a second response, that response was “November”). Likewise, the cues “North” and “winter” elicited the dominant responses “South” and “summer” rather than either of the other two directions or seasons that complete these lexical sets. Thus, such responses seem to reflect antonym production, in addition to taxonomic knowledge.

When these four cue-response pairs were removed from the analysis, the difference in response latency between taxonomic and linguistic associates was not statistically significant (linguistic-taxonomic: critical difference = 79.62; observed difference = 50.05). The shorter latency to produce taxonomic associates would appear to be driven by associates that could also be viewed as linguistic. This may reflect the double link effect on WA response times, as detected by Fitzpatrick and Izura (2011). They found that responses associated to cues by meaning and collocation were produced faster than responses which entered into one type of association with cue words. Overall, it would seem that at least some taxonomic associates can be processed via language, without activating conceptual processing.

The results reported in this chapter provide partial support for LASS theory (Barsalou et al., 2008), in that linguistic responses were more frequently the first response given and such responses were produced more quickly than conceptual ones, as the first response. However, the large number of conceptual associations produced suggest that WA behaviour is not dominated by linguistic processing, as posited by Santos et al. (2011). A possible explanation is that the time frame in which linguistic processing dominates conceptual processing may be too short to be detected in experimental conditions in which participants consciously process language and produce a response. Priming studies, such as those summarised in Paivio (1986), suggest linguistic processing commences prior to conceptual processing, but at a level of subconscious awareness, in the performance of receptive tasks. Productive tasks that require access to semantic meaning, not just word forms, may not be suitable to testing LASS theory at a behavioural level, even if they may be suitable for testing the theory at a neural level (Simmons et al., 2008). This explanation is discussed in further detail in Chapter 8 (Section 8.2).

The replication experiment reported in this chapter revealed three important limitations of the original experiment by Santos et al. (2011). As noted above, the first four categories of cue words were predicted, based on the dominant response in the USF norms (Nelson et al., 1998), to elicit linguistic responses, whereas only one category was predicted to elicit conceptual responses (two after the cues were recoded). The other three cue word categories in the original experiment can be viewed as taxonomic, although only one was labelled as such. One of these, brand cues, was removed in the replication experiment, as explained above (Section 3.2.1). The other category was described as eliciting "a dominant word associate that was a category from the same semantic field" (Santos et al., 2011, p. 94); these cues elicited predominantly taxonomic associates. This feature of the task design means that if WA behaviour conformed to the USF norms (i.e., if associations were consistent across time and participant population), it would *a priori* be expected that half of all associates produced on the WAT reported in Santos et al. and in this chapter should be linguistic. In other words, the imbalanced proportion of cues predicted to elicit a dominant linguistic associate compared to taxonomic and conceptual ones is a potential confounding factor in the design of the original and replication experiment. This concern is heightened by the strong influence of the cue words on responses that has been demonstrated in the analyses reported in this chapter. Further research is needed to test whether the finding that linguistic associates were produced earlier (or were not produced later) was due to the cue words used.

A related issue emerged with the coding system for response (and cue) words. In presenting their results (pp. 97 and 98), Santos et al. (2011) divided the general linguistic response category into six specific categories: the four cue word categories, as would be expected (forward and backward collocates, synonyms and antonyms), plus a minority of responses classified as similar root or sound to the cue word[5]. The general taxonomic category was broken down into three specific response categories, separating superordinate, coordinate, and subordinate responses. This was presumably based on authors such as Murphy (2002), who suggest superordinate relations reflect linguistic knowledge, but coordinate and subordinate relations reflect conceptual knowledge. In contrast, the general conceptual response category was described as "object or situation descriptor" and conflated with the specific response category. Yet, no cue words were chosen which were predicted to elicit situation descriptor responses. In addition, the term *semantic field* was applied to one cue type but was not used to code response types. Unfortunately, Santos et al. offered no explanation for these significant cue word choice decisions or response word type coding changes.

In addition, when considering linguistic, taxonomic, and conceptual associates (i.e., after removing responses coded as other), two responses were produced by participants to only 53% of all cues and a third response was provided to just 18%. Thus, the data available for analysing the types of response given across output positions was substantially reduced when compared with the number of participants and cue words. Moreover, inspection of the data collected in the replication revealed that the number of second and third responses varied considerably by both participant and cue word. It is unknown how this may have biased the results. These three observations suggest that generalisations based on the data collected may not be reliable.

In sum, the replication of Santos et al.'s (2011) WA experiment, reported in this chapter, is important as, to the best of my knowledge, it is the only investigation of WA behaviour that has used a coding system motivated by cognitive science theory (LASS theory; Barsalou et al., 2008). Santos et al.'s coding system, distinguishing between linguistic, taxonomic, and conceptual responses, provides a

[5] The inclusion of these two categories as linguistic responses is unexpected, given that a response word which is connected to the cue word via shared sound or root is typically classified as a form or clang (error) response in word association studies.

very different coding system to the syntagmatic-paradigmatic coding system proposed by Osgood et al. (1954) that has been most frequently used in L2 research since that date (for a review, see Fitzpatrick & Thwaites, 2020; see also Chapter 2, Section 2.2.2). This chapter has established the potential of the new coding system to provide insight into how words in the mental lexicon are processed. As highlighted by De Deyne et al. (2018), to understand the semantic representations underlying lexical knowledge, models need to combine multimodal knowledge from direct experience of the world (and affective reactions to such experiences) with linguistic knowledge. In view of the increasing empirical evidence in support of this assertion (reviewed in Chapter 2, Sections 2.5 and 2.6), coding WA data in terms of linguistic, taxonomic, and conceptual responses warrants further consideration. Yet, the three important limitations that emerged in the replication, regarding cue word selection, coding of responses, and elicitation of multiple responses, require addressing before any reliable conclusions can be drawn regarding the insights the new coding system can provide into WA behaviour. This is the aim of the next experiment, reported in Chapter 4.

Chapter 4. A New Linguistic-Conceptual Coding System for WA Data

The previous chapter reported the replication of the WA experiment conducted by Santos et al. (2011) to test key claims of LASS theory (Barsalou et al., 2008). In the replication, as in the original experiment by Santos et al., participants produced more conceptual than linguistic associates, counter to the prediction that WA would primarily elicit linguistic processing. However, the median output position of unique conceptual associates was, on average, later than linguistic ones, in line with the original study and LASS theory. Due to inconsistencies with the original experiment, further analyses were also conducted, which questioned the interpretation that linguistic associates were produced earlier than conceptual ones. Instead, it seemed that linguistic and conceptual associates were produced in similar proportions as a first response, but linguistic associates were less likely to be produced later. Nevertheless, when the time to produce the first response was analysed, linguistic and taxonomic associates were produced significantly more quickly than conceptual ones. Overall, the results suggest that linguistic and conceptual associations may reflect different processing or semantic information, supporting dual processing theories (e.g., Barsalou et al., 2008; Paivio, 1986). However, important limitations were detected in the replication experiment that require addressing.

Therefore, a new WAT is conducted in this chapter, with two important methodological changes, to provide a more reliable test of the claim that linguistic and conceptual associations are distinct. First, in conducting the replication experiment, it was found that, based on the dominant response in the USF norms (Nelson et al., 1998), there were over twice as many cue words that were predicted to elicit linguistic associates as conceptual ones (27 vs 12). This means that, even if participants were equally likely to access linguistic and conceptual knowledge in producing responses, if the cue word influences responses (a general assumption of the WA paradigm) and WA data is reliable, most responses should have been linguistic associates. This suggests that it would be informative to explore the types of associates that are produced on WATs using a more balanced set of cue words. Second, there was considerable intra- and inter-individual variation in the number of responses given to each cue word. Many participants gave only one associate to most or all of the cues, reducing the data that was available for analysis of multiple responses and introducing a response bias. A task format is needed in which multiple responses can be elicited to all cues from all participants.

Therefore, as detailed below, in this experiment, cue words were chosen that were predicted to elicit a balanced number of linguistic and conceptual associates, according to dominant responses in the USF norms (Nelson et al., 1998). Each participant was requested to provide three associates to each cue word, in line with previous studies that have elicited multiple responses (e.g., De Deyne et al., 2013, 2018; Wolter, 2002). The main hypothesis is that the first response will be more likely to be a linguistic than a conceptual associate, whereas more conceptual than linguistic associates will be given as second and third responses. In addition, exploratory analyses are performed to examine whether there are any patterns in response types across output positions.

4.1 Method

The WAT reported in this chapter is similar to the task conducted in the replication experiment. All materials and the procedure were approved by the Ethics Committee at Swansea University. Two important differences were made in the classification and selection criterion for the cue words. Three changes were also made to the task design to increase the data for analysis, thus improving the power and reliability of the findings. These five changes are listed here and explained in detail below:

1. The cue words are equally divided into two main categories (linguistic and conceptual), each with four subcategories (with 6 cues in each).
2. The cue words are selected to minimise their effects on responses (each cue word was determined to elicit a primary and secondary response of the opposite type in the USF norms).
3. Each participant provided three associates to each cue word.
4. The task was changed to a written format.
5. Each participant responded to all 48 cue words.

4.1.1 *Cue Word Classification*

In the WA experiment by Santos et al. the specific categories of the cue words were divided across three general categories or types. There were four linguistic categories, one taxonomic category, and one conceptual category, plus a semantic field and a brand category, which both seemed to be interpreted as taxonomic in the data analysis (although semantic field was considered a conceptual cue category in the methods). In recoding the cue words (see Chapter 3, Section 3.4.1), this was changed to four linguistic categories, two taxonomic categories, and two conceptual categories, with differing numbers of cue words in each category.

In this experiment the same four specific categories were retained for linguistic associations, namely forward collocates, backward collocates, synonyms, and antonyms. However, these were matched with four specific categories for conceptual associations, which were (stereotypical) object property and an amended semantic field category, plus two new categories, labelled situational descriptor (thematic) and cue property. The cue property category is simply the converse of the object property category (for example, the response “poor” to the cue word “beggar”). *Situational descriptor* is a term used in Santos et al. (2011) and in LASS theory (Barsalou et al., 2008), yet was not used as a cue category, although it was incorporated within the cue property response category. An example of this would be the response “desert” to the cue word “mirage”. Using a balanced selection of cue words, classified into the type of association they are predicted to chiefly elicit, should be more informative about relations between words and concepts in semantic memory (Nelson, 1977).

As noted in the previous chapter, the semantic field category in Santos et al. (2011) consisted of items from a narrow and fixed lexical set, such as a day of the week, the responses to which were predominantly categorised as taxonomic, but which could arguably be considered to be learned as linguistic lists. Therefore, in this experiment the definition of semantic field was based on that given in Vigliocco et al. (2004), namely “a set of lexicalized representations that bear some conceptual similarity” (p. 439). Associations categorised as semantic field were those in which the conceptual similarity was not captured by property or situational descriptor relations.

Given that neither LASS theory (Barsalou et al., 2008) nor Santos et al. (2011) provided a characterisation of taxonomic associations or a prediction regarding such responses, it was decided not to select cue words which elicit a taxonomic associate as the dominant response. This is chiefly due to the lack of consensus over whether such responses reflect linguistic or conceptual processing, or even both types of processing dependent on the type of taxonomic response (Barsalou et al., 2008; Murphy, 2002). In addition, taxonomic associates, by definition, are associations between two nouns that refer to natural entities or human-made objects. Given that word class influences WA responses (Nissen & Henriksen, 2006), a reliable comparison between linguistic, taxonomic, and conceptual cue words would require all cue words to be concrete nouns. Such a design would not allow investigation of the whole mental lexicon. Finally, in order to take the cue word type into account in some analyses, an equal number of cue words are needed in each general and specific category. As there are only three types of taxonomic relation (superordinate, coordinate, and subordinate), this would have meant reducing the specific linguistic and conceptual categories from four to three.

4.1.2 Cue Word Selection Criterion

In the WA experiment conducted by Santos et al. (2011) and replicated in Chapter 3, cue words were chosen based on a dominant response in the USF norms (Nelson et al., 1998). Thus, as discussed in the previous section, all cue words were predicted to elicit responses that entered into a linguistic, taxonomic, or conceptual association with the cue. Although in both Santos et al.'s original experiment and the replication, linguistic cue words also elicited a substantial number of conceptual associates, it could be argued that the results, in part, reflect the choice of cue words. Indeed, this is a key reason for conducting a new WA experiment with an equal number of linguistic and conceptual cues. In addition, it is possible that the observed tendency for participants to produce the same general type of associate as response one and two (when they gave two or more responses) was due to the influence of the cue words. The main prediction of LASS theory (Barsalou et al., 2008) is that linguistic processing is more active more quickly than conceptual processing. If the cue words influenced the dominant associates in the way predicted based on the USF norms, this may have obscured any processing differences due to the type of association.

Therefore, a more rigorous design would be to choose words that are predicted to elicit both linguistic and conceptual associates as dominant responses. To the extent that WA responses are stable, this selection criterion entails that both types of response should be almost equally available to participants to all cues. This has two key advantages: (a) all cues should elicit many responses of both types across all participants, and (b) many cues should elicit both types of response from the same participant (as three responses are required per cue). The first advantage should minimise the influence of the cue words on responses in terms of the general categories of linguistic and conceptual associates. The second advantage enables the assessment of whether participants tend to produce linguistic associates before conceptual ones when they produce both types of response to the same cue, providing a better test of LASS theory (Barsalou et al., 2008).

4.1.3 Task Design

Three changes were made to the procedure used in the replication experiment (Chapter 3). First, three responses were collected for each cue word from each participant so that every cue word and participant were equally represented. One of the main aims of the experiment by Santos et al. (2011) that was replicated in the previous chapter was to test the hypothesis that linguistic associates are produced earlier than conceptual ones. This aim was confounded by the design, which in the replication resulted in many participants producing only one associate to many cue words. This skewed the data in the analysis of response type by output position. In addition to the data loss, it also introduced potential participant and item bias into the results. The most straightforward approach to remedy this issue is to collect a fixed number of associates to each cue word. The number three is slightly arbitrary, but has been used in previous WA research (e.g., De Deyne et al., 2013, 2018; Kim, 2013; Schmitt, 1998a, 1998b; Wolter, 2002). Collecting two associates per cue may yield insufficient data for analysing responses across output positions, whereas collecting four or more associates could exacerbate issues related to chaining of responses, considered next.

The possibility that participants will give chained responses is a concern when collecting multiple associations (i.e., that the second response will be an associate of the first response given, rather than the cue word). However, in a comparison of a study in which participants gave two associates to each cue with an earlier data collection in which only one associate per cue was produced, Nelson et al. (2000) estimated that the potential for chaining was only about 18 per cent, with actual chaining of responses presumably lower than this maximum possible value. This value was lower (8%) when considering cue words that elicited a range of responses with no extremely dominant associate, which was a selection criterion for the cue words used for this experiment. However, as noted by Nelson et al., although this means that the majority of second and third responses were related to the cue word, it does not mean that the second and third responses were unrelated to the previous response(s).

The second design change was partly a result of the decision to collect three responses per cue word. To facilitate the regular production of three responses, the format was changed from oral mode (spoken cues and responses) to a written mode (printed cues and written responses). The effect of delivery and response mode on response words and types is unclear, but previous research suggests that there is an influence. In a small-scale study with L2 learners of English (15 participants and 18 cue words in each format), Suzuki-Parker and Higginbotham (2019) analysed differences in response type, using the main categories created by Fitzpatrick (2006). They found that in the written-written format, compared with an aural-oral format, participants gave more meaning-based responses, with a concomitant decrease in position- and form-based responses as well as erratic responses.

However, the results from an earlier, much larger study, in terms of both participants and cue words, suggest that the effect of delivery and response format on WA behaviour might not be due to the format change, but to a confounding change in the level of task pressure. Entwisle (1966) examined differences between in WA responses given by children aged 11 when the task was conducted in an individual spoken format and a group written format. She found children gave more responses of the same word class as the cue to verb and adjective cue words in the individual spoken format. In a detailed analysis of the adjective cues, she found this was due to children from a higher socio-economic status (ascertained by caregiver income) providing twice as many antonym responses in the individual spoken format. This led her to speculate that response differences between the two formats were caused by unequal power relations between the adult experimenter and respondents, rather than the task. Under pressure, these children provided the response they felt was more correct or desirable. To confirm this hypothesis, she compared responses of female college students under both administrative formats of the WAT and when the interviewer was a peer or a male professor. The students gave almost identical proportions of antonyms to adjectives when interviewed by a peer or performing the task in a group written format. However, they gave a significantly higher proportion of antonyms when responding to the male professor.

This implies that under high pressure conditions, response conformity increases (i.e., some participants try to provide a “correct” response rather than the first response that comes to mind). Importantly, this interpretation provided by Entwisle (1966) could also be applied to Suzuki-Parker and Higginbotham’s (2019) study. Their finding that L2 learners provided more form-based and erratic responses in the spoken format than the written format could indicate pressure effects. Importantly for the experiment reported in this chapter, this implies that the format change should not cause significant differences in responses, given it is being conducted by a student with university student participants. Moreover, this finding suggests that any reduction in pressure resulting from changing from a spoken to a written format will increase the tendency for participants to produce the first responses that come to mind (the intended responses).

Finally, to increase the amount of data available for analysis, each participant was asked to respond to all 48 cue words, rather than dividing the cue words into four sets, as in Santos et al. (2011) and the replication experiment. Santos et al. divided the cue words to minimise any potential effects of previous cue and/or response words on later responses. However, in the property generation task they reported in the same paper, which used the same cue words as in the WAT (reduced from 64 to 60), Santos et al. collected responses to 30 stimulus words from each participant. There is no reason why the influence of previous cue or response words should be diminished in property generation compared with WA. Indeed, participants were given more time to produce properties (15 seconds) than associations (stopped when they paused), thus produced an average of 5.87 properties but 1.74 associates. This suggests any influence should have been greater in the property generation than the WA task.

Analyses conducted by Entwisle (1966) also suggest that such effects are minimal. In her study with L1 English-speaking children, using 96 cue words, a set of 15 semantically related cues were deliberately embedded within the list, the "butterfly" words. On different administrations of the WAT, the order of these words was changed, so that order effects could be investigated. The butterfly words included similar coordinates (e.g., bird, insect, moth), features (e.g., pretty, yellow, wing), and conceptually or semantically related words (cocoon, fly, flower). She found no systematic tendency for participants to give previous cue words as responses to later cues beyond the tendency for these words to be provided (i.e., the number of responses of "insect" to "butterfly" did not differ when the cue "insect" was presented before or after the cue "butterfly"). However, she cautioned against placing related words close to each other on a cue word list and speculated that order effects may be more pronounced with adults. Nevertheless, these results suggest that order effects should not be problematic when using cue words that have not been chosen from semantically related sets. As in the replication experiment, ensuring that no two consecutive cues elicit a dominant associate of the same type and creating two differently ordered lists of cue words, then also reversing each, should be sufficient to control for any influence of previous cue words on later responses.

4.1.4 Cue Word Choice

The new set of cue words selected for this experiment were again taken from the USF norms (Nelson et al., 1998), to maintain consistency and because these norms contain over 5000 cue words and have frequently been used in psycholinguistic research (see Chapter 2, Section 2.1). In order to select words that might elicit both a linguistic and a conceptual response from participants, the norms lists were exhaustively searched for cue words for which the dominant associate was linguistic but the second most dominant associate was conceptual, and vice versa. Strikingly, it was difficult to find six words in each category that satisfied this criterion from the 5019 cue words in the USF norms database. This provides indirect support for the finding in experiment one that most people gave two associates of the same general type—it would appear this could be a property of the cue words, rather than being related to their processing by participants. Note that using the selection criterion automatically excluded cue words for which one of the two dominant responses was a taxonomic response, or a response based on root or sound similarity (although such responses were still coded accordingly in the experiment). The list of potential cue words was further reduced as it became apparent it would be desirable to exclude certain words for a variety of reasons, as explained next.

Firstly, words that elicited an extremely dominant associate that accounted for over 66% of all responses were excluded, as for such cue words it would seem likely that the dominant associate reflects a deeply entrenched association, presumably shared by most L1 speakers of English. Such cue words are unlikely to be very informative about participants' ease of, or preference for, activating linguistic or conceptual processing to produce an associate.

I also chose not to include words that elicited dominant linguistic and conceptual responses to two different meanings of the same form. This meant excluding homographic cue words, such as "lead", /led/ or /li:d/, for which the two dominant responses were "pencil" (object property) and "follow" (antonym), and also homonyms, such as "croak", which elicited the dominant responses of "die" (synonym) and "frog" (object property). Such words were excluded as the different types of response result from responding to different meanings, rather than being different responses to the same item. Unfortunately, this proved insufficient to completely eliminate responses to different meanings of a cue word in all instances, most notably for the cue word "siren", which in the USF norms (Nelson et al., 1998) was responded to almost exclusively as in the meaning alarm, with the dominant associates of "police" (backward collocation) and "loud" (cue word property), but in this experiment unexpectedly elicited a large number of associations to the mythological creature.

I further decided not to include cue words which were a proper noun, such as "Einstein" or "Pepsi", or words which elicited a dominant response that reflected a high cultural load, as UK students responding 20 years later were unlikely to respond to such cues in similar ways as psychology students in Florida in the 1990s. This excluded words that showed geographical load (e.g., "reservation: Indian"), historical load (e.g., "record: tape"), or religious load (e.g., "creation: god"), and antiquated cue words, such as "nightgown" or "galoshes".

In addition, cues and associates that were verb: noun or noun: verb combinations were also excluded (e.g., "scratch: itch") as these are common discourse collocates that also evoke an image of the situation they describe, so it would be problematic to judge that such an associate was produced linguistically or conceptually. This problem in classifying responses was discussed in the previous chapter and is revisited in the discussion in this chapter and Chapter 8.

The list of potential cue words resulting from this selection process was then given to a second researcher[6] who was asked to comment on the various choices and discuss any deemed to be problematic. This feedback was used in deciding on the final list of 48 cue words used in the experiment. This list, the two dominant associates in the USF norms, and their classification, is given in Appendix 6.

4.1.5 Participants

I recruited participants from seminars and lectures or in quiet campus café areas and asked if they were willing to take part in a short language experiment. In total, 54 participants completed the WAT, all of whom were L1 English-speaking students at Swansea University; 40 were aged under 25, and 10 were aged over 25 (four participants did not complete this information). All participants provided informed written consent before proceeding with the WAT and were given a debrief about the experiment afterwards, which included my email address. No participant received any financial compensation for taking part in the experiment.

4.1.6 Procedure

The oral instructions given to each participant were to write three responses to each cue word as quickly as possible, and it was stressed that there were no correct answers, they should just write the first words that came to mind, spontaneously. The cue words were presented in written format, in a table over two sides of A4 paper, with the cue words written in the first column and three blank columns for responses. The cue words were semi-randomly ordered in two different presentation orders, with each list also being reversed. No two words from the same response category nor two words that were semantically related (such as “heart” and “stomach”) were adjacent on the lists. Brief instructions were given at the top on both sides, telling participants, “for each word write the first three words you think of, as naturally and spontaneously as you can.” These instructions were in line with those given orally in the replication experiment (see Chapter 3). At the end of the task participants were

[6] A fellow PhD candidate conducting research in a similar area, who was acquainted with Santos et al.’s (2011) paper and LASS theory (Barsalou et al., 2008).

asked to circle whether they were under or over 25 and monolingual or bilingual. No time limit was set to complete the word association task.

4.1.7 Exclusions

Before analysis, one cue word, *inertia*, was excluded as over half the participants did not know the word at all and made no attempt to respond to it, or their responses showed they had misunderstood its meaning. In addition, data from one participant was removed prior to analysis, as 25% of associates were idiosyncratic and a further 15% were either only very loosely associated (e.g., “bicycle: man”), associated via root or sound similarity, or chained associates given to the previous response rather than the cue word.

4.1.8 Data Analysis

Analyses were conducted in RStudio (RStudio Team, 2020), run on R (R Core Team, 2021), using the tidyverse (Wickham et al. (2019) and vcd packages (Meyer et al., 2006; 2022). Figures were created using ggplot2 (Wickham, 2016) and the wesanderson package (Ram & Wickham, 2018), following Healy (2019).

4.2 Results

In line with the replication experiment (Chapter 3), first the number of responses of each specific and general association type was calculated as a function of the cue word category type. This was done for the first response only and for all three responses. Next, the data were explored to examine whether there were any systematic differences in linguistic and conceptual associates across output positions. On analysing the data, a substantial proportion of responses were evaluative properties of the cue words, such as the responses “yum” and “happy” to the cue “ice-cream”. Although such responses are a type of cue property, they are not always informative about the meaning of the cue word, so were coded separately. In addition, an *other* category was again used. This included idiosyncratic associates, associates with a similar form to the cue word, chained responses, and no responses. Such responses were evenly spread across linguistic and conceptual cue words.

4.2.1 *Distribution of Response Types*

First, the number of responses of each specific association type were calculated per cue word type. As all participants provided three responses to each cue word, the number of responses is shown for the first response only in Table 4.1 and for all three responses in Table 4.2. Both tables show that participants produced more conceptual than linguistic associates to all cue words. However, each specific type of associate was given more frequently to the cue word predicted to elicit that type of associate than to any other cue type. This indicates responses were somewhat consistent with the USF norms (Nelson et al., 1998).

Table 4.1

Cue Word Response Categories for Response One

Response category	Cue word category								Total	
	Linguistic				Conceptual					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
FC	**92**	6	18	25	12	21	44	27	245	0.098
BC	18	**97**	3	3	7	18	0	0	146	0.059
SN	14	24	**56**	23	43	29	37	23	249	0.100
AN	0	1	1	**31**	3	0	4	4	44	0.018
OP	0	2	43	48	**142**	2	0	2	239	0.096
CP	45	42	14	74	11	**79**	28	3	296	0.119
EP	2	21	25	18	4	**21**	14	3	108	0.043
SD	99	100	114	68	31	112	**137**	26	687	0.276
SF	5	6	29	10	39	19	14	**162**	284	0.114
TX	28	5	0	3	7	3	29	0	75	0.030
OR	15	14	15	15	19	14	11	15	118	0.047
Linguistic	124	128	78	82	65	68	85	54	684	0.275
Conceptual	151	171	225	218	227	233	193	196	1614	0.648

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = object property, CP = cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = Other response, No. = number, Prop. = proportion. The number of responses that are of the same category type as the cue word are in bold. There were 5 semantic field cues, compared to six of the other types, meaning this cue category has 53 fewer responses.

Considering first responses only, conceptual associates dominated responses, accounting for 65% compared with just 28% linguistic associates. Situational descriptors were the dominant type of association to most cue types, with this one specific response type accounting for 28% of all responses. Only object property and semantic field cues elicited more responses of the predicted type. In addition, antonym cues elicited more cue property associates. The remaining seven target response types all accounted for approximately 10% of responses, except for backward collocates and antonyms, which represented just under 6% and 2% of responses, respectively. More in line with the prediction that cue words would elicit both types of response, forward collocation and synonym associates were relatively frequently given in response to all four types of conceptual cues. Conversely, object properties tended to be produced to synonym and antonym cues and cue properties to collocation cues.

Across all three responses to each cue, as shown in Table 4.2, conceptual associates were slightly more dominant, at 68% of responses, whereas the proportion of linguistic associates decreased to 21%, partly also due to a slight increase of other responses. The proportion of situational descriptors was also slightly higher, at 32%. Similar to first responses, all cue types elicited mostly situational descriptor associates, except object property and semantic field cues, which predominantly elicited associates of those types. In general, the proportion of responses of each specific conceptual type remained similar across all three responses as in response one, whereas the proportion of each specific linguistic type decreased slightly.

Table 4.2

Cue Word Response Categories Across all Three Responses

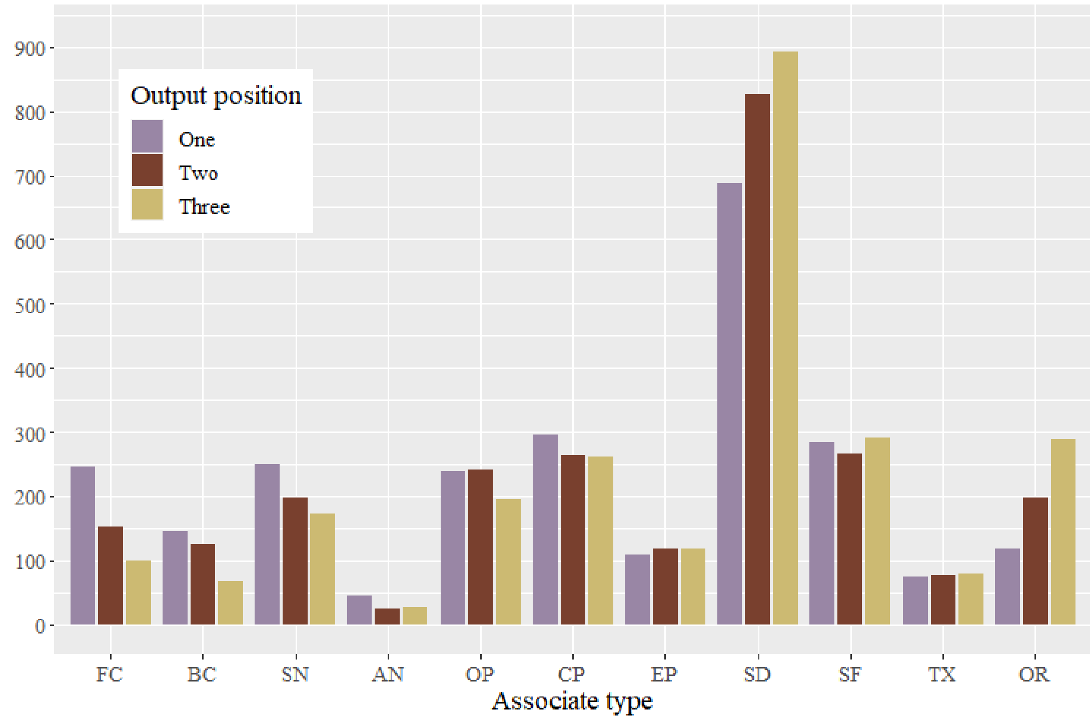
Response category	Cue word category								Total	
	Linguistic				Conceptual					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
FC	**172**	13	33	70	24	38	96	51	497	0.067
BC	39	**202**	10	4	15	64	3	0	337	0.045
SN	36	68	**146**	61	89	61	87	73	621	0.083
AN	2	1	2	**59**	5	1	13	12	95	0.013
OP	3	7	110	121	**427**	3	2	4	677	0.091
CP	143	135	57	183	36	**177**	78	13	822	0.11
EP	9	62	86	50	7	**61**	49	18	342	0.046
SD	369	359	325	285	132	404	**409**	121	2404	0.322
SF	41	22	82	48	106	60	51	**433**	843	0.113
TX	77	13	0	11	14	10	105	2	232	0.031
OR	63	72	103	62	99	75	61	68	603	0.081
Linguistic	249	284	191	194	133	164	199	136	1550	0.208
Conceptual	565	585	660	687	708	705	589	589	5088	0.682

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = object property, CP = cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = Other response, No. = number, Prop. = proportion. The number of responses that are of the same category type as the cue word are in bold. There were 5 semantic field cues, compared to six of the other types, meaning this cue category has 159 fewer responses.

To better understand this data, the number of responses of each specific associate type by output position (first, second, or third response) is shown in Figure 4.1. This visualisation of the data was possible in this experiment as the task was modified to elicit three words to each cue from each participant, rather than as many associates as possible without pausing, so the number of responses in each position is constant (117 instances of no response are included within other responses).

Figure 4.1

Associate Type by Output Response Position



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = object property, CP = cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response.

The most striking finding is the far greater number of situational descriptor associates produced across all three output positions than any other type of associate. However, there was a decrease in all four types of linguistic associate across output position, particularly forward collocations. In contrast, the number of most types of conceptual associate (and the few taxonomic associates given) was relatively stable across output position, except for situational descriptors, which increased. These results suggest that rather than producing linguistic associates earlier, as predicted by LASS theory (Barsalou et al., 2008), a more accurate interpretation of the data is that linguistic associates were less likely to be produced at later output positions. In other words, participants produced substantially fewer linguistic associates and these were more likely to be produced early or not at all.

4.2.2 Relationships Between Multiple Responses

Due to the predominance of conceptual associates, the strongest relationship between multiple responses is evidently the tendency to produce two or three conceptual associates. However, comparison of the association types produced as response one and two by each participant allows the examination of whether a linguistic associate is more likely to precede or follow a conceptual one, providing an assessment of the key claim of LASS theory (Barsalou et al., 2008) that linguistic associates tend to be produced earlier. For this analysis, all response pairs that included a taxonomic associate or an associate categorised as other were excluded. Descriptive statistics for each of the four possible response patterns of linguistic and conceptual associates across first and second responses is shown in Table 4.3, based on each individual participant. The important comparison is whether participants produced a linguistic followed by a conceptual associate significantly more frequently than they produced a conceptual followed by a linguistic associate. A paired samples t-test showed this trend was highly significant, with a moderate effect size (t = 4.813, df = 52, p.<.001, Cohen's d = 0.66). This suggests that, despite the overall trend to produce conceptual associations, when linguistic associates were produced, they were produced earlier than conceptual ones.

Table 4.3

Descriptive statistics for Response One and Two

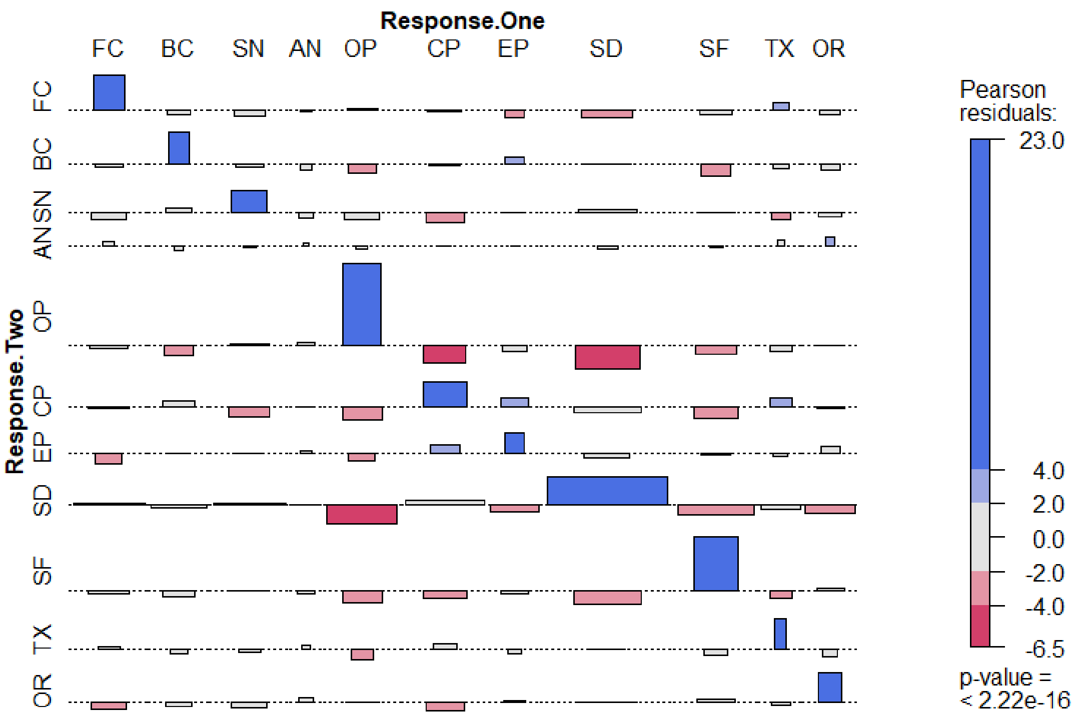
Response one	Response two	Mean	SD
Linguistic	Linguistic	3.87	2.24
	Conceptual	7.96	3.07
Conceptual	Linguistic	5.00	2.49
	Conceptual	22.38	4.47

To further understand this association between the first and second responses given, a chi-square test was performed to assess whether the specific type of association to the cue word evinced by response one was associated with response two type. This revealed a highly significant association ($\chi^2(100) = 1704$, $p. < 0.00001$), which was explored in more detail via an association plot, as shown in

Figure 4.2. It is notable that every individual category of first response except antonyms (which were elicited in very small numbers) was followed by the same type of response as the second response at a level that was statistically significant. Although the majority of responses were two situational descriptors (as indicated by the width of the bar), this association was particularly specific for object property associates (as indicated by the height of the bar). There was also a statistically significant association between cue and evaluative property associates, suggesting these two types of association are processed as similar constructs. The strongest dissociation was between situational descriptor and cue property associates.

Figure 4.2

Association Between Response One and Two



Note. Blue bars indicate higher numbers than expected and pink bars indicate lower numbers; deep hues indicate a significant difference. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = object property, CP = cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response.

Of note, the influence of the cue word type on the type of association produced as the first or second response was not as restrictive (see Appendix 7). This indicates that the type of associate participants produced as their first response influenced the next response type. Moreover, this influence of the first response was greater than that of the cue word. This suggests that once an association is given to a cue word, participants may *self-prime,* producing a second response of the same type.

4.3 Discussion

The main finding in the experiment reported in this chapter was that participants produced a much greater proportion of conceptual associates than linguistic ones, across all three output positions. Nearly 70 per cent of all responses were conceptual, compared to only 20 per cent linguistic responses. In addition, there was a much greater proportion of situational descriptor associates than predicted, and this was the only category within which the number of associates increased steadily and to a large extent, from first to third output position.

These results were unexpected, given that the cue words were predicted to elicit approximately equal proportions of linguistic and conceptual responses, based on the dominant responses given in the USF norms (Nelson et al., 1998; see Appendix 6). They refute the hypothesis that WATs predominantly activate shallower linguistic processing (Santos et al., 2011) because they are a language task (Clark, 1970; Osgood et al., 1954; Wettler & Rapp, 1993). Instead, participants tended to access deeper conceptual processing to respond to cue words, especially knowledge of the contexts in which the word or its referent typically occurs.

Although linguistic associates decreased over later output positions, the results did not produce a clear picture of linguistic responses preceding conceptual ones, as predicted by LASS theory (Barsalou et al., 2008) and claimed in Santos et al. (2011). Instead, they suggest that participants were not only more likely to produce conceptual associates, but that once this knowledge was activated, they tended to keep using this system to respond. In contrast, when participants accessed linguistic processing to produce the first response, they easily switched to conceptual processing to produce later responses. These response patterns support the idea that conceptual processing relies on situated simulation, implying higher costs in switching from richer, more detailed conceptual representations to shallower linguistic processing than in switching in the opposite direction. Thus, the already dominant conceptual processing became even more dominant as more associates were produced to each cue (see also De Deyne et al., 2013).

In addition, the results suggest that the first response to each cue primes further responses, not only in general terms, but according to specific type. This indicates that the categories used to code the association between the cue and response words may be psychologically real. Moreover, that such priming effects were slightly stronger for conceptual responses than linguistic ones suggests that connections grounded in conceptual representations may be somewhat stronger than connections grounded in linguistic representations. This further supports the claim that conceptual associations are based on situated simulation and thus reflect deeper processing than linguistic associations (Barsalou et al., 2008; Santos et al., 2011).

Overall, the results presented in this chapter support the distinction between linguistic and conceptual associations, suggesting that they may reflect different types of processing or knowledge. However, this claim is tentative due to the much larger proportion of conceptual than linguistic associates produced to the cue words selected. The tendency not to produce a linguistic associate after a conceptual one could just reflect the low number of linguistic associates produced. A WAT that elicited similar proportions of linguistic and conceptual associates would provide a more robust test of this trend. In addition, the results raise the possibility that WATs may be more informative about conceptual than linguistic knowledge. If this is the case, WA may not be an appropriate paradigm for exploring the mental lexicon of L2 learners, given that conceptual associates may not be very informative regarding lexical knowledge (Meara, 1983). For both these reasons, a new WAT is conducted in the next chapter, with the aim of eliciting a more balanced proportion of linguistic and conceptual associates.

Chapter 5. The Linguistic-Conceptual Coding System Revisited

This chapter presents a new iteration of the WAT reported in Chapter 4, in which the coding system remains the same but new cue words are used, based on different selection criteria (explained below). This stems from a theoretical and practical motivation: To attempt to select cues that elicit a balanced proportion of linguistic and conceptual associates and that can be used with English language learners of moderate proficiency level.

5.1 Background and Aims

The experiment reported in Chapter 4 incorporated two substantial improvements on the WAT conducted by Santos et al. (2011) and replicated in Chapter 3: the use of a balanced set of cue words predicted to elicit language and situated simulation, in line with LASS theory (Barsalou et al., 2008; see Chapter 2, Section 2.7); and the elicitation of three responses to each cue word to enable more robust analyses. The main finding was that conceptual associates were more frequent than linguistic ones across all three output positions. In addition, participants were significantly more likely to produce a conceptual associate after a linguistic one than they were to produce a linguistic associate after a conceptual one. The eight specific categories used within the general linguistic and conceptual categories also showed evidence of self-priming, in that there was a tendency for the second response to enter into the same type of association to the cue word as the first response.

These findings suggest that the distinction between linguistic and conceptual associates could be psychologically valid. If so, coding word associations as linguistic or conceptual associates may provide novel insights into the lexical processing and/or storage underlying the performance of such tasks. However, these findings should be interpreted with caution due to an unanticipated effect of the design of the previous experiment. In order to minimise the influence of the cue words, words were selected that elicited dominant linguistic and conceptual associates in the USF norms (Nelson et al., 1998). The primary response to each cue word was identified as holding the opposite type of association as the second most frequently given response (i.e., conceptual then linguistic or linguistic then conceptual). Rather than minimising the influence of the cue, such that similar proportions of linguistic and conceptual associates were produced, this led to a much higher proportion of conceptual than linguistic associates. The extreme imbalance between the proportion of linguistic and conceptual associates may have distorted the analyses, as linguistic associates may only have tended not to be produced after conceptual ones because they tended not to be produced in general. It also meant the high number of conceptual associates created almost a ceiling effect, overwhelming the reliable detection of any systemic pattern beyond the tendency for associates to be conceptual. This issue was compounded by the high proportion of situational descriptors within the general conceptual associate category, in comparison with the other three specific types of conceptual associate (object and cue properties and semantic field). Moreover, this finding further contradicts the assumption of Santos et al. (2011), based partly on LASS theory (Barsalou et al., 2008), that WATs are primarily shallow language tasks, which tend not to recruit deeper conceptual processing (situated simulations).

Collecting a more equal number of linguistic and conceptual associates will enable the research question posed in the previous experiment to be answered more convincingly: Does coding WA data as linguistic or conceptual lead to findings that support the key claim of LASS Theory (Barsalou et al., 2008), that words are processed by accessing (shallow, symbolic) linguistic or (grounded, multimodal) conceptual knowledge? In other words, does the production of linguistic and conceptual associates reveal systematic patterns which suggest that they reflect access to different types of knowledge? This is the main research question addressed in this chapter. To do so, as explained in the methods, cue words were chosen for which, as far as possible, the first three dominant responses in the USF norms (Nelson et al., 1998) were all the same general type (i.e., all linguistic or all conceptual). Selecting cue words that have previously shown a tendency to constrain the type of associate produced in a WAT (according to the distinction between linguistic and conceptual associates employed throughout this thesis) is predicted to result in balanced numbers of each associate type. If the two types of associates reflect different processing systems, the distribution of both types across multiple responses should not be uniform. Note that this also creates a more stringent test, as the cue words selected for this experiment are anticipated to strongly influence the type of associate produced. If the first associate produced to a cue word by participants tends to influence the subsequent associates, and this effect is not determined by the cue word, such self-priming would provide much stronger support for the hypothesis that it reflects an underlying difference between processing linguistic and conceptual associations.

If the cue words used in the experiment reported next elicit responses that show systematic differences between linguistic and conceptual associates, the subsequent test of the utility of this new coding system, developed from that presented in Santos et al. (2011), should be to compare it with established coding systems that have been used in WA research (see Chapter 2, Sections 2.2 and 2.3). It is highly plausible that any patterns revealed by the linguistic-conceptual distinction emerge with other coding systems, but that previous studies haven't analysed and reported multiple responses in this way. To foreshadow the results reported in this chapter, this comparison of coding systems is the topic of the following chapter.

Further, if different coding systems reveal different patterns in WA behaviour across multiple responses, a logical next step would be to compare them in an area that has been a focus of much WA research (see Chapter 2, Section 2.3), namely, to explore associations made by language learners. In anticipation of such a step, high frequency words that are more likely to be familiar to learners were selected as the cue words for the experiment reported in this chapter. Although this decision places an extra constraint on the cue words, it enables WAs to be collected that could subsequently be compared with those produced by learners.

In summary, the methodological aim of the experiment reported in this chapter is to collect a more balanced proportion of linguistic and conceptual associates to a new set of cue words, all of which are high frequency words, retaining the requirement for three associates per cue. The research question is the same as in the previous chapter: Are there systematic differences between linguistic and conceptual associates that suggest word meanings are processed via these two different semantic systems, rather than one semantic system?

5.2 Method

The WAT reported in this chapter is similar to that reported in Chapter 4, in terms of both materials and procedure. However, in order to achieve the aims stated above, the selection criteria for the cue words were modified, as described next. All materials and the procedure were approved by the Ethics Committee at Swansea University.

5.2.1 *Cue Word Selection Criteria*

Cue words for the word association task reported here were again chosen from the USF norms (Nelson et al., 1998), according to the dominant associates given. In the previous experiment, 48 cue words were selected, half of which were determined to elicit a dominant linguistic associate and half a conceptual associate. The cue-response association was defined as linguistic if the association was that of forward collocation, backward collocation, synonymy, or antonymy. Conceptual associates were related to the cue such that the cue was a property of the associate (object property), the associate was a property of the cue (cue property), the associate was a situational descriptor of the cue (thematic associate), or by semantic connection (semantic field). These categories were a modification of those reported in Santos et al. (2011; see Chapter 3). A total of six words were chosen for each of the eight specific coding categories, with the added criteria that the second most frequently given associate was of the opposite type (i.e., it was a conceptual associate when the dominant associate was one of the four linguistic types, and vice versa). In this experiment, the cue words were divided into the same four linguistic and four conceptual categories, and the following selection criteria were retained (see Chapter 4, section 4.1.2 for more details):

- An equal number of cue words were chosen for each specific linguistic and conceptual category.
- Cue words that elicited a single highly dominant associate were avoided.
- Cue words whose dominant associate formed a verb + noun pair were avoided, due to the ambiguity as to whether such associations are collocates (linguistic) or situational descriptors (conceptual).
- Homographic cue words to which the dominant associates reflected two different meanings were avoided.
- Cue words for which the dominant associate displayed specific cultural load were avoided.
- Cue words that were proper nouns were avoided.

However, the following changes were made to the selection criteria, as explained in more detail below:

- The dominant associates to the cue words in each specific linguistic category were all linguistic.
- The dominant associates to the cue words in each specific conceptual category were all conceptual.
- All cue words were headwords in the 1000 or 2000 most frequent word families (Nation, 2017).

For this experiment, 80 cue words (10 in each category) were chosen such that the three or four most frequently elicited responses were all linguistic or all conceptual associates. This means that the cue words are anticipated to constrain many of the response words to conform to the predicted general category, if not the specific category. To the extent that similar associates are produced to those found in the USF norms (Nelson et al., 1998), participants should produce a balanced proportion of linguistic and conceptual associates. This will provide a large dataset to test between two competing claims: (a) linguistic associates are produced earlier than conceptual ones, as suggested by Santos et al. (2011), or (b) linguistic associates are less likely to be produced after a conceptual associate has been provided to a cue word, as suggested in the previous experiment reported in Chapter 4. In addition, the use of cue words that are anticipated to elicit predominantly linguistic or predominantly conceptual associates, rather than eliciting both type of associates, allows these potential outcomes to be compared to a null hypothesis, (c) the type of associate produced as the first, second, and third response is a function of the cue word, such that cue word selection is the most important factor in adult L1 speaker word association.

Given word class has been shown to influence the proportion of paradigmatic and syntagmatic associates given (Nissen and Henriksen, 2006), as far as possible, nouns, verbs, and adjectives were included in each specific category. For example, it would have been easier to select cue words that would elicit the predicted type of response if only adjectives and verbs were chosen for the synonym and antonym categories and nouns in the conceptual associate categories, but care was taken to avoid creating this confound.

Due to the added restriction that all cue words were headwords from the 2000 most frequent word families in English (see below), this division between linguistic and conceptual cue words was difficult to meet with conceptual cues. Consequently, seven of the 40 cue words that were chosen as they elicited a dominant conceptual associate in the USF norms (Nelson et al., 1998), elicited a linguistic associate as the second most frequent response. Four of these cue words were within the semantic field relation category, which proved problematic for further reasons, as discussed later in this chapter. However, as the number of conceptual associates exceeded linguistic ones in both the replication experiment (Chapter 3) and the previous experiment (Chapter 4), this issue is not expected to result in more linguistic than conceptual associates overall.

To select cue words that could be used with English language learners, the 1000 and 2000 highest frequency English word family lists created by Nation (2017) were used. These lists are composed of headwords, which are the base form of the word, without any prefixes or suffixes (i.e., the headword "believe" also represents such forms or tokens as believed and unbelievable). They were made using the British National Corpus (BNC; Davies, 2004) and Corpus of Contemporary American English (COCA; Davies, 2008-). The two highest frequency word lists were created using a subset of these two corpora (lower frequency words were selected from the whole of both corpora). To compile these first two lists, Nation produced a ten-million-word-token corpus, of which six million words were spoken British and American English. He also included sources of English language aimed at children, such as school journals and movies. These adaptations were made so that the word families in the first two frequency bands would more closely resemble everyday language use, rather than be distorted by words that are high frequency in formal written texts, thus providing a useful pedagogical tool for teaching English language learners (for more information, see Nation, 2016).

These design choices make these lists appropriate for selecting cue words that English language learners would be likely to know. In addition, psycholinguistic research into priming effects has shown that semantic models constructed from a relatively small corpus based on American film and television subtitles (SUBTLEX-US; Brysbaert & New, 2009) correlate better with human performance than a much larger corpus based on written texts (Brysbaert et al., 2011; Mandera et al., 2017), further supporting Nation's design choices for the first 2000 word families.

At the first cue word selection attempt, only the 1000 word list was used. All words in this list were entered into a spreadsheet, then all associates given to each word by at least five participants in the USF norms (Nelson et al., 1998) were entered alongside (109 words in the 1000 word family list were not used as cue words in the USF norms). Potential cue words were then selected based on the type of association with the most highly dominant response words (using the four linguistic and four conceptual categories to identify the association). As noted above, the selection criteria were difficult to meet in trying to obtain 80 cue words from this restricted set of words, so the 2000 word list was also consulted and 19 cue words were chosen from this frequency band (six linguistic and 13 conceptual cues). The specific cue word category with the highest number of lower frequency words was object property (conceptual), in which six of the ten cue words appeared in the 2000 word band.

As a result of this compromise, two further word lists were consulted: The New General Service List (NGSL; Browne et al., 2013) and the 1000 and 2000 word lists from the JACET List of 8000 Basic Words (committee of revising JACET basic words, 2003). The two highest frequency word lists from the JACET 8000 were used as I expected to administer a subset of the cue words as a WAT to Japanese learners of English as a foreign language (see Chapter 8). Of the 80 cue words selected, 77 appeared on both these lists, with 60 words present on the JACET 1000 word list. The only exceptions were "elephant", which is not on the NGSL, "belt", which is on the 3000 word list of the JACET 8000, and "toilet", which is also on the 3000 word list of the JACET 8000 and is not on the NGSL.

These additional word lists were subsequently referred to whenever a potential cue word was identified that was a derivative of a headword on Nation's (2017) word family lists. Only six of the 80 cue words were not headwords, and all of these words are included on the JACET 1000 or 2000 word lists in the form used in this experiment. The six words were the noun forms of two verb headwords, "government" and "education", which both appear on the 1000 word list of the JACET 8000 and on the NGSL; the cue words "interesting", which appears on the 1000 word list of the JACET 8000 (only the headword "interest" is on the NGSL), and "healthy", which appears on the 2000 word list of the JACET 8000 (the headword "health" is on the 1000 word list) and on the NGSL; and the cue words "weekend" and "tourist", which appear on the 2000 word list of the JACET 8000 (the headword "week" is on the 1000 word list and "tour" is also on the 2000 word list). By consulting three different lists of high frequency words that were all created to aid in the teaching of English as a foreign language, all the cue words selected for this experiment have a high probability of being among the earliest words learned in this context. In addition, the WAT administered in this experiment is viewed as a task to assess lexical processing, not as a test of word knowledge.

5.2.2 Materials

The 80 cue words were randomly divided into two lists, with five cue words from each predicted response category on each list, so that each participant provided associates to half the cue words (see Appendix 8). This was done to prevent fatigue and keep the time required by most participants to complete the task within 10 minutes. In the previous experiment (Chapter 5), which had 48 cue words, some participants took longer than 10 minutes and this seemed to approach the limit of their willingness to continue. Therefore, I decided that greatly expanding the length of the WAT to 80 cue words could reduce the reliability of the data collected. After randomly creating the two lists, I checked to ensure any two cue words with similar meanings or whose referents typically co-occur appeared on different lists. For example, the cue words "fire" and "burn" and "station" and "ticket" were separated. The 40 cue words on each list were semi-randomly ordered, so that no two words from the same predicted response category occurred next to each other. Each list was then reverse-ordered to reduce any priming effects from previous cues, creating four different word lists. The lists were printed in two tables on one double-sided sheet of A4 paper, with three spaces provided next to each cue word for participants to write their response words. The task was preceded by short and simple written instructions and at the end of the task, participants were asked to circle their gender, age, and first language(s).

5.2.3 Participants

After receiving ethical clearance from Swansea University Ethics Committee, I administered the WAT to students at Swansea University. Participation was voluntary and was not remunerated. Participants read and signed a consent form first and then proceeded to complete the task, after which they were given a short debrief about the aims of the experiment. In total 101 L1 English speakers completed the WAT, of whom 68 were female and 33 were male and 90 were aged under 25. Most participants (87) spoke only English as a native language, 11 were bilingual in English and Welsh, and three participants were bilingual in English and another language.

5.2.4 Procedure

Participants were recruited from seminars and lectures or approached in a campus café area and asked if they were willing to take part in a short language experiment. All participants were asked to give informed written consent before proceeding. The instructions given to each participant were to write three responses to each cue word as quickly as possible, and it was stressed that there were no correct answers, they should just write the first words that came to mind, spontaneously. However, no time limit was set to complete the word association task. After completing the WAT, participants were given a written debrief about the purpose of the experiment, which included my email address.

5.2.5 Coding of Responses

As in both the previous experiments, response words were first entered into a spreadsheet and then each response was coded according to association type. The main association types of interest were the four linguistic and four conceptual ones, as these are the two semantic systems posited in LASS Theory (Barsalou et al., 2008). The four linguistic codes used were forward collocation, backward collocation, synonym, and antonym and the four conceptual codes were stereotypical object property, stereotypical cue property, situational descriptor, and semantic field. As in the previous two experiments, participants also produced some taxonomic associates, in which the response was a superordinate, coordinate, or subordinate of the cue word. Cue words were not chosen that were predicted to elicit taxonomic relations (as discussed in Chapter 4) because Santos et al. (2011) were uncertain whether such associations were produced via linguistic or conceptual processing.

In addition, a small proportion of response words (7.3%) were coded as other (erratic or clang) associations, largely in line with the other category used by Nissen and Henriksen (2006). This category included no or illegible responses and chained responses (responses that seemed to be associated to the previously given associate rather than the cue word, such as the associate “ability” given after the associate “mobility” to the cue word “body”; or which were derived from the previous associate, such as the associate “taking” after the associate “take”). Such responses accounted for 2.6% of all responses. It also included response words for which the linguistic association did not form a semantic link to the cue word but was a word of the same root as the cue (the addition of a standard affix, such as “helpful” in response to “help”, or a change of verb tense, such as “thought” in response to “think”), or a translation of the cue word (into Welsh or French), which composed just 1.7% of all responses. The other 2.9% of responses in this category were conceptual associates that did not provide a meaningful context for the cue. These were conceptual associations that were idiosyncratic and so were assumed to reflect personal semantics (for a review, see Renoult et al., 2012), meaning the association was unclear (such as the associate “protection” in response to the cue word “bag”) or reflected personal experience (such as the response “brother” given to the cue “soldier”); as well as associations which held only the most generalised relation[7] to the cue and so were judged to be non-informative (such as the response “people” to the cue “boat” or prepositions, pronouns, or articles which did not alter the meaning of the cue word in any way or form a meaningful phrasal collocate, such as the response “it” to the cue “measure”). However, unlike in Nissen and Henriksen (2006), phonological associates were included in the other category (e.g., “habit-rabbit”) but evaluative properties were not, instead they were analysed incorporated within cue properties and as a distinct category.

Some further coding decisions that should be noted are that responses coded as synonym associates included some words that were synonyms within a restricted sense (such as the response “addiction” to the cue word “habit”) or a broad sense, such as “big” to “fat”) and some collocations were completions of a compound word (most notably the dominant response “hand” to the cue word “bag”[8]). In addition, examples or types of an entity referred to by the cue word (such as “wedding” in response to the cue “event”) and brand names (such as “Apple” in response to “computer”) were coded as taxonomic associations. A few associates reflecting

[7] A larger set of non-specific and thus uninformative associates were excluded by Kent & Rosanoff (1920).

[8] In the USF norms, the dominant responses to “bag” were “lunch” and “paper”.

popular media were coded in two different ways: When such associates completed a name, they were coded as backward or forward collocations (such as "mirror-black" or "chance-rapper"). All other instances were coded as situational descriptors (such as the associate "Tinkerbell" to the cue word "wish"). In addition, as in the previous two experiments, many responses that were coded as cue word property associates were evaluative rather than stereotypical properties (such as "amazing" in response to "elephant"). Given the large number of evaluative properties, this type of response is discussed in more detail below.

5.2.6 Data Analysis

The word association data was analysed in RStudio (RStudio Team, 2020), run on R (R Core Team, 2021). Coding was based on various sources, notably Levshina (2015). Tables and analyses were produced using the tidyverse (Wickham et al. (2019) and vcd packages (Meyer et al., 2006; 2022) and figures using ggplot2 (Wickham, 2016), following Healy (2019). The colour scheme for some of the graphs was taken from the wesanderson package (Ram & Wickham, 2018).

Although some inferential statistics are reported, most analyses are descriptive, for two main reasons. First, as in the previous two experiments, the unit of analysis is the word, which means all analyses are based on thousands of data points and so even very minor differences will be statistically significant, merely reflecting this large sample size and not necessarily indicating an appreciable effect. Second, and more importantly, the questions being asked of the data do not conform to standard analyses, as it is the nested nature of the data that is being explored. The main question is not how participants' responses are influenced by the cue words, but how their first responses influence subsequent ones, independent of the cue word. In other words, rather than determining if there is a systematic effect of the stimuli on responses, the main question is whether there is a systematic effect of response one on responses two and three. In an important sense, the stimulus becomes the main confounding variable, rather than the independent or predictor variable.

5.3 Results

A total of 12,120 word associations were collected for the 80 cue words, 4040 in each output position (responses one, two, and three). To better understand this data, summary descriptions of the coded word associations are reported first. Further descriptive analyses are then reported to explore patterns within the multiple responses provided by each participant to each cue word, to assess (a) whether there is evidence that the type of associate given as the first response influences later responses and (b) whether, on average, a conceptual associate is less likely to be followed by a linguistic associate than a linguistic associate is to be followed by a conceptual one.

5.3.1 Associate Types

First, as in the previous WATs (reported in Chapters 3 and 5) and described above, the number of each of the four types of linguistic and conceptual associations, plus taxonomic and other associates, was calculated, according to the cue word types. Given that Nelson et al. (2000) suggested that first responses are more reliable than subsequent responses on WATs because the second response is more likely to be given by only one participant compared with the first, the number of each type of associate to each type of cue word is shown for first responses only in Table 5.1.

Table 5.1

Cue Word Response Categories for Response One

Response Category	Cue word category								Total	
	Linguistic				Conceptual					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
FC	**235**	57	143	127	53	90	48	46	799	0.198
BC	17	**240**	23	24	42	28	33	38	445	0.110
SN	56	18	**179**	91	23	16	17	35	435	0.108
AN	5	0	9	**87**	0	4	0	4	109	0.027
OP	0	0	8	36	**186**	7	3	1	241	0.060
CP	36	24	33	28	45	**145**	27	62	400	0.099
SD	108	79	38	39	97	136	**316**	194	1007	0.249
SF	32	18	44	44	28	37	11	**97**	311	0.077
TX	0	54	11	4	5	23	38	18	153	0.038
OR	16	15	17	25	26	19	12	10	140	0.035
Linguistic	313	315	354	329	118	138	98	123	1788	0.443
Conceptual	176	121	123	147	356	323	357	354	1959	0.485

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = Other response, No. = number, Prop. = proportion. The number of responses that are of the same category or concept type as the cue word are in bold.

A balanced proportion of linguistic and conceptual responses were produced as the first associate, at just over 44% and 48%, respectively. Taking into account the cue word type, 64% of the first responses given to linguistic cue words were linguistic and 70% of first responses to conceptual cues were likewise conceptual. Furthermore, when considering the specific categories, the number of responses in each category is generally highest for the cues word of the same type, as shown by the numbers in bold. For example, of the 505 responses produced to cue words predicted to elicit forward collocates, almost half were of this type. One exception was antonym cue words, which would appear to be due to the very low number of antonym associates produced by participants overall (note that although only 87 responses to antonym cue words were antonyms, only 109 antonym responses were provided in total). The other exception was semantic field associates, but this, as discussed later, was because many of these cue words actually elicited situational descriptors as the dominant response type in the USF norms Nelson et al. (1998).

Overall, the type of association between the cue word and the first response produced by students at a university in the UK in 2018 conformed quite highly to that produced by students in Florida in the 1990s. This is important as it suggests that word association data is consistent across different generations of students in different Western countries. Indeed, just over half the words given as the dominant response were the same (42 out of 80) and a further 15 of the dominant responses given in this experiment featured within the three most frequently given response words in the USF norms (Nelson et al., 1998; and see Appendix 9). Despite most cue words tending to elicit a large number of response words that are given only by one or two respondents (Nelson et al., 2000), it would seem that dominant responses given by L1 speakers are reliable and thus can be used to explore the mental lexicon.

When all three responses given to each cue word are considered, more conceptual associates were produced than linguistic ones, consisting of approximately 54% and 35% of all responses, respectively, as shown in Table 5.2. Thus, the tendency, on average, for participants to produce more responses that enter into a conceptual association with the cue words, was replicated in this experiment, with different cue words. However, this increase in conceptual associates across all responses indicates that such associations are more frequently given as responses two and three, rather than that they are given later than linguistic associates, as claimed by Santos et al. (2011). This confirms the findings reported in the replication study and the previous experiment (Chapters 3 and 4; see also De Deyne et al., 2013).

Table 5.2

Cue Word Response Categories Across all Three Responses

Response Category	**Cue word category**								**Total**	
	Linguistic				**Conceptual**					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
FC	**549**	114	350	352	119	165	135	98	1882	0.155
BC	46	**581**	64	49	103	87	80	89	1099	0.091
SN	137	57	**417**	258	61	42	42	90	1104	0.091
AN	11	0	16	**145**	0	7	0	16	195	0.016
OP	0	11	28	134	**409**	24	9	1	616	0.051
CP	135	116	108	95	197	**448**	126	251	1476	0.122
SD	402	311	180	169	363	470	**906**	522	3323	0.274
SF	133	72	192	154	101	115	37	**299**	1103	0.091
TX	11	145	27	6	22	62	100	68	441	0.036
OR	91	108	133	153	140	95	80	81	881	0.073
Linguistic	743	752	847	804	283	301	257	293	4280	0.353
Conceptual	670	510	508	552	1070	1057	1078	1073	6518	0.538

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = Other response, No. = number, Prop. = proportion. The number of responses that are of the same category or concept type as the cue word are in bold.

Moreover, 52% of associates to linguistic cue words were linguistic (as opposed to 37% conceptual) and 71% of associates to conceptual cue words were conceptual (with only 19% linguistic). This suggests that even if the actual words given as second and third responses are less reliable, as claimed by Nelson et al. (2000), the associate type may remain reliable, especially for cue words that tend to elicit conceptual associates. This contrasts with the previous experiment, in which later associates tended to be conceptual, regardless of the cue word type. It would seem that for linguistic cue words to reliably elicit linguistic associates over multiple responses they must show a tendency to elicit several different linguistic associates, but that this is much less important for conceptual cue words.

Considering the eight specific categories, as indicated in bold in Table 5.2, many response words continued to be of the same specific type as the cue. In line with the increase in conceptual associations across the second and third responses, the proportions of the specific types of conceptual associates that are the same as the cue word are mostly consistent with the proportions for response one only (i.e., there are three times as many cue word property, situational descriptor, and semantic field associates to those cue words). In contrast, for linguistic cue words, there are just over twice as many associates of the same type as the cue. This further suggests that conceptual responses are more consistent over multiple responses than linguistic ones.

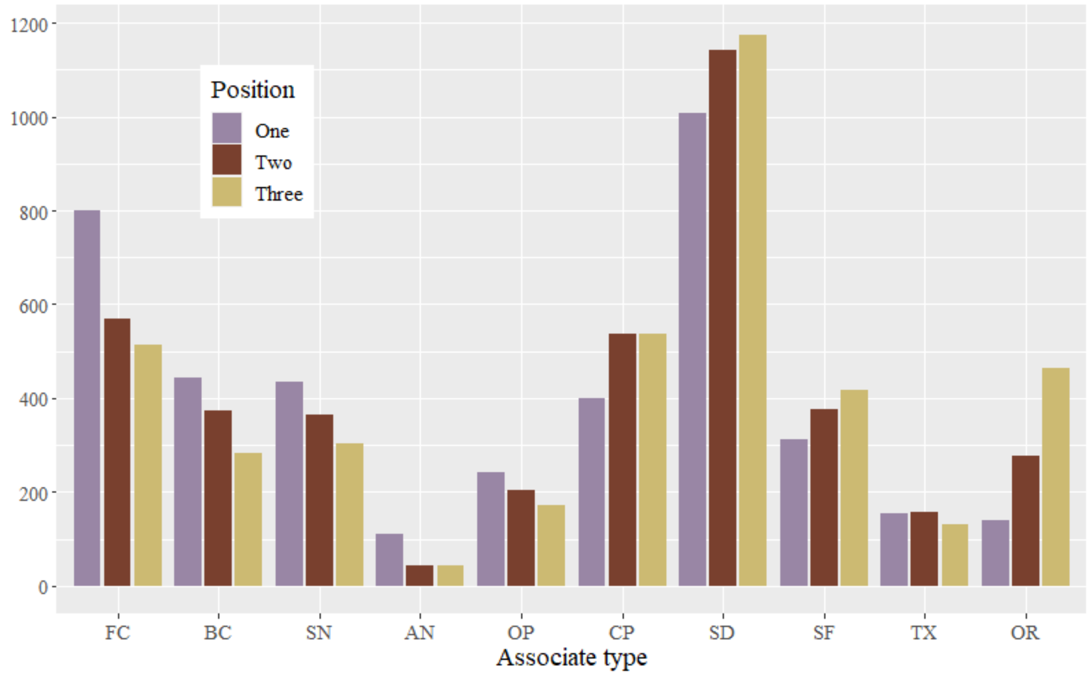
In addition, as in the previous experiments, the dominant specific response category was situational descriptors (27.4% of all responses). This type of association was frequently given to all cue word types except those predicted to elicit synonym and antonym associates. Instead, these two types of linguistic cue word tended to elicit forward collocations, contributing towards their high frequency of occurrence (15.5% of all responses). This result supports the classification of synonyms and antonyms as linguistic associates, a point that will be returned to in the next chapter when the coding system used in this thesis is compared with the paradigmatic and syntagmatic system.

The variation between the proportion of linguistic and conceptual associates produced in each output position (responses one, two, and three) is shown in Figure 5.1. This figure shows a similar pattern as in the previous experiments (Chapters 3 and 4): The number of each type of linguistic associate decreased as second and third responses, with a corresponding increase in most conceptual category types (except object property). This chart also illustrates the predominance of situational descriptors, as well as the high number of forward collocations given as the first associate. Overall, it confirms that the increase in conceptual associates as second and third responses is related to the decrease in linguistic associates, but does not suggest a tendency to produce conceptual associates later than linguistic ones.

The increase in responses categorised as other response in output positions two and three was partly due to response chaining, but also due to associates which were idiosyncratic to the participant. For such responses it was often impossible to identify the connection that was made to the cue word, such as “event: triangles” or “elephant: p53”. Others were clearly personal or individual associations such as “soldier: brother”. In all such cases, the response word did not show a generalisable, meaningful connection to the cue, which is frequently an aim of word association research (Nelson et al., 2000, 2004).

Figure 5.1

Bar Chart Showing Count of Associate Type by Output Position



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field, TX = Taxonomic, OR = other response.

In summary, the descriptive data presented above suggests that the specific types of associates produced were constrained to a considerable extent by the cue words that were chosen, especially for the first associate produced. This shows that the high frequency cue words selected for this experiment elicit stable associations from different populations, indicating that word association data can be very reliable, at least with carefully chosen cue words. In addition, the first associate elicited by each cue word tended to belong to the predicted general category, linguistic or conceptual, even when it was not of the specific predicted category. This supports the validity of the categorization system: The eight specific categories are consistent with the general category to which they have been assigned. However, in contrast to the claim made by Santos et al. (2011; replicated in Chapter 3), linguistic associates are not more likely to be produced as the first response compared with conceptual ones. In this experiment, approximately equal numbers of linguistic and conceptual associates were given as response one, although more conceptual associates were given to both types of cue words as second and third responses.

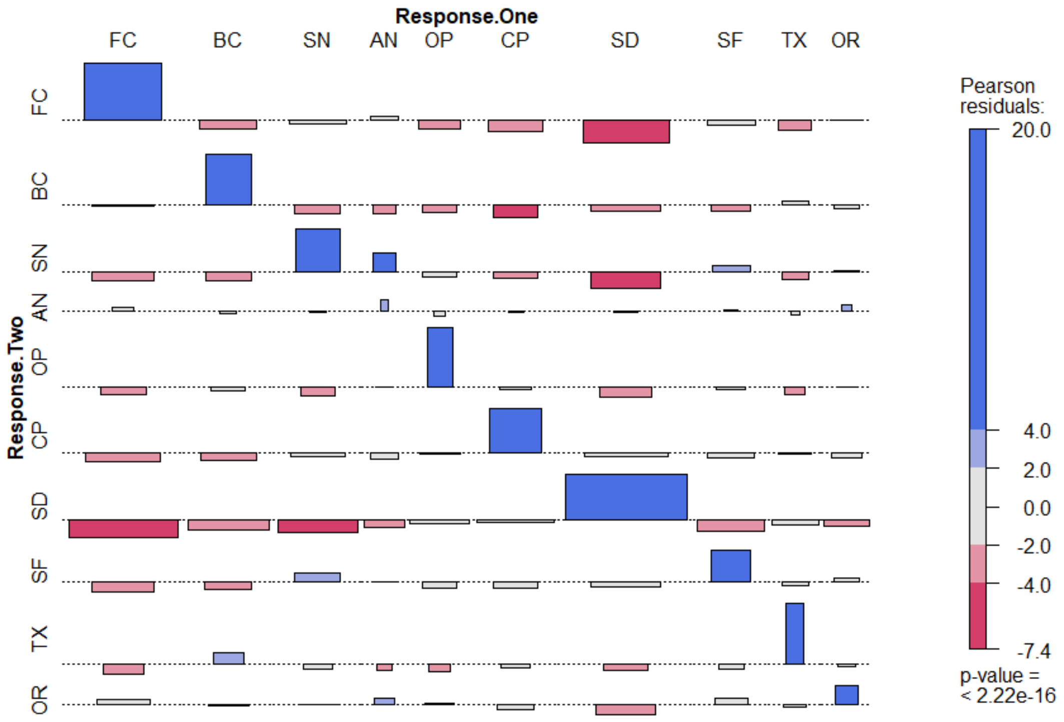
5.3.2 Relationship Between Multiple Responses

In this experiment, the cue words constrained the type of associate responses produced, yet responses two and three were still more likely to be conceptual rather than linguistic associations. The next step is to explore whether there are any systematic differences between the two types of association across multiple responses, independent of the cue words. As noted in the introduction to this chapter, any patterns across linguistic and conceptual associates may indicate that this distinction captures some information about word knowledge or processing. This would then also support the key claim of LASS Theory (Barsalou et al., 2008) and earlier dual coding theories (e.g., Paivio, 1986), that semantic knowledge can be processed through language and multimodal conceptual processing. If there are two distinct semantic systems, it is possible that later associates will be influenced by the type of association given previously, in addition to the influence of the cue word (the association it is predicted to elicit most frequently), indicating priming effects at the more abstract level of the type of link, not just at the word level. In addition, if language processing is relatively shallow compared to multimodal conceptual processing, it should be relatively more difficult to switch from a conceptual to linguistic associate than vice versa, reflecting the higher processing cost in moving from deeper to shallower processing.

To address the first question, a chi-square test was performed to assess whether the specific type of association to the cue word evinced by response one was associated with response two type. This revealed a highly significant association ($\chi^2(81) = 2697.5$, *p.* < 0.00001), which was explored in more detail via an association plot, as shown in Figure 5.2. The results were in line with those in the previous experiment, in that the likelihood that response two held the same specific type of association to the cue word as response one was significantly greater than would be expected by chance, as indicated by the dark blue colour of the bars. This tendency was found across all response types except for when an antonym was given as response one, in which case the second associate showed a significant probability of being a synonym. In addition, there was a significant trend for situational descriptors not to co-occur with forward collocations or synonyms, either as response one or two, as indicated by the deep pink shading. Note that the width of the bars represents the number of observations whereas the height of the bars indicates selectivity of the association and so higher bars are partly due to fewer observations of that type of association overall, which is why the bar representing taxonomic-taxonomic associates is highest.

Figure 5.2

Association Between Response One and Two



Note. Blue bars indicate higher numbers than expected and pink bars indicate lower numbers; deep hues indicate a significant difference FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field, TX = Taxonomic, OR = Other. The width of the bars depicts the number of responses.

One reason for this relation between response one and two was that both responses were connected to each other as well as the cue, such as “root” eliciting “plant” and “tree.” However, there were also instances where response one and two showed the same type of association to the cue word despite being unrelated to each other. For example, in response to “line” one participant gave two backward collocations, “fishing” and “telephone”, words which have little, if any, connection to each other beyond collocating with the word line. As discussed below, these results support the hypothesis that the type of association given as the first response affects the second response, not just the actual word produced (Nelson, 2000).

There remains the possibility that the association between the type of associate given as response one and two, as shown in Figure 5.2, is an artefact of both responses being the predicted response type for the cue word. In other words, the cue word concept type could be a confounding factor underlying this association. To check this possibility, the specific type of association of the first and second responses was calculated as a function of each specific cue word type and each other, as shown in Table 5.3 (responses coded as taxonomic and other associates have been removed). For each associate type the number of responses is shown in the top row and the proportion of responses is shown below.

Table 5.3

Predictive Value of the Cue Word and Response One on Subsequent Responses

Association	Linguistic categories				Conceptual categories				Total
	Type of association of response one as a function of cue word type								
	FC	BC	SN	AN	OP	CP	SD	SF	
Same specific type	235 .481	240 .550	179 .375	87 .183	186 .392	145 .313	316 .695	97 .203	1485 .396
Same general type	78 .160	75 .172	175 .367	242 .508	170 .359	180 .389	41 .090	257 .539	1218 .325
Opposite type	176 .360	121 .278	123 .258	147 .309	118 .249	138 .298	98 .215	123 .258	1044 .279
	Type of association of response two as a function of cue word type								
Same specific type	169 .360	203 .480	129 .287	35 .076	126 .280	155 .343	293 .660	107 .234	1217 .337
Same general type	57 .122	48 .113	134 .298	234 .506	239 .531	219 .485	64 .144	260 .569	1255 .348
Opposite type	243 .518	172 .407	186 .414	193 .418	85 .189	78 .173	87 .196	90 .197	1134 .314
	Type of association of response two as a function of response one type								
Same specific type	308 .427	146 .379	127 .322	5 .052	80 .360	158 .425	539 .573	85 .310	1448 .425
Same general type	128 .178	64 .166	77 .195	49 .510	101 .455	149 .401	227 .241	98 .358	893 .262
Opposite type	285 .395	175 .455	191 .484	42 .438	41 .185	65 .175	175 .186	91 .332	1065 .313

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field. Same specific type refers to the eight distinct categories used to code the associations. Same general type and opposite type refer to linguistic and conceptual associates.

Importantly, there were more instances in which response one and two showed the same specific type of association to the cue word than those in which either response one or response two were of the type predicted for the cue word. This shows that, overall, the influence of the type of association produced for response one on that of response two was independent of the predicted dominant association type of the cue word: response one type was a better predictor than cue word type of response two type. However, this influence of response one on response two was not a uniform effect across all response types. Based on absolute numbers, response one and two tended to be the same type when both were situational descriptor and forward collocation associates. However, these were the highest frequency conceptual and linguistic responses. When the proportion of responses is considered, the influence of situational descriptor associates on the second response is reduced, whereas synonym, object and cue properties, and semantic field associates produced as a first response exert greater influence on the second response type than the cue words. It can also be seen that these trends emerged despite the high probability of producing a conceptual associate after a linguistic one.

In interpreting the total numbers of each pair of types, displayed in the final column in Table 5.3, it is important to consider the difference from the expected values if there was no effect of association type (i.e., if all combinations were equally frequent). There is only one combination in which the specific type of association is the same, compared to three combinations in which the general type of association is the same and four combinations in which it is the opposite type. This means that, under an equal distribution, only 430 pairs would be of the same specific association type, 1290 pairs would be of the same general type, and 1720 pairs would be of the opposite type[9]. Despite the large numbers of responses that were of the opposite type to the predicted dominant association type of the cue word or to the response given first, these numbers are considerably lower than chance values. Overall, these patterns within specific and general response type across responses one and two support the claim that word association reflects two semantic processing systems.

To further assess the tendency to produce a conceptual associate after a linguistic one but not to produce a linguistic associate after a conceptual one, the general type of association (conceptual or linguistic) given to each type of cue word was calculated across each output position (response 1, 2, and 3). These figures are

---

[9] These numbers are slightly higher for the association between the cue words and response one, at 432, 1296, and 1727, respectively.

shown in Table 5.4 (taxonomic and other associates have been removed), along with the difference between the actual numbers obtained and the number that would be expected by chance (i.e., if all cells for each cue word type contained an equal number of response types). This shows that there was an extremely strong tendency for conceptual cue words to elicit three conceptual associates, with this response pattern accounting for 57% of all responses to this cue type. Although linguistic cue words also elicited three linguistic associates more frequently than any other response pattern, this amounted to only 29% of the responses. Moreover, linguistic cue words also elicited three conceptual associates slightly more frequently than would be expected if all combinations of responses were equally common.

Table 5.4

Linguistic and Conceptual Responses as a Function of Cue Word Type

Response one	Response two	Response three	Cue word type		Difference from equal distribution		Combined difference
			Linguistic	Conceptual	Ling	Conc	
Linguistic	Linguistic	Linguistic	431	53	245	-137	108
		Conceptual	233	71	47	-119	-72
	Conceptual	Linguistic	138	59	-48	-131	-179
		Conceptual	226	183	40	-7	33
Conceptual	Conceptual	Conceptual	221	868	35	678	713
		Linguistic	78	129	-108	-61	-169
	Linguistic	Conceptual	84	113	-102	-77	-179
		Linguistic	80	45	-106	-145	-251

Note. Ling = linguistic, Conc = conceptual.

For conceptual cue words, given the number of instances that three conceptual associates were produced, it is not surprising that all other response combinations occurred less frequently than expected by chance. However, the production of a linguistic associate followed by two conceptual ones was almost at chance level. It is of note that for linguistic cue words, other than producing three associates of the same type (i.e., all linguistic or all conceptual), the production of two linguistic followed by one conceptual associate and one linguistic followed by two conceptual associates also occurred above chance levels. Overall, participants were much less likely to produce a conceptual associate after producing a linguistic one than they were to follow a linguistic associate with a conceptual one, independent of the cue word type, suggesting that conceptual associates reflect deeper, multimodal processing (situated simulation) compared with linguistic associates, which are produced through shallower lexical semantic processing (Barsalou et al., 2008).

5.4 Discussion

The methodological aim of the WAT reported in this chapter was achieved; a more balanced proportion of linguistic and conceptual associates was elicited than in the previous experiment (reported in Chapter 4): 35% of responses were linguistic associates, compared with 54% conceptual associates. The main research question guiding this experiment, based on theories of two separate semantic systems, specifically LASS theory (Barsalou et al., 2008), was also answered in the affirmative: analysis of the WA data revealed systematic differences in the distribution of linguistic and conceptual associations. Moreover, the specific type of association produced in response one influenced subsequent responses, at least in part independently of the cue word type. These two patterns across multiple responses to each cue are discussed next.

The distribution of linguistic and conceptual associates across output positions (i.e., response one, two, and three to each cue) showed that participants produced fewer linguistic associates as second and third responses, with a corresponding increase in conceptual associates (as was also found in the experiments reported in Chapters 3 and 4; see also De Deyne et al., 2013). More detailed analysis of this distribution revealed that this relative reduction in linguistic associates at later output positions was at least partly due to participants being less likely to produce a linguistic associate after a conceptual associate than vice versa. This finding suggests that switching from conceptual to linguistic processing was relatively more effortful. This supports the claim of LASS theory (Barsalou et al., 2008; see Chapter 2, Section 2.7) that linguistic processing of semantic meaning is shallower than the situated simulations that constitute conceptual processing.

The second question was whether the type of association made in the first response influenced later response types. Nelson et al. (2000) noted that when participants produced two associates per cue word, both associates were often related to each other as well as to the cue word. Such biasing effects of earlier responses on the current one have also been shown in category generation tasks (e.g., Troyer et al., 1997), in which the exemplars produced for a given category influence subsequent exemplars (a possible explanation is explored in Chapter 8, Section 8.3). In the experiment reported in this chapter, analysis of WAs revealed that such biasing effects, which could be considered a type of self-priming, were found at the more abstract level of association type, not solely at the level of the actual words produced. Crucially, the influence of the specific associate type produced in the first response on that of the second response was greater than the influence of the cue word on the second response. This is notable given that responses showed considerable consistency with the predicted category type for the cue words. Participants not only provided context for the cue words through their association words, but also through the type of link they produced. This supports the idea that WATs may be more informative about processing than storage of semantic knowledge (Santos et al., 2011).

However, the high number of conceptual associates produced as the first associate does not support the claim of LASS theory (Barsalou et al., 2008) that linguistic knowledge is accessed more quickly than conceptual knowledge. Although this claim cannot be ruled out by WATs, the findings of this and the previous experiments together provide robust evidence that associations between individual words may be dominated by conceptual knowledge. This challenges the assumption that WATs are mainly a linguistic task (e.g., Osgood et al., 1954; Santos et al., 2011), which may have implications for interpreting word association data, as discussed in Chapter 8. Overall, the findings support theories which posit that semantic meaning is processed through linguistic and conceptual systems, which make separate contributions to such knowledge (e.g., Barsalou, 1999; Barsalou et al., 2008; Paivio, 1986; see Chapter 2, Section 2.6).

In addition, the primary responses to the high frequency words chosen as cues for this experiment were relatively consistent with those given in the USF norms (Nelson et al., 1998). Almost half of the three primary responses were the same as in the USF norms, when considering only the first responses given and all three responses to each cue word (see Appendix 9). This suggests that collecting multiple responses does not reduce the reliability of WA data, in line with the conclusions of De Deyne et al. (2018), but contrary to those of Nelson et al. (2000).

Finally, the relatively low proportion of synonym responses compared with the previous experiments (Chapters 3 and 4) may be due to the choice of high frequency words as cues. This meant that the cues tended to be the highest frequency synonym available within any cluster of word meanings. For example, although the cue “fat” makes one higher frequency synonym, “big”, available (which was produced as a response), this synonym has many meanings beyond that indicated by “fat”. All other potential synonym responses are lower frequency words or marked forms (see Benor & Levy, 2006, and references therein), such as “obese” or “lard”, which, due to their lower accessibility might be less likely to be given as associates.

5.4.1 Coding Issues and Next Steps

Although the cue words used in this WAT elicited a more balanced mix of linguistic and conceptual associates, the antonym and semantic field categories were somewhat problematic. Firstly, an extremely small number of antonym responses were obtained (as also observed in the previous experiment, see Chapter 4), despite antonym cues eliciting many of the most robust associations in the USF norms (Nelson et al., 1998), such as “good: bad” or “slow: fast”. However, such antonyms can be categorised as binary pairs, necessarily interdependent concepts (Keil & Kominsky, 2015). This finding suggests that antonym associations may be quite restricted to such binary pairs. Although most of the antonym responses were elicited by the cue words chosen to elicit antonyms, these cue words showed a greater tendency to elicit forward collocations and synonyms. Overall, this type of association appeared to be less available for other word cues.

On coding the WA data, I realised that some of the cue words chosen within the category of semantic field actually elicited a situated descriptor as the dominant associate in the USF norms (Nelson et al., 1998). In part this was due to the difficulty in finding cue words that met the requirements of eliciting an associate that was related to the cue by semantic field as the dominant response in the USF norms and that were within the first 1000 or 2000 most frequent word families. Rather than the associate being related to the cue through an abstracted conceptual association, the cue word formed an abstracted semantic association to an experiential response. For example, the dominant associates elicited by the cue word “education” were “school”, “university”, and “learn”, which may be considered to be related by simulation based on the abstract concept of education. This issue partially accounts for the large number of situational descriptors produced in response to semantic field cue words.

Moreover, coding responses as semantic field associations also revealed possible contradictions. This may, in part, have resulted from how this category was operationalised in the study by Santos et al. (2011), as the replication of the WAT therein (see Chapter 3) informed the categorisation and coding choices made in this experiment and the previous one (Chapter 4). In Santos et al. (2011) the cue words that were chosen for the semantic field category were all members of limited lexical sets, such as winter, Monday, and Jupiter. In analysing their data, the associates to semantic field cues were typically coded as taxonomic associations. However, as a cue word type, semantic field was included as a conceptual response category by Santos et al. Based on their decision regarding the cue words and on the responses provided by participants, in replicating their WA experiment, associates were coded as related by semantic field if they had a conceptual relation to the cue word that did not reflect situated simulation or a property.

In effect, this means semantic field associates were words that were semantically related to the cue but may lack strong sensorimotor properties. In other words, such responses were semantically related in an abstract manner, such as the response “religion” to the cue “church.” This type of abstract, conceptual relation was also coded as distinct to situational or featural relations by Lenci et al. (2013) in a property generation task. Given that a central tenet of LASS theory is that abstract knowledge may be grounded by language as much as (or more than) experience (Barsalou et al., 2008; see also Paivio, 1986), it is thus unclear whether semantic field associates should be coded as linguistic rather than conceptual. This issue is discussed in more detail in Chapters 6 and 8; here it is merely noted that if semantic field associates reflect linguistic knowledge as much as, or more than, conceptual knowledge, this would have made it more difficult to detect the patterns reported in this chapter. In other words, if semantic field associates have been wrongly coded, this will have lessened the distributional differences between linguistic and conceptual associates, resulting in underestimation of the true effect.

Overall, the findings reported in this chapter suggest that there are systematic differences between linguistic and conceptual associates, supporting the idea that they represent two different semantic systems (Barsalou, 1999; Barsalou et al., 2008; Paivio, 1986). The question remains as to whether there are advantages to coding

WA data in this way. Many previous studies in which free word associations have been collected and coded have used a syntagmatic-paradigmatic classification (see Chapter 2, Sections 2.1.2 and 2.2), proposed by Osgood et al. (1954). However, studies using the paradigmatic-syntagmatic distinction have produced inconsistent results (for a review, see Fitzpatrick & Thwaites, 2020). In the following chapter the new coding system used here will be directly compared to the syntagmatic-paradigmatic coding system. In addition, the specific categories within linguistic and conceptual associates are compared with those within meaning- and position-based responses (Fitzpatrick, 2006, 2007, 2009; see Chapter 2, Sections 2.1.2 and 2.2). If similar patterns are found when applying these two language-based coding systems as with the linguistic-conceptual coding system, it would be difficult to claim that the new coding system reveals new insight into WA processes. Such a null result might raise questions about what coding WA data contributes to understanding the underlying processes. However, if the patterns uncovered in this chapter differ, or are not found at all, when using the syntagmatic-paradigmatic and the meaning- and position-based coding systems, this might further support that the linguistic-conceptual distinction captures a genuine difference in how semantic knowledge is processed and/or represented.

Chapter 6. Analytical Comparison of Two Word Association Coding Systems

The results reported in the previous chapter suggest that the distinction between linguistic and conceptual associates on WATs may reflect differences in processing or representing meaning. This assertion is based on two key findings, related to the general categorisation of the association type as linguistic or conceptual and to the four specific categories used within each of these two types.

First, as predicted in the choice of cue words, linguistic and conceptual associates were produced in almost equal proportions by participants as the first associate. In this position, conceptual associates were only slightly more frequent than linguistic ones. However, despite the tendency to produce either three conceptual or three linguistic associates in response to the cue words, conceptual associates were much more frequently given in later output positions. This reflected the decreased likelihood that a conceptual associate would be given after a linguistic one and the much greater relative probability that a linguistic associate would be followed by one that was conceptually related to the cue word. Thus, switching from a linguistic to a conceptual associate may be easier for participants than switching from a conceptual to a linguistic associate. This supports the claims of dual processing theories that linguistic processing is shallower than multimodal conceptual processing (e.g., Barsalou et al., 2008; Paivio, 1986).

Second, there was a significantly greater likelihood than would be expected by chance that the first two responses produced held the same specific type of association to the cue word. This trend was more pronounced than the tendency that the second response held the association to the cue word that had been predicted based on the USF norms (Nelson et al, 1998). This finding was interpreted as a form of self-priming of association type.

In summary, the use of a coding system based on distinctions between linguistic and conceptual associations revealed two patterns in the WA data that indicate they may reflect two different types of processing and/or knowledge. However, before the distinction between linguistic and conceptual associates can be claimed to be psychologically valid, it should be considered whether similar response patterns could be uncovered using different coding systems. This is the purpose of this chapter.

6.1 Data Analysis and Aims

This chapter reports the comparison of the new linguistic-conceptual coding system with two established coding systems that are both grounded in linguistic theory. The data collected in the previous experiment (Chapter 5) is used for these comparisons. Given that, to the best of my knowledge, no WA study has directly compared different coding systems for analysing the data, the analyses reported in this chapter are considered exploratory, and no hypotheses are made. Overall, the aim is to assess whether the novel coding system developed in Chapters 4 and 5, based on Santos et al. (2011), may offer advantages towards understanding the mental lexicon through WA. Coincidences between coding systems may imply they offer similar insights into WA data, as used to explore the development of the mental lexicon, which may question the informativeness of coding WAs. In contrast, any significant differences between coding systems would suggest they capture different aspects of WA, supporting such analytical approaches.

First, I compare the general categories of linguistic and conceptual associates with the paradigmatic-syntagmatic coding system proposed by Osgood et al. (1954). As described in Chapter 2 (Sections 2.1 and 2.2), this system has been used extensively in studies that have coded WA data. The aim is to explore whether the participants also tend to produce three paradigmatic or three syntagmatic responses and whether this system reveals a similar trend to the tendency not to produce a conceptual associate after a linguistic one.

Second, the eight specific categories within linguistic and conceptual responses are compared with a specific coding system developed to expand the paradigmatic-syntagmatic framework (Fitzpatrick, 2007; see Chapter 2, Sections 2.1.2 and 2.2). Fitzpatrick used nine categories (described in Section 6.3) to divide responses into meaning-, position-, and form-based responses (plus an erratic response category, similar to the other category used in this thesis). The aim is to assess the equivalence of the specific categories used in both coding systems. As the four sub-categories used to code linguistic associates in this thesis are similar to categories used by Fitzpatrick, it is expected that both systems will reveal self-priming of specific types. However, the theoretically-driven nature of the sub-categorisation of conceptual associates, based on Santos et al. (2011), may lead to some differences in interpreting the data.

In addition, in Section 6.5, further analyses are reported to test the claim that conceptual associations reflect deeper, multimodal processing compared to linguistic ones (Barsalou et al., 2008; Santos et al., 2011). The response words, as categorised using both the linguistic-conceptual and the paradigmatic-syntagmatic coding systems, are assessed using the Lancaster Sensorimotor Norms (LSN; Lynott et al., 2020). If response words coded as conceptual associates are rated higher for sensorimotor experience than those coded as linguistic, this would support this distinction between associate types. In addition, there should be no significant difference in strength of sensorimotor experience between responses coded as paradigmatic compared with those coded as syntagmatic, as both associate types are assumed to reflect linguistic knowledge (Osgood et al., 1954).

All analyses were conducted in RStudio (RStudio Team, 2020), run on R (R Core Team, 2021). Coding was based on various sources, notably Levshina (2015) and Winter (2020). Tables and analyses were produced using the tidyverse (Wickham et al. (2019) and vcd packages (Meyer et al., 2006; 2022) and figures using ggplot2 (Wickham, 2016), following Healy (2019). The colour scheme for some graphs was taken from the wesanderson package (Ram & Wickham, 2018).

6.2 Paradigmatic-Syntagmatic vs. Linguistic-Conceptual Associates

The novel classification system used in the two experiments reported in Chapters 4 and 5 was an adaptation of that used by Santos et al. (2011), in which the key dichotomy is between linguistic and conceptual associates. This framework, stemming from embodied cognition theories of language and concepts (see Chapter 2, Sections 2.5 and 2.6), is based on the assumption that word meanings can be processed by activating linguistic semantic representations or conceptual multimodal simulations (see Chapter 2, Sections 2.6 and 2.7). However, the analyses conducted do not indicate whether this new system should be preferred to the paradigmatic-syntagmatic distinction that has guided the coding system used in multiple WA studies within applied linguistics (Osgood et al., 1954; see Chapter 2, Sections 2.1 and 2.2). Therefore, a direct comparison of the two coding approaches is needed to assess the informativeness of each system.

The paradigmatic-syntagmatic coding system assumes WA can be analysed purely as a language task and so makes a distinction between whether the relationship between the cue and response words is such that the response could replace the cue in a sentence or that the response could appear before or after the cue word in a sentence. Importantly, the relations classified as linguistic in the coding system proposed in this thesis include both paradigmatic and syntagmatic relations: Forward and backward collocations are syntagmatic associates, whereas synonyms and antonyms are paradigmatic. Moreover, all four specific types of conceptual associations (cue word properties, object properties, situational descriptors, and semantic field) may be either paradigmatic or syntagmatic. The labels divide responses in very different ways and are not interchangeable.

6.2.1 Application of the Syntagmatic-Paradigmatic Coding System

The paradigmatic-syntagmatic coding system was proposed as a linguistic coding system for WA data (Osgood et al., 1954). Paradigmatic associations are response words that might replace the cue word in a sentence, so are most likely to be the same part of speech (i.e., the noun "dog" might replace the noun "cat"). In contrast, syntagmatic associations are response words that might follow or precede the cue word in a sentence, so are most likely to be different parts of speech (i.e., the noun "dog" might be followed by the verb "bark" or preceded by the adjective "friendly").

As highlighted by Osgood et al. (1954) very early in the adoption of this coding framework, the degree to which a response is paradigmatically or syntagmatically related to a cue word is highly variable. For example, "cat" is more similar to "dog" than is "octopus" and is more likely to replace it in a sentence. Likewise, the verb "bark" is typically associated with activities of dogs and with few other entities, whereas "friendly" could be used to modify a large variety of nouns. In this way, syntagmatic and paradigmatic associates may be viewed as more or less informative about the cue word. In addition, sometimes the distinction is difficult to determine, for example, should "leash" be coded as a paradigmatic associate of "dog" as both are meaningfully related nouns or as a syntagmatic associate, in the sense of the dog's leash?

Although coding decisions are also not always straightforward with the linguistic-conceptual classification, a systematic method is needed for applying the paradigmatic-syntagmatic coding system in this chapter. The simplest system would be to determine association type purely according to word class, as has been done previously (e.g., Ervin, 1961; Piranian, 1983). However, this raises issues in coding noun pairs, as it would result in associations such as “belt: seat” being coded as paradigmatic despite the most likely association being that of backward collocation and thus syntagmatic, as noted by previous researchers (e.g., Nissen & Henriksen, 2006: Wolter, 2001). If collocational noun pairs are coded as a syntagmatic relation, this raises questions about other noun pairs in which the most probable relation seems to be one of co-occurrence rather than equivalence of meaning, such as those involving an animate and an inanimate entity (e.g., child: cot).

In previous studies the application of the paradigmatic-syntagmatic distinction has been illustrated by clear examples of each type of relation, such as “dog: cat” (paradigmatic) or “read: book” (syntagmatic). Given that many studies suggest paradigmatic relations are more frequently given by adults and that they have been interpreted as evidence of better developed language networks (e.g., Burke & Peters, 1986; Ervin, 1961; Sandgren et al., 2021), it is assumed that researchers have tended to resolve problematic instances by categorising relations as paradigmatic rather than syntagmatic. Therefore, noun-noun associates were coded as syntagmatic only if they formed a noun pair, as in “station: train”; if they referred to inanimate and animate entities; or if they contrasted time or place with an entity or thing. Examples of noun pairs that were coded as syntagmatic are “elephant: zoo”, “soldier: war”, “beach: sunglasses”, and “weekend: camping”.

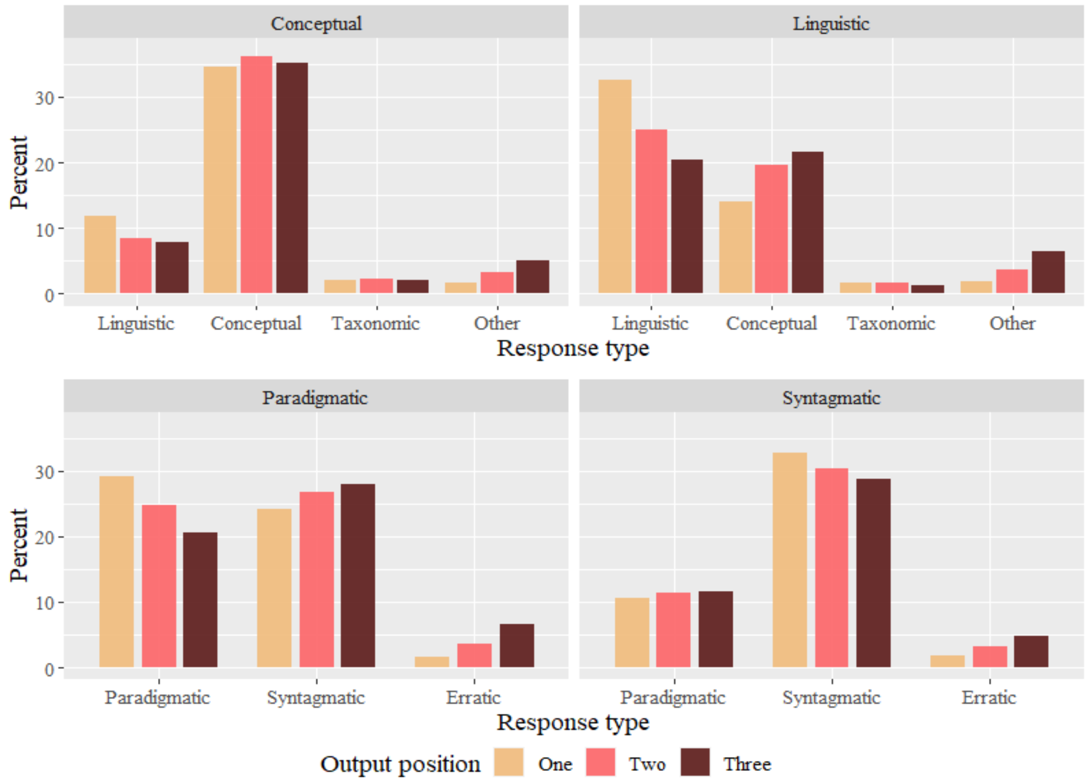
After applying the paradigmatic-syntagmatic coding system to the WA data collected in Chapter 5, erratic associations were removed, to enable direct comparison of paradigmatic and syntagmatic associates with conceptual and linguistic ones. Exploratory analyses were conducted to examine whether both systems reveal similar patterns, such as the tendency for linguistic associates to precede rather than follow conceptual ones.

6.2.2 *Relationships Between Multiple Responses*

Figure 6.1 shows the general response types across all three responses, according to both coding systems (linguistic-conceptual and paradigmatic-syntagmatic). There were 40 linguistic and 40 conceptual cue words, but these were recoded as 44 paradigmatic and 36 syntagmatic cues, so the percentage of responses given to paradigmatic cues at each position equals 55% compared with 45% to syntagmatic cues. Surprisingly, participants produced more syntagmatic than paradigmatic associates. The distribution of paradigmatic associates resembled that of linguistic ones, whereas syntagmatic associates were more similar to conceptual ones. However, paradigmatic associates did not show such a marked decrease in in later response positions as linguistic associates.

Figure 6.1

General Response Types for Each Cue Type



Note. Bar charts show the percent of response types as a function of the cue word type and output position, using the two different coding systems. For each output position, the responses across both cue types (conceptual and linguistic in the top panels and paradigmatic and syntagmatic in the bottom panels) sum to 100 per cent.

The relationship between the three response words given as a function of the predicted cue word type (paradigmatic or syntagmatic) was calculated, as shown in table 6.1 (erratic associates have been removed). There were 1817 responses to paradigmatic cue words compared with 1486 responses to syntagmatic cue words. Despite the larger number of cues for which the dominant associates in the USF norms (Nelson et al, 1998) was paradigmatic, participants tended to produce three syntagmatic associates. This trend dominated response behaviour to both cue types. For syntagmatic cue words, all other response patterns were less likely than would be predicted by chance, although the production of two syntagmatic associates followed by a paradigmatic associate was at chance level. For paradigmatic cue words, participants also tended to produce three paradigmatic associates.

Table 6.1

Paradigmatic and Syntagmatic Responses as a Function of Cue Word Type

Response			Cue word type		Difference from chance		Total difference
One	Two	Three	Paradigmatic	Syntagmatic	Para	Syn	
Paradigmatic	Paradigmatic	Para	323	71	96	-115	-19
		Syn	249	89	22	-97	-75
	Syntagmatic	Para	166	69	-61	-117	-178
		Syn	259	136	32	-50	-18
Syntagmatic	Syntagmatic	Syn	367	693	140	507	647
		Para	141	181	-86	-5	-91
	Paradigmatic	Syn	165	134	-62	-52	-114
		Para	147	113	-80	-73	-153

Note. Para = paradigmatic, Syn = syntagmatic.

When associates were coded as linguistic or conceptual (see Chapter 5, Table 5.5), participants tended to produce three associates of the same type as the first and second response, most frequently giving three conceptual associates. In addition, there was a clear trend not to follow a conceptual associate with a linguistic one, although the reverse did not hold. In contrast to these response patterns, there was no trend not to produce a paradigmatic associate after a syntagmatic one, but rather just an overall trend to produce syntagmatic associates.

The percentages of linguistic and conceptual and paradigmatic and syntagmatic associates across each response position are shown in Table 6.2, without considering the cue word type. It shows that the tendency to produce three associates of the same type is more pronounced when responses are coded as linguistic and conceptual. Moreover, the tendency not to produce a linguistic associate after a conceptual one is not mirrored by a tendency not to produce a syntagmatic associate after a paradigmatic one.

Table 6.2

General Relationship Between Responses in Both Systems

Response One	Response Two	Response Three	Pc.	Response One	Response Two	Response Three	Pc.
Conceptual	Conceptual	Conceptual	36.1	Paradigmatic	Paradigmatic	Paradigmatic	11.9
		Linguistic	6.9			Syntagmatic	10.2
	Linguistic	Conceptual	6.5		Syntagmatic	Paradigmatic	7.1
		Linguistic	4.2			Syntagmatic	12.0
Linguistic	Linguistic	Linguistic	16.0	Syntagmatic	Syntagmatic	Syntagmatic	32.1
		Conceptual	10.1			Paradigmatic	9.7
	Conceptual	Linguistic	6.5		Paradigmatic	Syntagmatic	9.1
		Conceptual	13.6			Paradigmatic	7.9

Note. Pc. = percent. Percentages are displayed because the number of responses provided within each coding systems is unequal due to the exclusion of taxonomic associates in the linguistic-conceptual coding system.

To confirm the validity of these differences in linguistic and conceptual associates compared with paradigmatic and syntagmatic associates, the dominant or primary associates produced as the first response and across all three responses were analysed. This showed substantial differences: For 25% of response words, the dominant associate given as the first associate was different from the dominant associate overall. For example, the dominant first associate to the cue “model” was “super”, whereas overall, the dominant response was “fashion”.

However, it could be argued that assessing the two coding systems based solely on the dominant response to each cue word may not be very valid. Thus, the analysis was also conducted on the three primary first responses and the four primary responses across all output positions. These decisions were made based on inspection of the data, which showed that across all cue words, the top three first associates showed high stereotypy (were produced by a large proportion of participants) as did the top four responses overall.

As shown in Table 6.3, the linguistic-conceptual coding system revealed a reversal in dominant or primary associate types that was not found with the paradigmatic-syntagmatic one. When considering only the first responses produced, more of the primary responses were linguistic than conceptual, but this trend reversed across all output positions. In contrast, when coded as paradigmatic and syntagmatic associates, there was no difference between the proportion of primary associates of each type when considering only the first response or all three responses.

Table 6.3

Dominant Associates in Each Coding System

Associate classification	Dominant associate Response One	 All Responses	Top 3 Response One	Top 4 All Responses
Linguistic-Conceptual				
Linguistic	41 (51%)	36 (45%)	121 (50%)	140 (44%)
Conceptual	36 (45%)	41 (51%)	107 (45%)	167 (52%)
Taxonomic	2 (3%)	2 (3%)	9 (4%)	8 (3%)
Erratic	1 (1%)	1 (1%)	3 (1%)	5 (1%)
Paradigmatic-Syntagmatic				
Paradigmatic	39 (49%)	40 (49%)	110 (46%)	143 (45%)
Syntagmatic	40 (50%)	39 (50%)	127 (53%)	172 (54%)
Erratic	1 (1%)	1 (1%)	3 (1%)	5 (1%)

6.3 Expanding the Dichotomous Coding Systems

In the experiments reported in this thesis, four subcategories of linguistic and conceptual associates were used to gain a more detailed understanding of the data. Based on these subcategories, it was claimed that response behaviour showed evidence of self-priming, in that participants tended to produce the same type of association in the first and second response. However, such self-priming may be apparent when applying any WA coding system with detailed response categories, rather than implying the linguistic and conceptual categories yield new information about WA behaviour. Therefore, a comparison was conducted at this finer level of analysis. For this purpose, the categorisation system developed by Fitzpatrick was used (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015; see Chapter 2, Sections 2.1 and 2.2), as it is the only detailed coding system that has been applied in multiple studies.

In this framework, Fitzpatrick (2006, 2007, 2009) redefined the paradigmatic-syntagmatic division into the more user-friendly terms of *meaning-based* and *position-based associations*, as well as including form-based and erratic associations (in this thesis both these latter associations have been coded as other associations). These broad category types were used consistently in all three studies (and in Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015), although the sub-category types have varied between studies (as did the language(s) spoken by participants, the cue words used, and the research questions). In Fitzpatrick (2006) 17 categories were employed, making comparison with the 10-category system used in this thesis less straightforward. However, Fitzpatrick (2007, 2009) used 10 and 11 categories, respectively, with the only difference being that in the 2009 study, the erratic association category was divided into erratic and no association. Therefore, the 2007 system, shown in Table 6.4, was used for comparison (see also Chapter 7).

Table 6.4

Classification of Responses in Fitzpatrick (2007)

Category	Subcategory	Definition
Meaning-based association	Defining synonym DS	x means the same as y
	Specific synonym SS	x can mean y in some specific contexts
	Lexical set/context related LS	x and y same lexical set/coordinates/meronyms/ superordinates/provide context
	Conceptual association CA	x and y have some other conceptual link
Position-based association	Consecutive xy collocation FC	y follows x directly (includes compounds)
	Consecutive yx collocation BC	y precedes x directly (includes compounds)
	Other collocational OC	y follows/precedes x in phrase with word(s) between them
Form-based association	Change of affix AC	y is x plus or minus affix
	Similar form not meaning FS	y looks or sounds similar to x but has no clear meaning link or is an associate of a word with a similar form to x
Erratic association	No link/blank EA	y has no decipherable link to x or no response given

Reproduced from Fitzpatrick (2007, Table 1).

It should be noted that the meaning-based and position-based categories created by Fitzpatrick (2006, 2007. 2009) do not correspond directly to paradigmatic and syntagmatic associations. Position-based associates are all types of collocation, thus they are all syntagmatic. However, within meaning-based associations, although synonym and lexical set associations are paradigmatic, conceptual associations could be either paradigmatic or syntagmatic. This suggests that when using Fitzpatrick's (2007) coding system, most responses will be coded as meaning-based. In addition, lexical set includes all taxonomic associates but also extends to incorporate a broader range of responses, including antonyms. Based on the definition, in applying Fitzpatrick's system, script relations (Schank & Abelson, 1977; see Chapter 2, Section 2.3.1) were coded as lexical set, such as "hospital: doctor".

In the experiments reported here, the cue words were selected according to the dominant associates in the USF norms, such that there were ten cue words in each of the eight categories. Prior to analysis, the cue words were recoded using Fitzpatrick's (2007) classification system. Although she divided associates into ten categories, only six of the specific concept or associate types were needed to code the cue words. Two of these types, forward and backward collocation, were used in the new system developed in this thesis, meaning that position-based cue words consisted of ten forward and ten backward collocation cues. The other 60 cue words were all meaning-based association types, with four defining and six specific synonyms, 24 lexical set cues (ten antonym, eight object property, two cue property, and four situational descriptor cues in the linguistic-conceptual coding system), and 26 conceptual association cues (two object property, eight cue property, six situational descriptor, and ten semantic field cues) All ten response categories employed by Fitzpatrick (2007) were used in coding responses.

6.3.1 Meaning-, Position-and Form-Based Associate Types

First, the number of each type of response per cue word type was calculated, both for the first response given and across all three responses. This information is identical to that reported in Chapter 5 (Section 5.3.1), for the specific categories within the linguistic-conceptual coding system. Given 50 of the 80 cue words were coded as lexical set and conceptual association when using Fitzpatrick's (2007) coding system, most responses might be expected to be of these two types.

Considering the first response only, as shown in Table 6.5, conceptual associations accounted for nearly 38% of responses. However, forward collocations as frequent as lexical set responses, with both types of responses comprising just under 18% of all responses. The results were similar when all three responses were considered, as shown in Table 6.6. However, the proportion of conceptual associations was even greater, amounting to nearly 45% of all responses. Concomitantly, the proportion of all other types of association decreased slightly, except for erratic associations, which increased to nearly 6% of responses.

Table 6.5

Cue Word Response Categories for Response One

Response Category	Cue word type						TOTAL	
	Position-based		Meaning-based					
	FC	BC	DS	SS	LS	CA	No.	Prop.
FC	233	44	29	96	162	149	713	0.176
BC	15	235	9	12	59	77	407	0.101
OC	4	18	1	19	48	34	124	0.031
DS	19	11	29	40	49	8	156	0.039
SS	37	7	43	67	66	59	279	0.069
LS	21	79	16	1	446	144	707	0.175
CA	160	96	67	59	346	786	1514	0.375
AC	9	2	6	1	18	24	60	0.015
FS	2	6	0	0	5	2	15	0.004
EA	5	7	4	6	14	29	65	0.016
Position	252	297	39	127	269	260	1244	0.308
Meaning	237	193	155	167	907	997	2656	0.658
Other	16	15	10	7	37	55	140	0.035

Note. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association. No. = number, Prop. = proportion. The number of responses that are of the same category or concept type as the cue word are in bold.

Table 6.6

Cue Word Response Categories Across all Three Responses

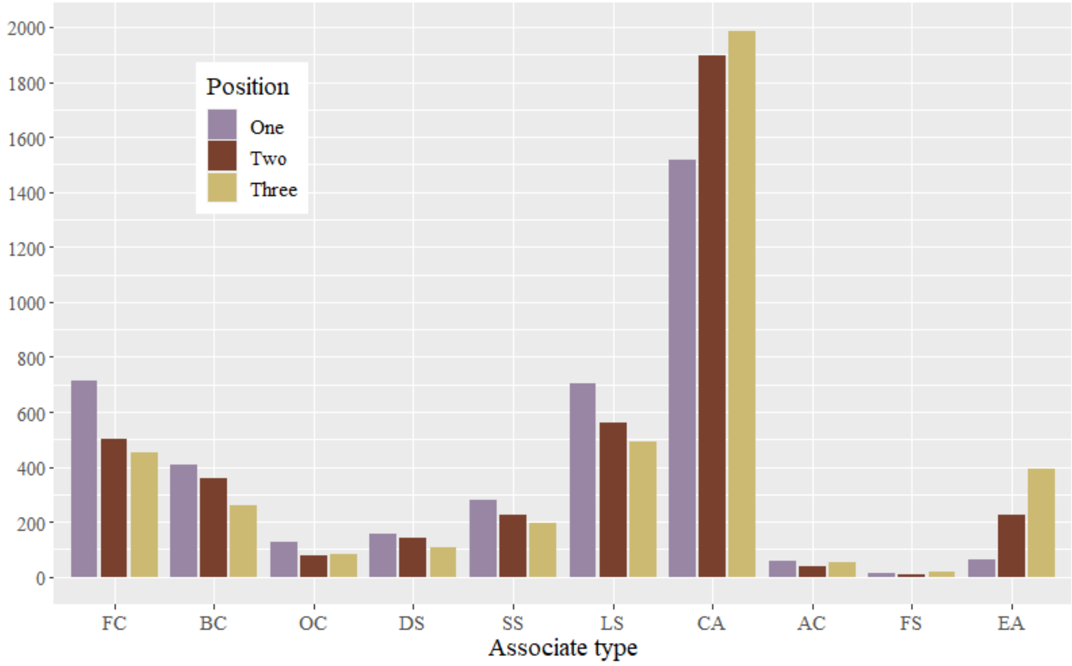
Response Category	Cue word type						TOTAL	
	Position-based		Meaning-based					
	FC	BC	DS	SS	LS	CA	No.	Prop.
FC	542	88	66	250	404	318	1668	0.138
BC	42	568	23	36	141	216	1026	0.085
OC	11	39	7	32	123	75	287	0.024
DS	41	31	72	82	139	38	403	0.033
SS	96	26	87	176	182	134	701	0.058
LS	73	199	31	7	997	452	1759	0.145
CA	619	456	265	248	1396	2411	5395	0.445
AC	22	9	16	5	55	48	155	0.013
FS	8	13	1	1	10	11	44	0.004
EA	61	86	44	66	192	233	682	0.056
Position	595	695	96	318	668	609	2981	0.247
Meaning	829	712	455	513	2714	3035	8258	0.681
Other	91	108	61	72	257	292	881	0.073

Note. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association. No. = number, Prop. = proportion. The number of responses that are of the same category or concept type as the cue word are in bold.

The predominance of conceptual associations is illustrated in Figure 6.2, which shows the number of each type of associate by output position. As noted above, apart from conceptual and erratic associations, all other types of associate decreased in frequency over later output positions. Form-based responses were particularly infrequent. The decrease in collocations and synonyms had already been shown using the linguistic-conceptual coding system, in which these three linguistic categories map onto the first five categories shown in Figure 6.2.

Figure 6.2

Count of Associate Type by Output Position



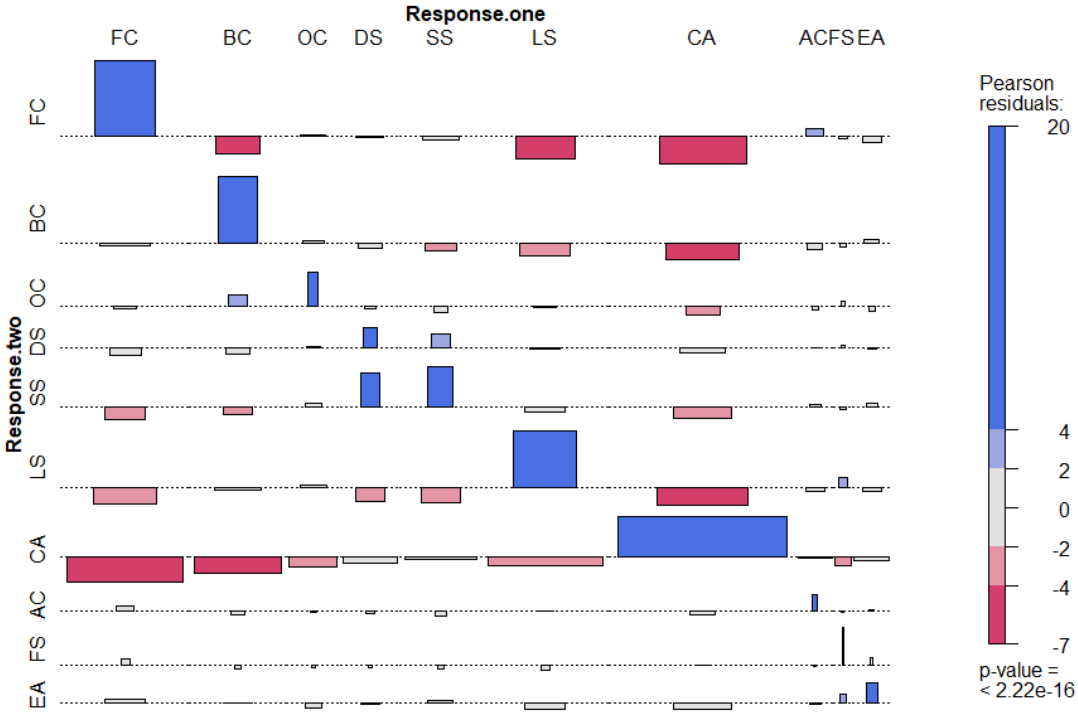
Note. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association.

6.3.2 Relationships Between Multiple Responses

One claim for the validity of the linguistic-conceptual coding system has been based on evidence of self-priming of the specific association types. This was examined using the meaning-, position-, and form-based subcategories, by performing a chi-square test to assess whether the specific type of association to the cue word evinced by response one was associated with response two type. This revealed a highly significant association ($\chi^2(81) = 1814.2$, *p.* < 0.00001), which was explored in more detail via an association plot, as shown in Figure 6.3. This shows that using this coding system produced a similar result: all cue types given in response two were of the same type as that given in response one significantly more frequently than would be expected by chance. Defining and specific synonyms were also likely to follow each other, suggesting that these two categories represent the same construct. The dissociations between collocation and conceptual and lexical set associates are also significant. The dissociation between lexical set and conceptual associations is somewhat more surprising, as these are both considered meaning-based responses.

Figure 6.3

Association Between Response One and Two



Note. Blue bars indicate higher numbers than expected and pink bars indicate lower numbers; deep hues indicate a significant difference. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association

6.4 Sensorimotor Ratings of Unique Associate Words

The analyses reported in the previous sections in this chapter suggest that the linguistic-conceptual classification system provides greater insight into word associations than the paradigmatic-syntagmatic classification. However, external evidence that the distinction between linguistic and conceptual associations may be a psychologically plausible construct would strengthen the claims made in this thesis.

To provide such a test, all response words coded as linguistic or conceptual were compared to the Lancaster Sensorimotor Norms compiled by Lynott et al. (2020). These norms, describe in more detail in the next section, are free to download from the Open Science Framework (http://osf.io/7emr6/). Briefly, they consist of 40,000 English words which were rated according to their strength of perceptual and action experience across 11 different dimensions. These norms will be used to test two main hypotheses. First, that response words coded as conceptually associated to the cue will have higher sensorimotor ratings than those coded as linguistically associated, indicating that conceptual associates are grounded by experience to a greater extent than linguistic associates. If this first hypothesis is supported, the second hypothesis is that this difference between linguistic and conceptual associates is not replicated between paradigmatic and syntagmatic associates. If paradigmatic associates are rated higher in sensorimotor strength than syntagmatic ones, this would indicate that this classification has equal psychological validity to the distinction between conceptual and linguistic associates.

In addition, given that conceptual associations coded as semantic field associates enter into a more abstract conceptual relation to the cue words than those coded as object or cue properties or as situational descriptors, semantic field associates are predicted to have the lowest sensorimotor ratings within the four types of conceptual associations. Likewise, if the coding of synonyms and antonyms as linguistic relations, rather than meaning-based or paradigmatic relations is valid, these associate words will have similar sensorimotor ratings to collocates.

It should be noted that this analysis is based on the response words given by participants, independent of the cue words. Thus, it is irrelevant whether the cue word was predicted to elicit predominantly linguistic or conceptual associates based on the USF norms. Moreover, this gives independence to the predictor variable, the type of association, as it is not a property of the response word in itself. The same associate word can enter into a linguistic and a conceptual association with two different cue words, for example, “bed” was coded as a forward collocation of the cue word “hospital” but a situational descriptor for the cue “dream”.

6.4.1 The Lancaster Sensorimotor Norms

The Lancaster Sensorimotor Norms (henceforth referred to as LSN) are a set of semantic norms for 40,000 English words created from participant ratings of each word according to 11 sensorimotor dimensions (Lynott et al., 2020). These norms were specifically created to test theories of embodied or grounded cognition, which is the theoretical framework that has guided the novel word association classification system developed here. In contrast to the concrete-abstract distinction made between words and concepts, they are multimodal or multidimensional ratings, so are less biased by visual perception. The largest set of concreteness ratings collected to date in English (also 40,000 words), by Brysbaert et al (2014), do not capture multimodal conceptual knowledge. Despite providing detailed instructions and examples to participants to rate words as more concrete to the extent they had more direct experience of their referents in any sensory modality, ratings indicated that most participants relied on sight and touch. For example, commonly experienced auditory stimuli tended to receive low concreteness ratings. This suggests that ratings based on concreteness may show similar variability to those based on imageability, which have recently been criticised (Dymarska et al., 2023).

Although sensorimotor experience account for ten of the experiential dimensions, the LSN also contains an interoceptive dimension. This is important as abstract concepts such as love have been shown to be grounded in interoceptive experience (Vigliocco et al., 2014). The LSN can potentially distinguish abstract concepts that are grounded in such direct experience from those that may be learned more through language, capturing the heterogenous nature of abstract concepts (Kiefer & Harpaintner, 2020). Interoceptive strength correlates negatively with concreteness ratings, yet predicts the concreteness effect in word processing, in which concrete words are processed more quickly than abstract ones (Connell et al., 2018). Thus, concepts that are more grounded in experience, whether concrete or abstract, should score higher on these dimensions than those that are learned more through language.

Experiential strength ratings were collected for each word/concept along six perceptual dimensions (the five senses plus interoception) and five action dimensions regarding the body part associated with the word (head, arm/hand, mouth, leg/foot, torso). Each participant rated a subset of the words according to either perceptual or action strength on a scale from 0 to 5, with an average of 18 and 19 ratings, respectively, per word. Inter-rater reliability (as calculated using Cronbach's alpha) was above .9 for all dimensions except head (.85) and torso action (.89).

Principal component analysis of the LSN (Lynott et al., 2020) showed that each dimension made a unique contribution to the variance between words, ranging from 12% for olfactory ratings to 43% for head action ratings. This shows these 11 dimensions all contribute to word or concept knowledge. However, ratings were highest for visual perception, with a mean rating of almost three along this dimension, indicative of the dominance of visual information in human perception (the only other dimension to receive a mean rating higher than two was head action). Moreover, although most concepts were multidimensional, vision was the dominant experiential dimension for over half (57%) of the words.

Lynott et al. (2020) compared various methods to reduce these 11 dimensions to the optimal single composite variable for use by subsequent researchers. They formed six different composite variables and judged them by their ability to predict response latency and accuracy on two lexical decision task data sets, which shared over 20,000 or 10,000 words with the LSN. In testing the LSN, they found that although all composite variables predicted response times and accuracy at above-chance levels, the best performance was obtained by Minkowski 3 distance reduction. Minkowski 3 distance is a metric calculated across all 11 dimensions in vector space, such that the dominant dimension always contributes to the composite variable, with other dimensions contributing to the extent they make a significant contribution (i.e., the closer their value is to that of the dominant dimension). In Minkowski 3 distance, contributing non-dominant dimensions are weakened compared to the dominant dimension, so that their influence within the composite variable is reduced.

They also tested the sensorimotor norms using this composite variable derived by Minkowski 3 distance, by creating three separate data sets: perceptual strength, action strength, and sensorimotor strength. They found that all three variables predicted lexical decision task performance, but that sensorimotor strength was a slightly better predictor.

In summary, the LSN constitute a large data set of words to compare with the word association data, thus providing high coverage. They are multidimensional ratings, guided by the theory that experience grounds conceptual knowledge, as claimed here. Moreover, the ratings have been shown to be reliable and the composite variable formed from the 11 dimensions has been demonstrated to predict word recognition. This makes the LSN an ideal set of norms to test the validity of distinguishing between associates produced by linguistic or conceptual processing.

6.4.2 Sensorimotor Ratings of Unique Associate Words

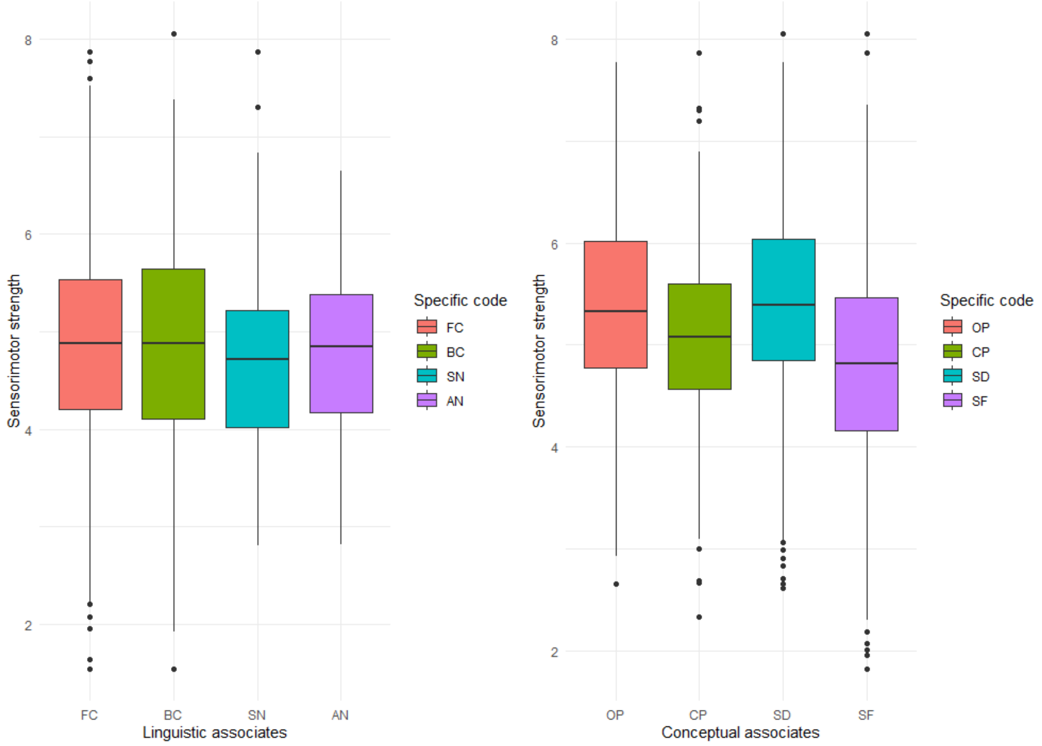
A new table was made of all the unique responses given to the cue words on the WAT. As the LSN do not contain plural nouns, all such responses were changed to the singular form (for example, all responses “friends” were changed to “friend”), unless they did not exist in singular form. Most such nouns did appear in the LSN, for example “goods” and “jeans”, although some did not, such as “supplies” and “credits” (end credits to a movie). The only exception was when this change altered the relation between the cue and associate, such as the single response of “trains” to “station”, which was coded as a situational descriptor, whereas all the responses of “train” were coded as backward collocations. No other changes were made to the WAT data, given that most relevant verb form changes are included in LSN (for example, “dance” and “dancing”, or “lose” and “lost”). This created a set of 5524 unique associations to the 80 cue words.

The unique word association data was joined with the LSN by response word, in R, adapting code provided in Winter (2020). This data set was then filtered to create a new data set including only those response words which appear in the LSN. In total, 5037 (91%) of the unique associations were matched to words in the LSN, of which 2452 were unique words (49%). To test the first hypothesis, that conceptual associates are rated higher on sensorimotor properties than linguistic ones, taxonomic and erratic associates were removed. The complete data set for this analysis consisted of 4322 unique associations and 2239 unique words (52%). Of these, 2781 (64%) of unique associations were conceptual associates, 57% of which were unique words. There were 1539 linguistic associates, of which 71% were unique words. These proportions are higher because some associate words were coded as linguistic and conceptual associates, depending on the cue word to which they were given, as explained above.

Figure 6.4 shows box-and-whisker plots of the sensorimotor strength ratings of associates within each specific conceptual and linguistic category. Although the majority of response words received an average rating between four and six across all eight associate types, the fairly large interquartile ranges and long whiskers show there was considerable variation around the median values, between ratings as low as two and up to the maximum rating of eight. Importantly, the median values for each specific category support the first hypothesis, being higher for conceptual associates than linguistic ones. In addition, semantic field associates are rated lowest within the conceptual associates.

Figure 6.4

Sensorimotor Ratings of Response Words in Linguistic and Conceptual Categories



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field

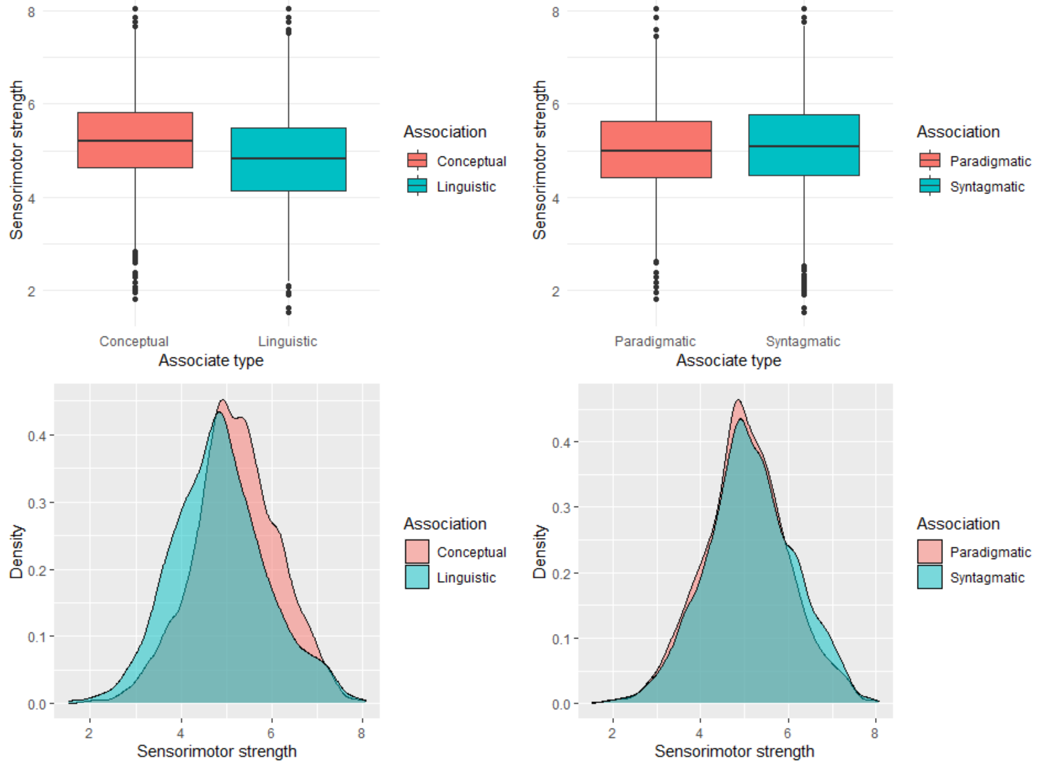
One-way ANOVAs were conducted followed by Tukey HSD post-hoc tests to control for multiple comparisons, separately on the sensorimotor ratings for response words coded as linguistic or conceptual associates and on all eight response categories. The results showed both the high similarity within the two general categories and the difference between the categories. The only statistically significant comparison within the linguistic categories was that synonym associate words were rated significantly lower on sensorimotor strength than the other types of associations. Conceptual associations showed greater within-category variation, with only object properties and situational descriptors showing no significant differences. Sensorimotor ratings for cue property words were different from these categories and from semantic field associates, which were significantly lower than all other response types. Between category comparisons showed that object property and situational descriptor associates were rated significantly higher than all four linguistic sub-categories. Importantly, synonyms were rated lower in sensorimotor strength than all four types of conceptual associate and both types of collocate. This further supports the classification of synonyms as linguistic associates, as discussed in Chapter 8.

As the first hypothesis was supported, the sensorimotor ratings for linguistic and conceptual associates were compared to those for paradigmatic and syntagmatic associates. As indicated by the previous analysis, and shown in Figure 6.5, conceptual associates were rated higher in sensorimotor strength than linguistic ones. In contrast, the difference between paradigmatic and syntagmatic associations is much smaller. In addition, contrary to expectation, this slight difference results from a tendency for syntagmatic associates to be rated higher on sensorimotor strength than paradigmatic ones.

Welch's independent *t*-tests were performed on both sets of data, as this test is more robust than Student's *t*-test (Delacre et al., 2017). The mean sensorimotor strength rating of conceptual associates was 5.21 (SD = 0.95), compared to 4.85 (SD = 1.04) for linguistic associates, which was a highly significant difference and a small to medium effect size, $t(2932) = 11.254$, $p < .0001$, Cohen's $d = 0.37$. Given the large sample size (number of unique associate words), the difference between paradigmatic and syntagmatic associates was also statistically significant ($t(2600) = -3.156$, $p = .0016$), despite the similarity of the mean for each group, 5.02 (SD = 0.96) and 5.12 (SD = 1.01) respectively. This was reflected in the very small effect size, Cohen's $d = 0.10$, which is generally considered an effect size of little practical importance (Cohen, 1992). As noted above, this difference is also in the opposite direction to expected.

Figure 6.5

Sensorimotor Ratings for Response Words According to Both Coding Systems



6.5 Discussion

In this chapter, I contrasted the linguistic-conceptual coding system developed and described in Chapters 4 and 5 (based on Santos et al., 2011) and the paradigmatic-syntagmatic coding system (Osgood et al., 1954) that has frequently been used in WA research (see Chapter 2, Section 2.2). I then compared the eight specific codes within this new system to the ten specific codes developed by Fitzpatrick (2007, 2009). Finally, the LSN (Lynott et al., 2020) was used as a measure of the multimodal content of the response words.

Firstly, it is of note that the high frequency cue words used here elicited a higher proportion of syntagmatic than paradigmatic associates. This is in contrast to the assumed response pattern (e.g., Deese, 1965) that has been found in previous WA studies with adult L1 speakers (e.g., Burke & Peters, 1986; Riegel & Riegel, 1964; Zareva, 2007), although other authors have reported similar findings (e.g., Hirsh & Tree, 2001; Nissen & Henriksen, 2006). This counters the assumption that paradigmatic associations dominate the structure of a well-organised mental lexicon and are in some way superior to syntagmatic associates (e.g., Deese, 1965; Sandgren, 2021).

More importantly, the patterns found for syntagmatic and paradigmatic associates did not parallel those found for linguistic and conceptual associates. Although syntagmatic cues tended to elicit associations of the predicted type, paradigmatic cues elicited similar proportions of both types of associates. In contrast, although more conceptual associates were produced overall, both linguistic and conceptual cues tended to elicit a first response of the predicted type of association. More importantly, whereas there was a tendency for participants not to produce linguistic associates after conceptual ones, the only clear pattern when using the paradigmatic-syntagmatic coding system was a tendency to produce syntagmatic associates, including producing three syntagmatic associates to both cue types.

Analysis of the dominant responses given to each cue revealed similar patterns. The dominant response words when considering only first responses were more frequently linguistic associates than conceptual ones, but when considering all three responses, this trend reversed, with conceptual associates being more frequent. No such pattern was found when responses were coded as paradigmatic and syntagmatic associates.

In sum, although conceptual associates were frequent across all output positions, there was a tendency for linguistic associates to be produced as the first response and for conceptual associates to dominate second and third responses (see also De Deyne et al., 2013). This pattern would seem to be partly driven by the lower likelihood of producing a linguistic associate after a conceptual one. However, using the paradigmatic-syntagmatic coding system the only tendency was for participants to produce more syntagmatic associates, a trend which was also contrary to that expected (Deese, 1965; Osgood et al., 1954).

Analysis of the specific codes showed similarities between the two systems. However, the proportion of responses in each specific category was more balanced across the categories in the new linguistic-conceptual system than when using the categories introduced by Fitzpatrick (2007), in which nearly half of all associates were coded as conceptual associations. This highlights the need to specify the construct of conceptual association in greater detail, as was achieved with the four specific codes within the conceptual associates used in the new system. In addition, the association between defining and specific synonyms as response one and two suggests that these responses reflect the same underlying construct, supporting their combination in one synonym category.

One possible explanation why the new system based on a distinction between conceptual and linguistic associates achieved a more even spread of associates across the categories than the system used by Fitzpatrick (2007) is that the cue words were very different. The data analysed here are responses to highly frequent words in the English language, as calculated based mostly on spoken corpora. In contrast, Fitzpatrick used words from the academic word list (Coxhead, 2000). These were abstract words which are likely to have been learned by participants in educational settings and thus through language rather than by grounding in experience. This may explain why forward collocates and defining synonyms were the most frequent response categories in that study. Using different cue words, Fitzpatrick and Izura (2011) also reported that the most frequent type of association produced were conceptual associates. The high proportion of conceptual associates indicates that this type of association has warranted the further investigation made in this thesis.

The analysis of responses using the Lancaster Sensorimotor Norms (Lynott et al., 2020) showed that conceptual associate words were rated higher on multiple multimodal dimensions of experiential strength compared to linguistic ones, providing further evidence that conceptual associates reflect multimodal processing to a greater extent than linguistic ones. There was no real difference between paradigmatic and syntagmatic associate words, as would be expected if this coding system captures language use (Osgood et al., 1954). This finding supports the psychological validity of the linguistic-conceptual distinction.

It should be noted that semantic field associates, which are assumed to capture more abstract semantic relations between words, had the lowest median score on experiential strength ratings of all eight associate types, further suggesting that the inclusion of this category within conceptual associates is problematic. It seems probable that this category is capturing semantic information about words and their concept referents that is learned through language to an equal or even greater extent than through direct, multimodal experiences. However, if responses coded as semantic field associates reflect linguistic knowledge, the true difference between cue and object properties and situational descriptors compared with linguistic associates is even greater than that shown in the analyses reported here.

Chapter 7. Testing the Linguistic-Conceptual Coding System with L2 Learners

The analyses reported in Chapter 6 suggest that the distinction between linguistic and conceptual responses on word association tasks (WATs) may be more psychologically valid than that made between paradigmatic and syntagmatic associates. Comparison of the specific linguistic and conceptual categories and the specific meaning-, position- and form-based categories created by Fitzpatrick (2007) revealed similar response patterns. However, the four conceptual categories captured distinctions not made in the one conceptual category used by Fitzpatrick, which accounted for most responses (some responses in Fitzpatrick's lexical set category would be considered conceptual under the new system). These findings regarding the WAs collected from L1 speakers suggest the three coding systems may reveal different insights into WAs produced by English language learners.

To the best of my knowledge, no previous study has analysed patterns in the types of associates produced by L2 learners across multiple responses. Moreover, no previous L2 WA study has compared different coding systems (see Chapter 2, Section 2.2.2, for a review of WA studies that have coded the responses of L2 learners). Therefore, the experiment reported in this chapter is again exploratory, with two main aims: (a) to uncover whether the linguistic-conceptual distinction reveals differences that are not indicated by the paradigmatic-syntagmatic distinction (Osgood et al., 1954); and (b) to discover whether there are systematic differences in the specific types of linguistic and conceptual associates produced by L2 learners compared with L1 speakers, and, if so, whether these are distinct from those found in specific types of meaning-, position-, and form-based responses (Fitzpatrick, 2006, 2007, 2009). These comparisons will enable assessment of whether the linguistic-conceptual coding system proposed in this thesis (based on Santos et al., 2011) can provide new insight into the L2 mental lexicon. The analyses reported in this chapter, combined with those reported in Chapter 6, also provide a detailed scrutiny of the information gained by coding WA data.

In this regard, it should be noted that L1 speakers produced more syntagmatic than paradigmatic associates to the high frequency cue words selected for these experiments. Although paradigmatic associates are assumed to dominate L1 adult WAs (e.g., Deese, 1965; Entwisle, 1966; Sandgren et al., 2021), Nissen and Henriksen (2006) also found participants produced mostly syntagmatic associates in their L1. Given the L1 data (see Chapter 6), L2 speaker responses will be more native-like if they produce more syntagmatic associates, in reverse of the assumed response pattern. In contrast, if the distinction between linguistic and conceptual associates is valid, L2 learners should produce a smaller proportion of linguistic associates than L1 speakers, reflecting their reduced knowledge of the English language. Accessing conceptual knowledge should be similar across both groups, so L2 learners may produce a similar proportion of conceptual associates as L1 speakers, although it is anticipated that L2 learners will produce more non- and other responses, in line with previous studies (e.g., Nissen & Henriksen, 2006; Norrby & Håkansson, 2007; Wolter, 2001).

Regarding the specific response categories, if conceptual responses dominate responses in the same way as with L1 speakers, most responses should be situational descriptors, based on experience of the referent, or, in the specific categories introduced by Fitzpatrick (2007), conceptual associates. Further, if all types of conceptual response are equally available to L2 learners and L1 speakers, the proportion of object and cue properties and semantic field associates should also be similar to that found with L1 speakers responding to the same cues. Alternatively, it could be that foreign words learned in a classroom context are associated with different types of conceptual knowledge, leading to differences in specific types of conceptual associations. This possibility is based on the results of Kim (2013), who analysed WA data from different groups of Korean learners using an adapted version of Fitzpatrick's (2006) coding system. Kim found that heritage learners of Korean in the US and learners in Korea produced more strong conceptual associates than non-heritage learners of Korean in the US. Within Fitzpatrick's (2007, 2009) coding system, any such difference may be captured by the distinction between conceptual associates and lexical set associates. If synonym associates are similar to conceptual and lexical set associates, L2 learners may produce similar proportions of synonyms to the L1 speakers. If the proportion of synonym associates and collocations are similarly reduced, this would support the claim that these relations reflect language knowledge.

7.1 Method

The experiment reported in this chapter followed a similar method to the one reported in Chapter 5, as described in detail below. The materials and procedure were approved by Swansea University Ethics Committee and permission was granted by the English department head at the university in Japan at which the experiment was conducted. All participants were Japanese learners of English as a foreign language (i.e., they had predominantly studied English in a classroom context). In addition to the WAT, a vocabulary test was also administered (see Section 7.1.3), so that the types of association produced by each participant could be correlated with an estimate of their English vocabulary knowledge.

7.1.1 Ethics Protocols

All participants were given a consent form, the WAT, and a vocabulary test. The consent form (provided in Japanese and English), explained the experimental tasks, why the research was being conducted, and the voluntary nature of participation, as well as providing contact details. After completing the WAT and the test, participants were given a debrief about the purpose of the experiment, including my email address. The consent form, WAT instructions, and debrief were translated into Japanese by an L1 Japanese language speaker and then checked by an L1 English speaker with professional Japanese-English translation experience. The consent form was provided in both Japanese and English, to ensure participants knew they could ask questions about anything they did not understand. As the experiment was conducted during English class periods, written informed consent was also obtained from the teachers.

7.1.2 Word Association Task

The WAT used cue words selected for the experiment reported in Chapter 5 (see Section 5.2.1). Although ten cue words of each specific linguistic and conceptual type were used in that experiment, this list was reduced to six cue words of each type for this experiment (see Appendix 10). This made a total of 48 cue words, of which 37 were in the first 1000 most frequent word family list created by Bauer and Nation (1993) and 37 were in the 1000 word list from the JACET List of

8000 Basic Words (Committee of revising JACET basic words, 2003), with 41 appearing within one of these two lists. All cue words were within the first 2000 word families and the first 2000 JACET Basic Words. In addition, every cue word except “elephant” was also on the New General Service List (NGSL; Browne et al., 2013). By category, six of the seven cue words that were on the 2000 word family and JACET word lists (i.e., were lower frequency words) were conceptual. Following coding, these cues were inspected to discover whether they elicited a higher proportion of no or other response, but no difference was found.

Cues were semi-randomly divided into two lists, with three cue words from each predicted response category on each list, so that each participant provided associates to 24 cue words (see Appendix 10). This was done to prevent fatigue and keep the time required by participants to complete the experiment within 15 minutes. The 24 cue words on each list were semi-randomly ordered, so that no two words from the same predicted response category occurred next to each other. Each list was then reverse-ordered to reduce any priming effects from previous cues, creating four different word lists. The lists were printed in one table on one side of A4 paper, with three spaces provided next to each cue word for participants to write their response words. The task was preceded by short and simple written instructions in Japanese and at the end of the task, participants were asked to circle their gender, first language(s), number of years of English study, and study abroad experience.

7.1.3 Vocabulary Size Test

To test whether the coded WA data was sensitive to differences in participants’ word knowledge I decided to administer a word knowledge test. Although all participants had similar English ability levels, students were streamed at the university based on their reading and listening ability and grammatical knowledge, as well as vocabulary knowledge. Therefore, vocabulary knowledge should vary across participants, and this might be reflected in the types of responses produced on the WAT. Should there be a systematic variation between WA types and vocabulary test scores, this might reveal differences between the three coding systems.

The Vocabulary Size Test (VST; Nation & Beglar, 2007), which is a freely available test for researchers (https://www.wgtn.ac.nz/lals/resources/paul-nationsresources/vocabulary-tests), was selected for 3 main reasons: (a) The VST is a comprehensive test intended to measure how many words a learner knows (i.e., their total vocabulary size); (b) it is based on word frequency, determined by word family lists created by Bauer and Nation (1993) from the British National Corpus (BNC; Davies, 2004); (c) its reliability has been validated by Rasch analysis (Beglar, 2010). The Japanese-English version of the VST was used, as research with Russian learners of English found that test-takers performed significantly better on the bilingual than the monolingual version, suggesting this format should be preferred (Elgort, 2013). This version of the VST has also been validated by Rasch analysis (Derrah & Rowe, 2015).

An additional advantage of using the VST was that it was possible to administer only those items in the higher frequency word bands. In this reduced format, the test is very simple and quick to complete, which was felt to be important after asking participants to provide 72 English words in response to 24 cues on the WAT, given that the upper limit for fatigue on this type of task has been estimated to be about 100 words (Fitzpatrick & Thwaites, 2020). Thus, participants completed the first four levels (40 items) from the English-Japanese version of the VST (Nation & Beglar, 2007), which tests knowledge of the first four 1000 word family levels. This limit was chosen based on administering the first 80 items to three student volunteers whose English ability was upper intermediate to advanced level, and also considering Nation & Beglar's (2007) observation that "under-graduate non-native speakers successfully coping with study at an English speaking university have a vocabulary of around 5,000-6,000 word families" (p. 12). Learners classified as having pre-intermediate to intermediate level English ability in Japan are unlikely to know this many word families, so 4,000 should be a sufficient number of test items to detect differences in word knowledge in this group.

In addition, due to the test's multiple-choice format, the potential distortion of test results due to guessing has been identified as an issue. In a study involving over 3000 Japanese university students, McLean et al. (2015) looked at the effects of guessing by examining scores at the 8000 word family level. They found that at this level, for the highest proficiency learners, over 40% of correct answers could be attributed to guessing. For students with an English ability level slightly above the average in Japan, this figure rose to 85%. A recent study that used a think aloud protocol with mature Japanese learners of English further confirmed that guessing on the VST inflated scores (Asquith, 2022).

One potential drawback of using the VST is that it is intended to measure written receptive knowledge of words, whereas WA is a productive task. However, in a study with Japanese university students, Webb (2008) showed that receptive knowledge of English vocabulary is predictive of productive knowledge. Nonetheless, this factor, in addition to any guessing, may reduce any correlations between VST scores and WA behaviour.

7.1.4 Participants

The experiment was completed by 108 Japanese university students, who all provided written informed consent prior to participating. Participation was voluntary and was not remunerated. All participants were in their first year at the same university and were studying English as a compulsory subject in addition to their major course of study. The students had completed an English placement test upon entrance at the university and had received similar scores. Their English proficiency was pre-intermediate level ranging towards intermediate level in the highest proficiency students. For reasons explained below (Section 7.1.7), data from five participants were discarded prior to analysis, reducing the final sample to 104 participants. Of these 46 were female and 58 male; 38 had studied English for 6 years or fewer; 58 had studied for 7 to 11 years; and 7 had studied English for more than 11 years. Ten participants had studied English abroad, eight for less than 6 months and only one for more than one year.

The WAT was administered during English classes, with permission from the teachers. As a result, data was also collected from 11 non-Japanese participants, but was not included in the analyses reported in this chapter (excluding these students from participating would have been perceived as discriminatory). In addition, six students did not attempt to complete the WAT (they wrote responses to just the first few cue words). This was considered a withdrawal of consent and their data was discarded.

7.1.5 Procedure

Participants completed the tasks at the beginning of an English class, on the date decided by their teacher. Participants were asked to read and sign an informed consent form first (provided in Japanese and English). In explaining the tasks to the students, I stressed that they could stop participating at any point during the tasks and that this experiment had no relation to their studies.

The instructions given for the WAT were to write three responses to each cue word as quickly as possible. They were told that it was not a test and there were no correct answers, they should just write the first words that came to mind. They were told to ignore any cue word they did not understand on the WAT and any word on the VST they did not know.

All materials were administered stapled together, such that the consent form was on top, followed by the WAT, and then the four levels of the VST. The presentation order of the WAT and VST was not counter-balanced, to ensure the words on the VST did not influence any responses given on the WAT. I monitored the students during the task, to ensure they completed the WAT and the VST without consulting a classmate or a dictionary. After each student had completed the task, I took their materials and gave them a short debrief about the aims of the experiment. No time limit was set to complete the WAT and the VST, but all participants completed both within 15 minutes.

7.1.6 Analyses

All materials were given an alphanumerical code to distinguish between classes and individual participants. Consent forms were then detached from the WAT and VST, to anonymise materials prior to data entry. The data from one participant were discarded as they completed the WAT in Japanese. Data from a further four participants were removed because the number of non-responses was more than three standard deviations above the mean (76 participants [70%] gave three responses to every cue word). As most of the non-responses by these four participants resulted from a failure to provide two or three responses to many cue words, rather than a failure to provide any response to some cue words, most of their data would not have contributed to the key analyses examining the relations between response words. Of the final 104 participants, 50 provided responses to the first set of cue words and 54 to the second set. However, as analyses are based on cue word type rather than cue word, this does not impact the analyses or interpretation of the results. Analyses including the VST results (Section 7.2.3) were conducted on 102 participants, as two participants failed to notice that the VST was printed double-sided.

Analyses were conducted in RStudio (RStudio Team, 2020), run on R (R Core Team, 2021), using the tidyverse (Wickham et al. (2019) and vcd packages (Meyer et al., 2006; 2022). Coding was based on various sources, notably Levshina (2015). Figures were created using ggplot2 (Wickham, 2016), following Healy (2019), and used the wesanderson package (Ram & Wickham, 2018) for bar charts.

7.1.7 Coding of Responses

In total, participants were asked to produce 7488 associations to the 48 cue words. The three responses provided by each participant to each cue were entered into a spreadsheet for analysis. Spelling mistakes were corrected when the intended word was evident based on knowledge of common mistakes made by Japanese learners of English. This group of learners find it difficult to distinguish between certain consonants, such as l and r and also may confuse b and d in writing. Thus, for example, “liver” in response to the cue word “swim” was corrected to “river”, and “deautiful” was corrected to “beautiful”. They may also use the incorrect vowel in spelling, especially when the vowel is part of a blend. For example, the response “cheep” to the cue “restaurant” was corrected to “cheap”. However, when the mistake was ambiguous it was coded as a spelling mistake in all systems (i.e., as an erratic or other response), such as the response “prace” to the cue “ticket”, in which it is unclear whether the intended associate was “price” or “place”. Responses for which the intended word could not be discerned, or which were written in Japanese characters were also coded as spelling mistakes. In total, 56 responses were coded as spelling mistakes (less than 1% of responses).

All responses were coded according to association type using three different coding systems (as in Chapters 5 and 6). First, they were coded using the new coding system proposed in Chapter 4, based on the one introduced by Santos et al. (2011). Applying this system, responses were coded as linguistic, conceptual, taxonomic, or other responses, including the specific subcategories of linguistic and conceptual associates. They were then coded as paradigmatic, syntagmatic, or other responses (based on Osgood et al., 1954). Finally, responses were coded as meaning-, position-, form-based, and erratic responses, applying the specific subcategories created by Fitzpatrick (2007; see Chapter 6, Section 6.4).

In Section 7.2 the results of the comparison of the linguistic-conceptual system and the paradigmatic-syntagmatic system are reported. The results of these analyses are also correlated with participants' scores on the first four levels of the VST. Section 7.3 reports the comparison of the eight subcategories used in the linguistic-conceptual system with the ten subcategories of the meaning-, position-, and form-based system developed by Fitzpatrick (2007). As noted above, given the lack of previous studies comparing multiple associations produced by L2 speakers or comparing different coding systems, analyses are exploratory rather than hypothesis-driven, so the emphasis is on descriptive data and visual representation of the findings, rather than inferential analyses.

As in the previous experiments, some associates reflected popular media. However, in this instance, most of these reflected Japanese culture that may not be familiar to many people living outside Japan, such as the associate "app" to the cue word "line" (Line is a popular app for communication in Japan, similar to WhatsApp). In addition, some responses showed influence from Japanese language, such as the response "live" to the cue word "ticket", given because the word "live" is used to refer to a concert in Japanese, as a loan word from the English expression "live music". Such responses were initially coded in two different ways: according to the intended meaning, such that "ticket: live" was coded as a backward collocation, the same as "concert", and as Japanese language or culture based, as a type of erratic or other response. However, given that responses based on English-speaking culture were not coded as other responses in the previous experiments, it was decided to analyse these responses according to the intended meaning. This issue is discussed in more detail in Section 7.4. It should be noted that such associations constituted just 1.5% of responses, so any influence on the interpretation of the WA data will be minor.

7.2 Results of the Comparison of the General Coding Systems

For these analyses, responses were coded as linguistic, conceptual, or taxonomic associates and as paradigmatic or syntagmatic associates. The proportions of each response type are presented, followed by relationships between multiple responses, and finally the correlation with scores on the VST.

In both systems, an erratic or other association category was also used, applied to responses that were idiosyncratic or had no meaningful connection to the cue word (e.g., "church: hill", "blood: people"), responses that involved the addition or subtraction of a regular affix or a change in verb tense (e.g., "cloud: cloudy", "telephone: phone", "have: had"), responses that were similar in sound or form only to the cue word (e.g., "chance: change"), responses based on a misunderstanding of the cue word (e.g., "root: journey"), responses that were connected to previous response words rather than the cue word (e.g., in response to the cue word "social", the words "network-computer-media" were provided by one participant, of which "computer" was coded as erratic), spelling mistakes that could not be corrected, and instances in which no response was given.

7.2.1 Associate Types

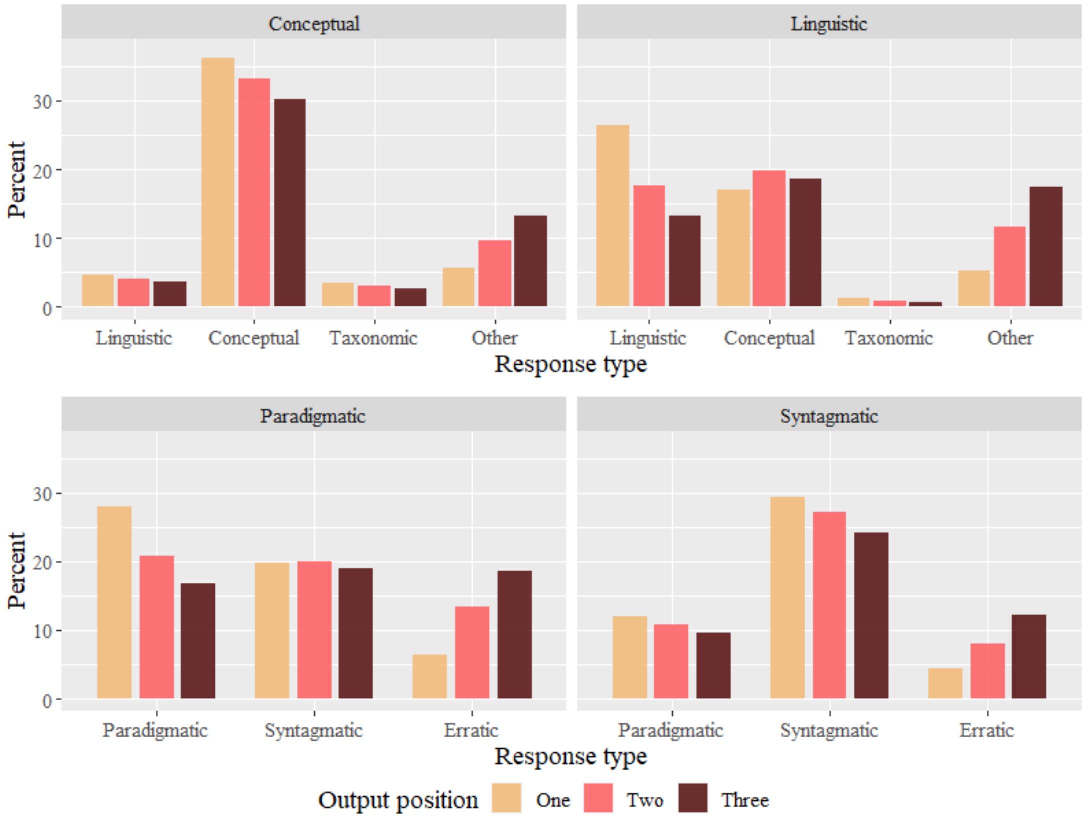
Figure 7.1 shows the response types across all three responses, when coded as linguistic, conceptual, or taxonomic, and paradigmatic or syntagmatic. There were 24 linguistic and 24 conceptual cue words but 26 paradigmatic and 22 syntagmatic cues, so the percentage of responses given to paradigmatic cues at each position equals 54% compared with 46% to syntagmatic cues.

Participants produced predominantly conceptual responses to the cue words predicted to elicit this type of association, with very few linguistic responses. Although they also tended to produce linguistic responses to linguistic cues as the first response, a large number of conceptual responses were given to cue words predicted to elicit linguistic associations, which became the dominant response type for second and third responses. Participants also tended to produce paradigmatic and syntagmatic responses to the cue words predicted to elicit each type of associate, although they produced more syntagmatic responses overall. Apart from the greater proportion of other associations, the response patterns for this group, when coded as paradigmatic and syntagmatic, were very similar to the L1 speaker data (see Chapter 6, Figure 6.1).

It should be noted that participants produced many responses that were coded as other, with such responses accounting for nearly 11% of all first responses and 18.5% of all third responses. This indicates that participants found it quite difficult to think of three words related to each cue word, even when they understood the cue. In addition, participants produced similar proportions of other responses to linguistic and conceptual cue words as the first and second response, but produced more non-target responses to linguistic cue words as the third response. In contrast, participants consistently produced more erratic responses to paradigmatic cue words than syntagmatic ones.

Figure 7.1

General Response Types as a Function of Cue Type



Note. Bar charts show the percent of response types as a function of the cue word type and output position, using the two different coding systems. For each output position, the responses across both cue types (conceptual and linguistic in the top panels and paradigmatic and syntagmatic in the bottom panels) sum to 100 per cent.

Overall, the influence of the cue words on the response type was greater when they were coded as linguistic or conceptual associates, despite the English language learners' reliance on conceptual knowledge to produce responses on the WAT. Learners produced substantially fewer linguistic responses and those given tended to be the first response to linguistic cue words. They also gave fewer linguistic associates to both types of cue word than L1 speakers (see Chapter 6, Figure 6.1), although they produced a similar number of conceptual associates. Thus, coding responses on the WATs as conceptual and linguistic associates discriminated between L1 and L2 English speakers, whereas coding them as paradigmatic and syntagmatic associates did not.

7.2.2 Relationships Between Multiple Responses

In the previous experiment, it was also shown that participants were less likely to produce a linguistic associate after a conceptual associate. As shown in Table 7.1, this was also true for the English language learners. Although the pattern is less noticeable due to the dominant tendency to produce conceptual associates, it is apparent in responses to linguistic cue words. In particular, this group hardly ever produced more than one linguistic associate in response to a conceptual cue word.

Table 7.1

Linguistic and Conceptual Responses as a Function of Cue Word Type

Response one	Response two	Response three	Cue word type		Difference from equal distribution		Combined difference
			Linguistic	Conceptual	Ling	Conc	
Linguistic	Linguistic	Linguistic	108	8	33	-74	-41
		Conceptual	88	9	13	-73	-60
	Conceptual	Linguistic	57	2	-18	-80	-98
		Conceptual	111	45	36	-37	-1
Conceptual	Conceptual	Conceptual	100	492	25	410	435
		Linguistic	49	53	-26	-29	-55
	Linguistic	Conceptual	50	42	-25	-40	-65
		Linguistic	37	7	-38	-75	-113
		TOTAL	591	658	75	82	

When coded as paradigmatic and syntagmatic, the main trend was for participants to give three responses of the type predicted to be elicited by the cue word, as shown in Table 7.2. Although they produced more syntagmatic associates overall, similarly to the L1 speaker group (see Chapter 6), there was no trend for syntagmatic associates to be less likely to follow paradigmatic responses. In addition, the learner group did not produce proportionately more syntagmatic and fewer paradigmatic associates compared with L1 speakers; there was no evidence of a syntagmatic-paradigmatic shift.

Table 7.2

Paradigmatic and Syntagmatic Responses as a Function of Cue Word Type

Response			Cue word type		Difference from chance		Total difference
One	Two	Three	Paradigmatic	Syntagmatic	Para	Syn	
Paradigmatic	Paradigmatic	Para	178	72	87	-15	72
		Syn	109	45	18	-42	-24
	Syntagmatic	Para	61	24	-30	-63	-93
		Syn	116	84	25	-3	22
Syntagmatic	Syntagmatic	Syn	107	300	17	213	230
		Para	52	75	-38	-12	-50
	Paradigmatic	Syn	40	64	-50	-23	-73
		Para	61	32	-29	-55	-84
		TOTAL	724	696	90	87	

Note. Para = paradigmatic, Syn = syntagmatic.

7.2.3 Associate Types and Vocabulary Knowledge

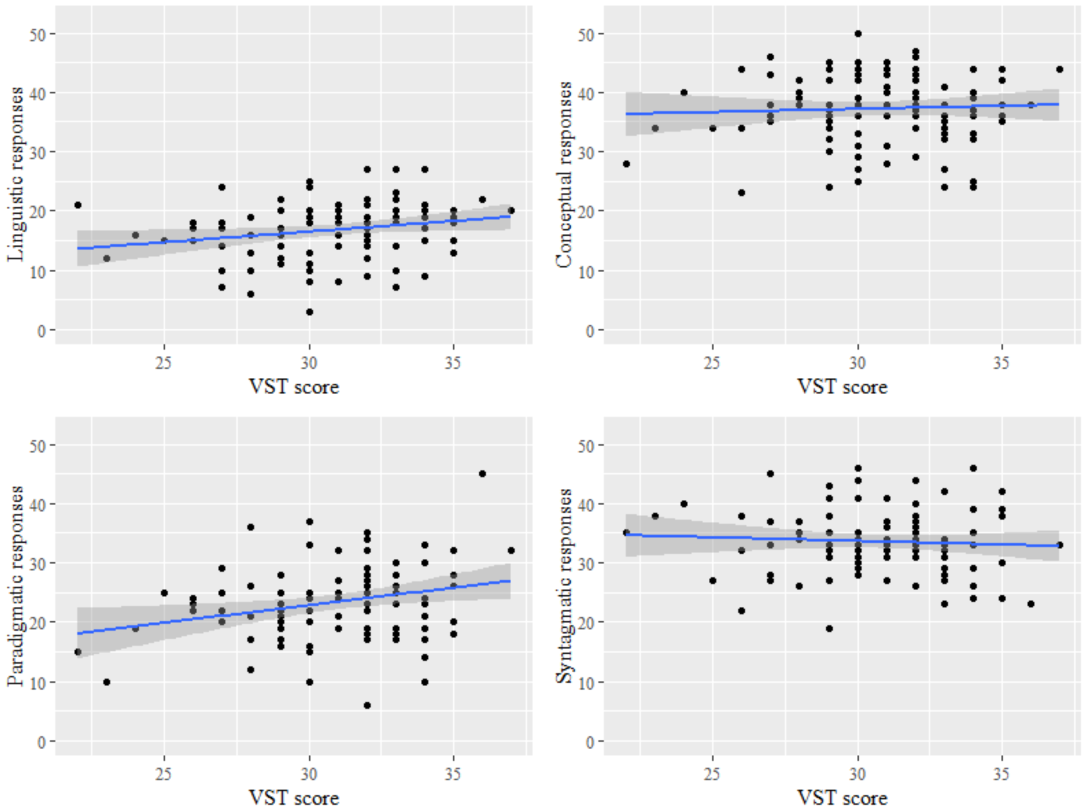
Next, for each participant, the number of associates given of each general response type was correlated with their VST score. The mean score on the first four levels of the VST was 30.83 (standard deviation 2.82). This suggests that most of the participants had similar vocabulary knowledge in English. Scores were correlated with the number of responses given by each participant, as shown in Figure 7.2.

The most frequently given response types in each coding system, conceptual and syntagmatic, showed no relation to VST scores. However, as predicted, the percentage of linguistic and paradigmatic associations both showed a modest positive correlation with VST scores. Kendall rank correlation analyses were conducted as many participants had the same VST scores and gave the same number of the different types of response. This showed that the correlation with linguistic responses was highly significant ($\tau_b = 0.180$, *p*. = 0.01) and that with paradigmatic responses was also significant ($\tau_b = 0.140$, *p*. = 0.05). As shown in Appendix 11, responses coded as other or erratic showed a slight negative correlation with VST scores, although this was not significant ($\tau_b = -0.110$, *p*. = 0.1).

Although the correlations are modest, the correlation between linguistic responses and VST score is notable, given that (as reported below) 68% of linguistic associations were collocations and therefore, were coded as syntagmatic in the traditional coding system. In addition, correlations with the general coding categories used by Fitzpatrick (2007; see Appendix 11), showed that the correlation between collocations and VST scores was significant (τ_b = 0.160, *p*. = 0.03), but slightly weaker than that with linguistic responses, indicating that synonym and antonym responses contributed to the correlation beyond collocation responses. It should also be noted that although paradigmatic responses showed a positive association with increased VST scores, it is difficult to interpret this result as a syntagmatic-paradigmatic shift with increased proficiency, given that the L1 speaker group gave a higher number of syntagmatic associates to the same cue words.

Figure 7.2

Association Between General Response Type and VST Score



Note. The upper plots show the number of responses per participant coded as linguistic and conceptual and the lower plots show the number of responses coded as paradigmatic and syntagmatic. The blue line shows the best fit and the shaded area represents 95% confidence intervals. Each participant provided 72 responses.

7.3 Results of the Comparison Using Specific Categories

The eight specific categories within the linguistic-conceptual coding system (linguistic categories: forward collocation, backward collocation, synonym, and antonym; conceptual categories: stereotypical object property, stereotypical cue property, situational descriptor, and semantic field) were compared with the nine meaning-, position-, and form-based categories Fitzpatrick (2007) proposed (position-based categories: forward collocation, backward collocation, other collocation; meaning-based categories: defining synonym, specific synonym, lexical set, conceptual association; form-based categories: affix change, form similarity). As participants gave many evaluative responses to cue words (e.g., scene-beautiful), these were separated from stereotypical cue property in the analysis. Both systems incorporate an erratic or other response category, for responses that have an idiosyncratic (non-discernible) relation to the cue word, spelling mistakes, chained responses, and non-responses. Note that in the linguistic-conceptual system, the other response category also includes affix changes and responses based on form similarity, including mistaking the cue word. More detailed information about the responses coded as "other response" is provided in Appendix 12.

7.3.1 Associate Types

Tables 7.3 to 7.6 show the number of each type of response given to each type of cue for both coding systems, for the first response only and across all three responses. Table 7.3 shows the first response given, as coded using the linguistic-conceptual system. Responses tended to be of the specific type predicted for all eight specific cue word types within this system (as explained in Chapter 5, some of the cue words chosen to elicit semantic field associates elicited situational descriptors as dominant associates).

Overall, the subcategorization of linguistic and conceptual associations shows that learners produced more situational descriptor associates compared to other types of response and that this type of response was given to all cue word types. If recombined, cue and evaluative properties were the second most frequently produced response type, at nearly 14% of all responses. The most frequently given type of linguistic associate were forward collocations, which were elicited by synonym and antonym cues as well as this cue type.

Table 7.3

Cue Word Response Categories for Response One (Linguistic-Conceptual)

Response Category	Cue word category type								Total	
	Linguistic				Conceptual					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
FC	**149**	1	49	37	8	11	13	15	283	0.113
BC	4	**146**	28	6	1	8	15	6	214	0.086
SN	29	6	**94**	44	15	12	5	6	211	0.084
AN	0	0	3	**64**	0	1	0	0	68	0.027
OP	0	0	0	27	**124**	4	0	0	155	0.062
CP	5	17	0	1	24	**122**	7	31	207	0.083
EP	18	14	11	16	12	**32**	9	21	133	0.053
SD	39	58	35	46	42	55	**212**	102	589	0.236
SF	34	18	55	31	22	12	7	**66**	245	0.099
TX	3	26	1	2	0	28	13	45	118	0.047
OR	31	26	36	38	64	27	31	20	273	0.109
Linguistic	182	153	174	151	24	32	33	27	777	0.311
Conceptual	96	107	101	121	224	225	235	220	1329	0.533
Taxonomic	3	26	1	2	0	28	13	45	118	0.047
Other	31	26	36	38	64	27	31	20	273	0.109

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response, No. = number, Prop. = proportion. The number of responses that are of the same category type as the cue word are in bold.

The response trends are similar over all three response positions, although the proportion of other responses nearly doubled, as can be seen in Table 7.4. Nevertheless, the proportion of situational descriptors, the dominant response type, continued to account for nearly 25% of all responses. The proportion of all four types of linguistic associate decreased, but within the conceptual categories, only object properties decreased to a similar extent. Although synonym and antonym cue words elicited many forward collocations, all four types of linguistic associate elicited a large number of situational descriptor and semantic field associates. In contrast, few linguistic associates were given to any of the specific types of conceptual associate. Although semantic field cue words elicited many taxonomic associates, closer inspection of the type of response given to each individual cue word showed that this was due to the cue word “family”, which elicited 122 of the 146 responses (84%).

Table 7.4

Cue Word Response Categories Across all Three Responses (Linguistic-Conceptual)

Response Category	Cue word category type								Total	
	Linguistic				Conceptual					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
FC	**339**	7	117	104	35	35	28	31	696	0.093
BC	13	**296**	62	17	9	24	48	20	489	0.065
SN	60	19	**181**	89	23	23	11	20	426	0.057
AN	2	0	8	**117**	0	1	0	0	128	0.017
OP	0	0	0	79	**253**	9	0	0	341	0.046
CP	38	54	0	4	60	**321**	35	68	580	0.077
EP	40	39	45	52	32	**73**	48	48	377	0.050
SD	149	197	89	131	160	204	**585**	280	1795	0.240
SF	113	77	167	114	62	40	25	**184**	782	0.105
TX	10	54	3	2	1	63	22	146	301	0.040
OR	172	193	264	227	301	143	134	139	1573	0.209
Linguistic	414	322	368	327	67	83	87	71	1739	0.232
Conceptual	340	367	301	380	567	647	693	580	3875	0.518
Taxonomic	10	54	3	2	1	63	22	146	301	0.040
Other	172	193	264	227	301	143	134	139	1573	0.209

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response, No. = number, Prop. = proportion. The number of responses that are of the same category type as the cue word are in bold.

In the meaning-, position-, and form-based system used by Fitzpatrick (2007), the first response also tended to be of the type predicted by the cue word. Note that, within this system, 30 of the cue words were predicted to elicit lexical set and conceptual associate responses, as these categories included all four specific conceptual categories in the linguistic-conceptual coding system as well as antonym cue words. This in part explains why over 40% of responses were conceptual associates, but it is also clear that many conceptual associations were given to all cue word types, whereas the only other cue word type that elicited many lexical set associates were those predicted to elicit conceptual associations.

The tendency to produce conceptual associates is even more pronounced over all three responses, with this being the dominant response type for all cue word types except forward collocations, which still elicited almost as many conceptual associates as collocates. Although form similarity responses were almost exclusively given to lexical set cue words, this is due to this category containing all the cue words that were most frequently mistaken (predominantly root for route, but also arrive for alive, quiet for quite, and cloud for crowd). Therefore, this should be interpreted as a chance effect rather than an interesting pattern.

Table 7.5

Cue Word Response Categories for Response One (Position-Meaning)

Response Category	Cue word concept type						TOTAL	
	Position-based		Meaning-based					
	FC (6)	BC (6)	DS (2)	SS (4)	LS (15)	CA (15)	No.	Prop.
FC	**149**	1	6	23	31	17	227	0.091
BC	4	**145**	10	7	21	14	201	0.081
OC	**0**	**1**	22	9	21	16	69	0.028
DS	8	2	**12**	12	21	5	60	0.024
SS	21	4	10	**60**	36	20	151	0.060
LS	5	31	5	3	**286**	123	453	0.181
CA	94	102	23	74	257	**512**	1062	0.426
AC	11	2	2	6	10	15	46	0.018
FS	0	2	3	0	39	1	45	0.018
EA	20	22	11	14	60	55	182	0.073
Position	153	147	38	39	73	47	497	0.200
Meaning	128	139	50	149	600	660	1726	0.691
Form	11	4	5	6	49	16	91	0.036
Erratic	20	22	11	14	60	55	182	0.073

Note. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association, No. = number, Prop. = proportion. The number of responses that are of the same category or concept type as the cue word are in bold.

As noted in the previous chapter, position- and meaning-based responses in this system do not coincide exactly with syntagmatic and paradigmatic responses. This is why, despite tending to give syntagmatic responses, participants game mostly meaning-based responses in this system. Almost 70% of the first responses given were meaning-based and although this figure drops to 63% over all three responses, this decrease is due to the increase in erratic responses, which caused all other specific response types except conceptual associates to decline.

Table 7.6

Cue Word Response Categories Across all Three Responses (Position-Meaning)

Response Category	Cue word concept type						TOTAL	
	Position-based		Meaning-based					
	FC (6)	BC (6)	DS (2)	SS (4)	LS (15)	CA (15)	No.	Prop.
FC	**335**	5	18	50	90	58	556	0.074
BC	13	**293**	27	20	64	46	463	0.062
OC	**4**	**5**	40	24	59	34	166	0.022
DS	18	6	**20**	22	35	10	111	0.015
SS	42	13	22	**117**	72	49	315	0.042
LS	26	67	13	10	**628**	341	1085	0.145
CA	326	354	70	219	832	**1418**	3219	0.430
AC	21	5	6	16	25	25	98	0.013
FS	0	5	4	0	116	1	126	0.017
EA	151	183	92	146	425	352	1349	0.180
Position	352	303	85	94	213	138	1185	0.158
Meaning	412	440	125	368	1567	1818	4730	0.632
Form	21	10	10	16	141	26	224	0.030
Erratic	151	183	92	146	425	352	1349	0.180

Note. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association, No. = number, Prop. = proportion. The number of responses that are of the same category or concept type as the cue word are in bold.

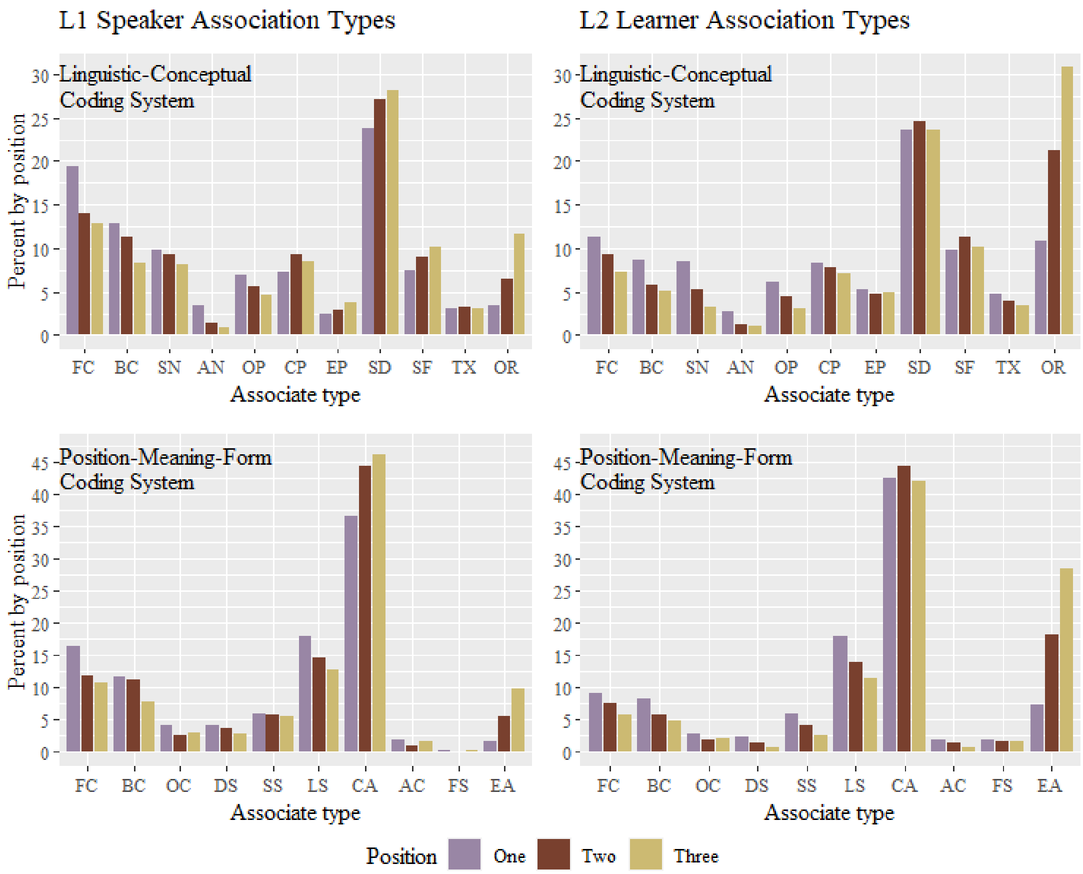
7.3.2 Relationships Between Multiple Responses

To better understand the numbers presented in the previous four tables, the proportion of associates in each output position (one, two, and three) is shown for both coding frameworks in Figure 7.3. The L1 speaker data (reported in Chapters 5 and 6) for the subset of 48 cue words used in this experiment is also shown for comparison. The percentage of each type of response is shown rather than the actual numbers as there were three more L2 speaker participants than L1 speakers. Note that the display size for both coding systems is maximised for clarity, which means that the y-axes are not on the same scale. The range for the specific categories within the linguistic-conceptual coding system is 0-30%, whereas for the specific categories within the position-meaning-based coding system it is 0-45%.

As noted above, the L2 learners produced a much larger percentage of other or erratic associations than the L1 speakers. However, L2 learners tended to produce similar proportions of conceptual associations, especially for the first response, and slightly more semantic field and evaluative properties. They produced notably fewer linguistic responses of all types, except for a similar number of antonyms (although these were infrequently given by both groups). Interestingly, the L2 learners also produced slightly more taxonomic responses than L1 speakers, further suggesting that taxonomic associations are not a dominant associative feature of a well-developed mental lexicon.

Figure 7.3

Count of Associate Type by Output Position for L1 and L2 Participants



Notes. Linguistic associates: FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym; Conceptual associates: OP = stereotypical object property, CP = stereotypical cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field; TX = taxonomic, OR = other response. Position-based associates: FC = forward collocation, BC = backward collocation, OC = other collocation; Meaning-based associates: DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association; Form-based associates: AC = affix change, FS = form similarity; EA = erratic association. The percentage of response types sums to 100 for each output position. To maximise display size, the y-axis extends to 30% for the conceptual-linguistic coding system but to 45% for the position-meaning-form coding system.

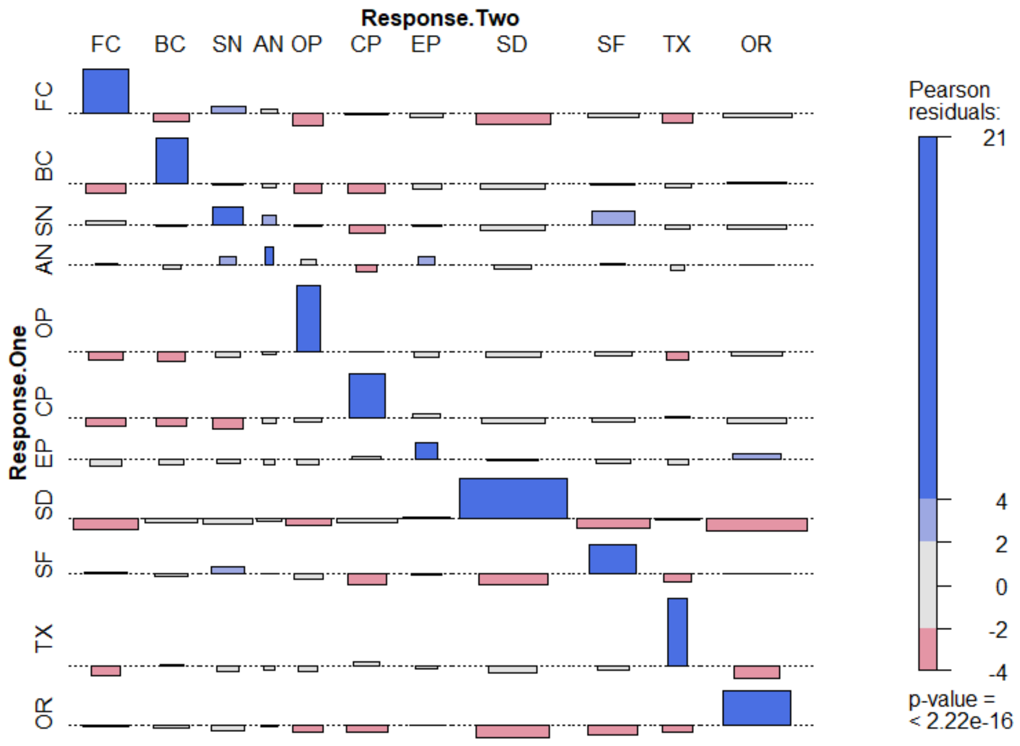
These differences are similarly captured in the specific categories of the meaning-, position-, and form-based coding system, albeit with a much higher proportion of conceptual associations compared to all other categories in both groups. The L2 Learners did not produce more affix change responses than the L1 speakers and the increased proportion of form similarity responses is due to the mistaken cue words being included in this category, indicating that, at least for familiar words, learners' lexical knowledge is focused on meaning not form. Importantly, learners produced similar numbers of lexical set and conceptual associations as L1 speakers, but fewer associates of all three collocate and both synonym types (except for specific synonyms as the first response), further supporting the claim that synonyms are part of linguistic knowledge and are qualitatively different to conceptual knowledge.

Next, the association between response one and response two was assessed for both coding systems, by performing chi-square tests to assess whether the specific type of association to the cue word evinced by response one was associated with response two type. This revealed a highly significant association for both the linguistic-conceptual ($\chi^2(100) = 2242.9$, $p. < 0.00001$) and the meaning-, position-, and form-based coding system ($\chi^2(81) = 2855.7$, $p. < 0.00001$), which was explored in more detail via association plots for both systems, as shown in Figures 7.4 and 7.5. These plots show that participants tended to produce two associates of the same specific type and that this relationship was captured by both systems. However, there were also some differences between the coding systems.

In the linguistic-conceptual coding system (Figure 7.4) the width of the situational descriptor bar reflects the prevalence of this associate type. In addition, there are smaller, but significant, associations between synonym and antonym responses and between synonym and semantic field responses, as well as a slight trend for a synonym to be produced after a forward collocation. The height of the bar representing taxonomic associates is due to most of this type of associate being produced to one cue word, as noted previously. There is also a slight tendency for evaluative property associates to be followed by an other associate, suggesting that producing evaluative properties may have been a strategy used by L2 learners when they found it difficult to think of an associative word in response to a cue.

Figure 7.4

Association Between Response One and Two (Linguistic-Conceptual)

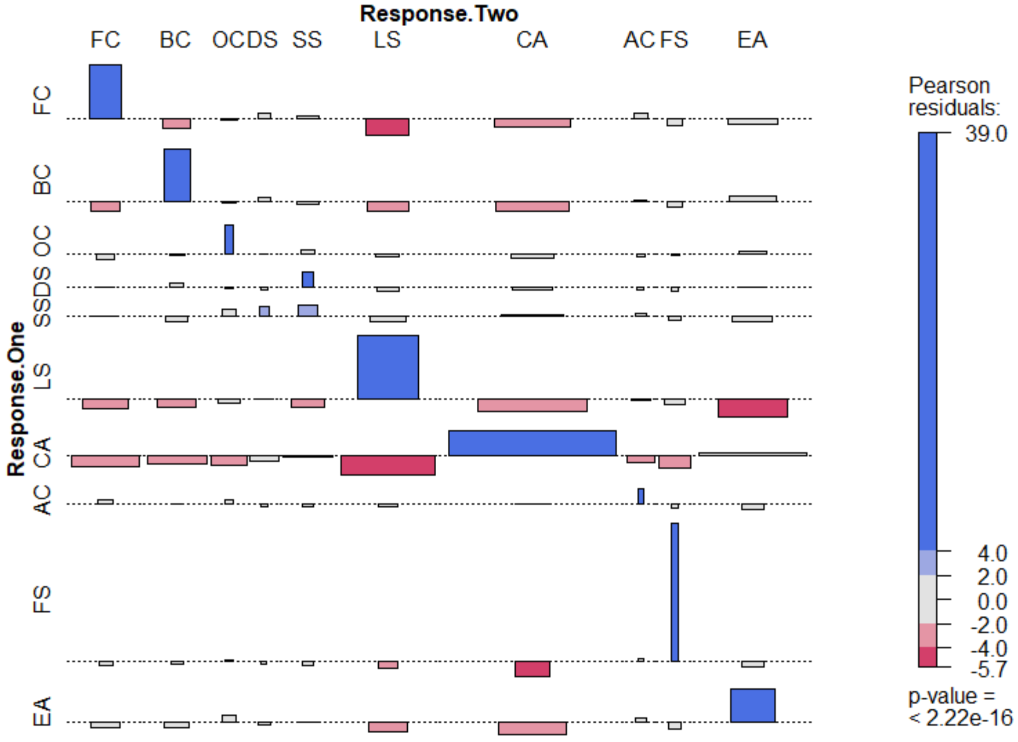


Notes. Blue bars indicate higher numbers than expected and pink bars indicate lower numbers; deep hues indicate a significant difference. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response.

In the meaning-, position-, and form-based coding system (Figure 7.5) the predominance of conceptual associates is clearly indicated. It also shows that the few defining synonyms given tended to be followed by a specific one and that, despite the predominance of lexical set and conceptual associations, these two types of response were negatively associated. This implies that these response types reflect different semantic knowledge. The extreme height of the bar representing form similarity associates is due to most of these associates being responses to a misinterpreted cue word, which meant that both responses were coded in the same way. If these mistakes had been coded as erratic associations they would have had much less influence on the shape of the plot.

Figure 7.5

Association Between Response One and Two (Meaning-Position)



Note. Blue bars indicate higher numbers than expected and pink bars indicate lower numbers; deep hues indicate a significant difference. FC = forward collocation, BC = backward collocation, OC = other collocation, DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association, AC = affix change, FS = form similarity, EA = erratic association.

Given that the specific type of cue word also influenced response one to be of the same type, these association plots are similar, as shown in Appendix 13. However, in the linguistic-conceptual coding system, the influence of the cue on response one shows somewhat more variability than the influence of response one on response two. Interpreting the influence of the cue word in the meaning-, position-, and form-based coding system is more difficult, due to the unequal distribution of cue words across the category types.

As a final analysis of any self-priming tendency in the L2 learner group, the number and proportion of each specific type of linguistic and conceptual associate given as response one and two was calculated as a function of the cue word type and of each other, as shown in Table 7.7. Prior to this analysis, all responses in which response one or two was a non-target response (i.e., taxonomic, other, form-based, or erratic) were removed, as none of these types were used as cue word categories. This means that this dataset is substantially smaller than that used in the previous analyses. This analysis is not shown for the meaning-, position-, and form-based coding system due to the unequal distribution of cue words across cue types. Within this coding system, most cue and response words were conceptual and lexical set associates, so these types would dominate the analysis even more than in the association plot above. However, it is noted that the forward and backward collocation and synonym cue and response categories are highly similar, so would yield similar results.

Table 7.7

Predictive Value of Cue Words and Response One (Linguistic-Conceptual)

Relation	Linguistic categories				Conceptual categories				Total
	Type of association of response one as a function of cue word type								
	FC	BC	SN	AN	OP	CP	SD	SF	
Same specific type	129 .549	110 .556	73 .363	50 .240	104 .525	121 .596	191 .809	51 .261	829 .495
Same general type	27 .115	7 .035	55 .274	73 .351	78 .394	55 .271	19 .081	125 .638	439 .262
Opposite type	79 .336	81 .409	73 .363	85 .409	16 .081	27 .133	26 .110	20 .102	407 .243
	Type of association of response two as a function of cue word type								
Same specific type	101 .430	81 .409	53 .264	25 .120	76 .384	106 .522	177 .750	57 .291	676 .404
Same general type	20 .085	11 .056	45 .224	61 .293	105 .530	77 .379	35 .148	116 .592	470 .281
Opposite type	114 .485	106 .535	103 .512	122 .587	17 .086	20 .099	24 .102	23 .117	529 .316
	Type of association of response two as a function of response one type								
Same specific type	96 .419	62 .392	29 .169	6 .111	60 .469	119 .467	290 .594	76 .398	738 .441
Same general type	35 .153	18 .114	41 .238	16 .296	57 .445	111 .435	117 .240	54 .283	449 .268
Opposite type	98 .428	78 .494	102 .593	32 .593	11 .086	25 .098	81 .166	61 .319	488 .291

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, SD = situation descriptor, SF = semantic field. Same specific type refers to the eight distinct categories used to code the associations. Same general type and opposite type refer to linguistic and conceptual associates.

The results, shown in Table 7.7, reflect the overall tendency to produce conceptual associates. In particular, when a linguistic associate was given as response one it tended to be followed by a conceptual associate as response two, as noted above. It also shows that situational descriptor cues exerted a very strong influence on both response one and two. In general, the cue word tended to influence the L2 speakers' responses more than their own first response, indicating a much-reduced tendency to self-prime association type compared to the L1 speakers (see Chapter 5, Section 5.3.2). However, for cue property and semantic field associates, the influence of the first associate type was stronger than that of the cue word.

7.4 Discussion

The analyses reported in this chapter compared the linguistic-conceptual system for coding word associations with that of the traditional paradigmatic-syntagmatic system (Osgood et al., 1954) and the finer-grained categories based on position, meaning, and form as used in Fitzpatrick (2007). The results showed that participants produced substantially fewer linguistic associates than conceptual ones, especially to cue words predicted to elicit conceptual associates. Although the L2 learners also produced more syntagmatic than paradigmatic associates, this trend was less pronounced, and was more similar to the L1 speaker response pattern, as reported in Chapter 6, and in line with the findings of Nissen and Henriksen (2006).

Importantly, when the WA data was compared to VST scores the linguistic-conceptual coding system was able to discriminate between learners based on the number of linguistic associations produced. This result was found despite the VST results indicating that word knowledge was fairly homogenous across participants and the VST measuring receptive vocabulary knowledge, compared to the productive nature of WATs. The positive association between paradigmatic associations and VST scores was also significant, however, the greater variability in the number of paradigmatic associations given by each participant as a function of their VST score made this trend less reliable. Moreover, this association cannot be interpreted as evidence of a syntagmatic-paradigmatic shift with increasing word knowledge, given that L1 speakers gave more syntagmatic than paradigmatic associations to the same cue words (Chapter 6).

Productive vocabulary knowledge has recently been shown to be related to productive collocation knowledge (Sonbul, El-Dakhs, & Masrai, 2022). These findings extend that relation to receptive vocabulary knowledge, and suggest there may be a similar relation with synonym knowledge. This finding that L2 learners produce fewer synonyms than L1 speakers was noted by Fitzpatrick (2006), and a similar observation was made regarding the development of children's associations (Entwisle, 1966). The difference is that in the new coding system, collocations and synonyms are both viewed as reflecting linguistic knowledge, thus providing a more parsimonious explanation for this finding. In Fitzpatrick (2006) the puzzle is why should position-based responses and one type of meaning-based response be difficult for L2 learners to produce. The answer may be that both types of association reflect linguistic links between words that are less accessible (or less well known).

These findings not only suggest that WATs can be coded and used as an indicator of language learners' word knowledge, but that choosing cue words that tend to elicit linguistic associations from L1 speakers will create more successful tools for this purpose. Further, the results provide additional support for the claim that the production of collocations and synonyms reflect linguistic processing, in support of LASS theory (Barsalou et al, 2008) and other dual coding theories (e.g., Paivio, 1986, Sadoski & Paivio, 2013). If, as Sadoski and Paivio (2013; p.50) claim, the meaning of linguistic units (*logogens*, in their terminology) is determined by linguistic and non-linguistic experience, these results suggest that L2 learners rely on their conceptual knowledge to a greater extent than L1 speakers, due to their relative lack of L2 exposure.

Research suggests that whereas L1 learners tend to process language holistically, L2 learners struggle to learn and use formulaic sequences such as collocations (Wray, 2002). This may be in part due to their relatively low frequency; the highest frequency collocations are spoken English interactional terms, such as "you know" or "I see", in which both words are individually very high frequency words (Shin & Nation, 2008). However, frequency does not seem to offer a complete explanation. Congruent collocations (i.e., word pairs that exist as translation equivalents in both the L1 and L2) may offer a processing advantage to L2 speakers compared with incongruent ones (e.g., Wolter & Yamashita, 2018). Although collocation knowledge improves with proficiency (e.g., Laufer & Waldman, 2011), even advanced learners make many errors. This may be partly explained by interference from L1 collocations (Nesselhauf, 2003) but also seems to reflect a lack of attention to patterns of word use, such as word pairs, in the L2 (Wray, 2002). This interpretation may also be informed by studies using natural reading tasks to examine the acquisition of the order of binomial pairs in an L1 and an L2. When L1 speakers were exposed to multiple instances of novel binomials embedded in texts, such as "wires and pipes", they showed immediate processing gains, as reflected in faster reading speeds for later instances of the novel word pairs (Conklin & Carrol, 2021). In contrast, in a similar experiment involving L2 learners, not only were they insensitive to fixed order binomials in the L2 (which may differ from the L1), such as "black and white", they also failed to show any advantage from repeated exposure to novel binomials (Sonbul, El-Dakhs, Conklin, & Carrol, 2022).

However, it is not yet known why L2 users do not share children's ability to acquire formulaic sequences or constructions (Goldberg, 2019). Two factors proposed by Goldberg (2019) for the acquisition and use of fixed constructions in an L1 are entrenchment and pre-emption. Goldberg and Lee (2021), drawing on Morgan and Levy (2016; see also Benor & Levy, 2006), suggest these two factors may also determine the establishment of binomial word order (as well as cognitive accessibility), explaining why the order can vary across languages and time. In this line, there are hints that some form of statistical learning (e.g., Saffran, 2003; Saffran & Kirkham, 2018; for a general review see Conway, 2020) may play a central role in learning collocations in L1 and L2, as such knowledge may be influenced by out-of-class exposure to the L2 (e.g., watching movies, reading) more than by corpus frequency (González Fernández & Schmitt, 2015). Durrant and Schmitt (2010) showed that L2 learners can begin to learn collocations with repetition, suggesting failure to produce such knowledge is due to insufficient exposure to consolidate the memory trace. Notably, in a WA study involving students at American universities, Zareva and Wolter (2012) found that intermediate level L2 learners produced significantly fewer collocations than advanced learners and L1 speakers. This suggests that, in immersion contexts, L2 learners may acquire collocation knowledge. However, Dąbrowska (2019) found that participants from a range of L1 language backgrounds, living in the UK, had low knowledge of English language collocations, both comparative to their English vocabulary knowledge and L1 speakers. At the university where the research reported in this chapter was conducted, exposure to English is largely confined to the classroom for most students, whose English level was judged to be pre-intermediate. It is, therefore, not surprising that the learner group produced fewer collocations than L1 speakers to the same cues. However, the patterning of collocates with synonym associates was less predicted and should be researched further.

The response patterns found in examining specific response types support the claim that collocations and synonyms share similar processing and suggest that these linguistic relations are less well developed in L2 learners than in L1 speakers. These results suggest that the notion of developing linguistic links may be a useful way to examine word association behaviour in language learners and may be more indicative of language ability than any syntagmatic-paradigmatic shift. In addition, the higher proportion of semantic field and evaluative properties within the conceptual associations produced by L2 learners, as well as somewhat fewer object properties, suggests qualitative differences may also exist in conceptual links. This contrasts with the proportion of meaning-based associates, which showed both groups produced similar proportions of lexical set and conceptual responses. Overall, the comparison of the three different coding systems suggests that coding responses as linguistic and conceptual, and using finer discriminations within these general response types, can reveal new insight into the development of the L2 mental lexicon.

Chapter 8. General Discussion and Conclusion

In this thesis, four word association tasks (WATs) have been reported and analysed in depth. The main aims were to develop and test a system for coding WA data that was grounded in current cognitive science theory rather than in linguistic theory. It was anticipated that this might provide new insights into the types of association that tend to be accessed in such tasks. To achieve these aims, Language and Situated Simulation (LASS) theory (Barsalou et al., 2008) was chosen for several reasons, summarised below:

- It adopts a (weak) embodied approach to cognition, which may constitute the best available framework for understanding cognitive functions (e.g., Kutas & Federmeier, 2011; Louwerse et al., 2015; Meteyard et al., 2012; Fernandino et al., 2016).
- It is a highly influential theory (Borghi et al., 2019; Vigliocco et al., 2009).
- It makes explicit predictions about linguistic and multimodal processing, in line with other influential dual processing theories (e.g., Paivio, 1986; Louwerse, 2011; Vigliocco et al., 2009).
- Its predictions had previously been tested in a WA experiment (Santos et al., 2011) which used general and specific categories, enabling comparison with the two main coding systems that have been used in L2 research, the paradigmatic-syntagmatic system (Osgood et al., 1954) and the meaning-, position-, and form-based system (Fitzpatrick, 2006, 2007, 2009).

A new coding system was developed (based on Santos et al., 2011) through three experiments reported in Chapters 3 to 5, and its informativeness was assessed by analysing multiple associates to each cue word. In Chapters 6 and 7, the new coding system was compared with two established language-based WA coding systems: the paradigmatic-syntagmatic coding system proposed by Osgood et al. (1954) and the meaning-, position-, and form-based system developed by Fitzpatrick (2006, 2007, 2009), using WA data collected from L1 and L2 speakers.

This chapter first summarises the experiments I conducted, focusing on the main hypotheses and findings, then discusses the results in relation to previous research and theories of linguistic and conceptual knowledge, or more broadly, semantic representations. It is intended to provide a broader consideration of the overall findings, building on the discussions in each individual experimental chapter. The central theme of this discussion is a consideration of the nature of WATs, attempting to elucidate the contribution of the new coding system for interpreting the kind of knowledge relations WATs access and how they do so. Finally, I consider the limitations of the experiments reported in this thesis, focusing on the challenges of coding WA data. I then suggest ways in which the new coding system could be used to further investigate linguistic and conceptual knowledge through WATs.

8.1 Summary of Main Findings

The four WATs reported in this thesis started with a close replication of the multiple-response WAT conducted with L1 English speakers by Santos et al. (2011), followed by reanalyses of the data collected. The aim of the original experiment was to test two key claims of LASS theory (Barsalou et al., 2008; see Chapter 2, Section 2.7): (a) that semantic knowledge can be processed through (symbolic) language or through multimodal situated (conceptual) experience; and (b) that language processing is shallower (hence quicker) than simulation-based conceptual processing.

The results of the replication WAT (reported in Chapter 3 of this thesis) were in line with those of the original study by Santos et al. (2011): Participants produced slightly more conceptual than linguistic associates, contrary to Santos et al.'s prediction. However, linguistic associates were produced earlier (in terms of median output position of unique responses) and faster (in terms of onset time of first response) than conceptual ones, supporting LASS theory (Barsalou et al., 2008). Further analysis showed that linguistic and taxonomic associates were produced much less often as the second or third response compared with conceptual associates (in line with De Deyne et al., 2013). This finding suggests that rather than being produced earlier, fewer linguistic responses were produced as later responses. Analysis of the first and second responses given by participants also showed that people tended to produce two associates of the same type, or, to a lesser extent, a linguistic or taxonomic associate followed by a conceptual one. These findings suggested that the distinction between linguistic and conceptual associates might be psychologically valid. However, potential confounds were identified in the cue word selection and response procedure, indicating further investigation was warranted before any firm conclusions could be drawn regarding the informativeness of Santos et al.'s coding system.

To this purpose, I conducted a subsequent experiment aimed at correcting the methodological issues found in the replication (see Chapter 4). In the original study, Santos et al. (2011) stated that taxonomic associates could be produced by either linguistic or conceptual processing, and the replication yielded ambiguous evidence in this regard. Therefore, I decided to focus on linguistic and conceptual associations, which reflect the focus of LASS theory (Barsalou et al., 2008): the distinction between simulated conceptual and symbolic linguistic knowledge. Thus, in the second experiment, cue words were chosen that were predicted to enable production of both linguistic and conceptual associates. Each cue word elicited response words that were identified as reflecting both types of association as the two primary responses in the USF norms (Nelson et al., 1998; see Appendix 6). Three responses were collected to each cue word so that there was an equal distribution of responses across output positions. This meant that every participant and every cue word had the potential to contribute equally to analyses, removing this source of participant and item bias.

The results showed that participants tended to produce multiple associates of the same type, particularly three conceptual associates. In addition, participants were substantially less likely to produce a linguistic associate after a conceptual one. This supports the distinction between these two types of processing and indicates that WATs may provide an effective method for investigating semantic access. Moreover, the results suggested that, when giving multiple associates, the type of association of the first response influenced the second response type, a finding that has received scant attention (but see De Deyne et al., 2013; Nelson et al., 2004; Spätgens & Schoonen, 2020). However, the approach used for selecting cue words appeared to lead participants to produce predominantly conceptual associates, which despite being an interesting finding (as discussed below), may have biased the comparison between linguistic and conceptual associates.

In Chapter 5 (experiment 3), I further investigated the distribution of linguistic and conceptual associates across multiple output positions. For this, I selected high frequency cue words that were predicted, based on the primary responses in the USF norms (Nelson et al., 1998), to predominantly elicit linguistic or conceptual associates, rather than both types of association. In this experiment, linguistic and conceptual associates were produced in similar proportions as the first response. However, the second and third responses were dominated by conceptual associates. As in the previous experiment, this appeared to be due to a tendency not to produce a conceptual associate after a linguistic one. In addition, the results again indicated that the type of response given first influenced the second response, and that this effect that was at least partly independent of the cue word type. The robustness of these response patterns suggests that coding associates as linguistic or conceptual (or taxonomic) could provide new insights into WA.

To test this claim further, in Chapter 6, I compared the distinction between linguistic and conceptual (and taxonomic) associates to the language-based distinction between paradigmatic and syntagmatic associates, proposed by Osgood et al. (1954), which has been used in many WA studies (e.g., Entwisle, 1966; Meara, 1978; Nissen & Henriksen, 2006; Sandgren et al., 2021; Wojcik & Kandhadai, 2020; Wolter, 2001; Zareva, 2007; see Chapter 2, Sections 2.1 and 2.2). I also compared the eight specific categories within the linguistic and conceptual categories to the nine meaning-, position-, and form-based categories proposed by Fitzpatrick (2007; see also Fitzpatrick, 2006; 2009; Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015; Kim, 2013; see Chapter 2, Sections 2.1 and 2.2).

These analyses revealed that the differences between linguistic and conceptual associates across multiple output positions were not apparent when coding responses as paradigmatic and syntagmatic. Furthermore, the validity of the distinction between linguistic and conceptual associates was supported by comparing the two coding systems with the Lancaster Sensorimotor Norms (LSN; Lynott et al., 2020; see Chapter 6, Section 6.5.1). This showed that response words coded as conceptual received, on average, higher strength ratings than those coded as linguistic, but that there was no such difference between paradigmatic and syntagmatic associates. The comparison of specific categories of response types revealed considerable similarities between the new coding system and that of Fitzpatrick (2007). However, the different subcategories used for conceptual associates were arguably more informative for coding responses to the high frequency cue words selected for this experiment. Importantly, these comparisons suggested that synonym associates patterned with collocations rather than conceptual associates, supporting the classification of synonyms and collocations as both being linguistic associations, rather than their division into paradigmatic/meaning-based and syntagmatic/position-based associates.

Finally, in Chapter 7 (experiment 4), I used a subset of the cue words used in the third experiment to conduct a WAT with L2 speakers. This enabled a further test of the explanatory power of the linguistic-conceptual coding system in comparison with the paradigmatic-syntagmatic and meaning-, position-, and form-based systems. The results showed that the L2 speakers produced fewer linguistic associates to the same cues than L1 speakers, despite producing similar proportions of paradigmatic and syntagmatic associates (total numbers were reduced due to a higher number of erratic or other responses). This supported the claim that collocations, synonyms, and antonyms reflect linguistic processing and therefore, should be considered similar associates on WATs, rather than being divided into different types. Moreover, the number of linguistic associates produced by each participant was positively correlated with scores on the first four levels of the Vocabulary Size Test (VST; Nation & Beglar, 2007). Although this correlation was rather weak, it was stronger and less variable than that between VST scores and paradigmatic or position-based associates. Finally, within the specific conceptual categories, the L2 learners produced more evaluative property and semantic field associates than L1 speakers.

Together, the experiments and analyses reported in this thesis suggest the distinction between linguistic and conceptual associates within the new coding system developed from Santos et al. (2011) is psychologically valid and reliable. As discussed below, these results provide a new perspective on WATs and the semantic processing they access, which might, if the linguistic-conceptual coding system is employed in future research, yield new insights into developmental changes in the mental lexicon.

8.2 Accessing Meaning in Word Association Tasks

One of the main outcomes of note was that participants produced a higher proportion of conceptual than linguistic associates across all four WA experiments. This robust finding is interpreted as evidence that WATs do not activate mainly language knowledge, as has been commonly assumed (e.g., Clark, 1970; McRae et al., 2012; Osgood et al., 1954; Santos et al., 2011; Wettler & Rapp, 1993). Instead, conceptual knowledge appears to play a larger role in WA, as suggested by Cremer et al. (2011). Conceptual representations seem to be more accessible for many words, rather than their use in language (De Deyne et al., 2015), even if both sources of word meaning are available (Hanks, 2013; Wittgenstein, 1953/2009). In this regard, in an early review of WA research, Nelson (1977) noted that many researchers interpreted their findings in terms of language processing, but asserted that "word associations are produced according to different principles than are sentences" (p. 101). As reviewed in Chapter 2 (Section 2.3), the assumption that WA is predominantly a language task is challenged by the substantial differences that have been revealed between WA and corpus data (e.g., Kang, 2018; Mollin, 2009), which are also reflected in semantic networks based on these two types of data (e.g., Steyvers & Tenenbaum, 2005). Given that the type of semantic knowledge accessed through WA is the key issue in interpreting such data (see Chapter 2, Sections 2.1, 2.2, and 2.3), the claim made in this thesis that WA activates conceptual processing more than linguistic processing is explored in depth below.

The strongest support for the predominance of conceptual processing in producing WAs was provided in the second experiment (Chapter 4). In this experiment, each cue word was predicted to elicit similar proportions of linguistic and conceptual associates, based on the two dominant associates given in the USF norms (Nelson et al., 1998). Each cue that elicited a primary response that was linguistic elicited a conceptual response as the second most frequent response, and vice-versa. It was hypothesised that this would facilitate the production of similar proportions of linguistic and conceptual associates to each cue word as the first response. As three responses were required to each cue, it was also predicted that many participants would produce both associate types in response to many of the cue words. However, 70% of associates were conceptual, most of which were situational descriptors (thematic associations). If this prevalence of conceptual associates reflects conceptual processing it suggests that, in producing WAs, such knowledge comes to mind more readily for most people than linguistic knowledge. This is in line with a recent WA study by Vivas et al. (2019), in which all cue words were high frequency concrete nouns. Based on one response per cue, they reported that just over 60% of responses were thematic associations.

Importantly, in all four experiments conducted in this thesis, as well as in the WA experiment reported by Santos et al. (2011), conceptual associates were not just more frequent than linguistic ones overall, they were also given nearly or as frequently as linguistic associates as the first response. If conceptual associations are produced via situated simulation, as claimed in LASS theory (Barsalou et al., 2008), this implies that people must be able to activate simulations and linguistic processing at similarly fast timescales. Neural studies have demonstrated that rapid, automatic processing of word meaning is possible (Hauk et al., 2004; Pulvermüller et al., 2001, 2005; van Dam et al., 2014). These studies found that processing action verbs can cause activation in the primary motor cortex within 200 ms of word presentation. Importantly, meaning was operationalised as multimodal or embodied processing, in line with the situated simulations that are asserted to underlie conceptual processing in LASS theory (Barsalou et al., 2008). Single word recognition has also been shown to simultaneously activate competing phonological and semantic neighbours, such that semantic activation onset precedes whole-word recognition (Lewis et al., 2017). Notably, this semantic activation was found in cortical areas implicated in perceptual processing, further supporting embodied cognition explanations.

As argued below, the high proportion of conceptual associates given as first responses in the experiments reported in this thesis is probably due to task demands. The neural evidence discussed above suggests rapid activation of simulation in word recognition, whereas responding on WATs is much slower and involves extra processes (decision making processes, motor production to speak or write the response word). Research has shown that linguistic processing can be detected as early as 50 ms post stimulus onset (Jones et al, 2017; see also Connell & Lynott, 2013). Notably, in the replication reported in Chapter 3, onset times to produce linguistic associates were significantly quicker than for conceptual associates. Thus, a plausible explanation is that linguistic processing may have an early activation advantage over conceptual processing, as claimed by LASS theory (Barsalou et al., 2008), but both types of processing can be activated with near enough simultaneity for this difference to be undetectable in terms of response type on WATs. In line with this interpretation, two examples are given below of priming studies that indicate that words can rapidly activate situated simulations on behavioural tasks.

First, in a picture naming priming study, Ostarek and Vigliocco (2017) found evidence for simulation in the semantic processing of single words. They compared the effect of word primes (displayed in the middle of the screen) that occurred in the same spatial context as target pictures, when the pictures were presented at the top or bottom of the screen (e.g., the word “sky” followed by a picture of a cloud). Priming only occurred when the target was shown in its spatially congruent location, suggesting automatic simulation of the prime word meaning. This is in line with earlier behavioural evidence in support of embodied cognition theories in studies by Zwaan and colleagues (Stanfield & Zwaan; 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012; see Chapter 2, Section 2.5.3).

Second, in an ERP priming study, Chwilla and Kolk (2005) found evidence for the rapid integration of word meanings based on world knowledge (thematic relations). They presented two prime words simultaneously, which were both semantically and associatively unrelated to the target word, but which on related trials could be integrated into a context into which the target could also be incorporated (e.g., the primes “director” and “bribe” followed by the target “dismissal”). On both a lexical decision and a judgement task, response latencies were shorter when the primes formed a congruent context for the target compared to when they were unrelated (the same words in randomised combinations). This was accompanied by the predicted N400 effect indicating facilitated semantic processing (Kutas & Hillyard, 1983; for a review, see Kutas & Federmeier, 2011). This study provides important support for rapid conceptual processing via situated simulation as the facilitation effect was due to online processing (contextual integration) rather than accessing information stored in semantic memory.

The explanation proposed here for the robust finding that WA accesses conceptual processing to a greater extent than linguistic processing is that this is due to the decontextualised nature of WATs, which consist of isolated words. It is argued that this makes WATs very different from languaging, which in its prototypical form, is a situated, interactive activity that unfolds in context (e.g., Linell, 2005; van Dijk, 2016). The lack of co-textual information (compared to linguistic utterances) seems to affect word processing, as reflected in the higher activation of conceptual than linguistic processing. This interpretation is in line with the claim that words do not have meaning, but rather the potential to make meaning (e.g., Anderson et al., 1976; Halliday, 2004; Hanks, 2013; Linell, 2005; Troyer & McRae, 2022). Below, I discuss studies, mostly involving priming tasks, that provide support for the claim that generating responses to single words, as in WATs, is different from language processing and production. These studies have directly compared processing of single words and words in context, demonstrating differences in the underlying mental processes.

A relatively early study that suggested that single words are not representative of language use was a priming experiment by Hess et al. (1995). They reported a series of experiments in which they examined word priming effects within short texts. They created paragraphs which provided a related or unrelated context for the target word (global context) and further manipulated whether the agent in the final sentence including the target was related to it or not (local context). They found that only related global contexts primed the reading of the target word over a baseline control paragraph; no priming occurred when the local context was related but the global context was unrelated. Based on their results, Hess et al. argued that lexical processing should not be examined at the level of isolated words or even single sentences, but within extended discourse.

Likewise, in an ERP study comparing single word and sentence priming, Coulson et al. (2005) found evidence that sentence level information supersedes word level information, when both are available. They used the EAT (Kiss et al., 1973) to select pairs of collocations which varied as to whether the second word was associated with the first (i.e., "olive oil" versus "olive shoes"). They then embedded these word pairs into pairs of sentences, such that each pair appeared in a congruent and incongruent sentence (creating four sentences in total). In the single word presentation, they found a reduced N400 for target words after an associated compared with a non-associated prime. However, when the words appeared at the end of sentences, the N400 was reduced by congruent sentences, with no or little effect of association (depending on the visual field to which the stimuli were presented). These two studies suggest that how words are processed depends on the level of information available (single words or sentences).

Moreover, in a word priming study, Sass et al. (2010) found that requiring participants to name pictures within the same simple sentence produced the opposite pattern of effects on picture naming latency compared to naming single pictures. Participants named pairs of unrelated pictures within a specific sentence stating the position of each object (e.g., "The ship is to the left/right of the button"). A distractor noun was presented on each trial, which was either thematically related to the first or second noun (e.g., "anchor" or "clothing") or unrelated to both (e.g., "wall").

Contrary to the expected facilitation effects for related nouns, they found interference effects for the first picture to be named. They therefore repeated the task with the single pictures as primes and targets, showing the related words facilitated picture naming, in line with previous priming studies (for a review see Hutchison, 2003; see also Chapter 2, Section 2.3.1). This unexpected reversal of the normal priming effect for related items, induced by repeating a simple sentence, further suggests that individual words are processed differently to words in sentential context.

Similarly, Just et al. (2017) examined neural activity associated with processing single words and words in proto-sentences (three-word sentences composed of subject-verb-object). They used fMRI to detect differences in patterns of neural activation in both conditions while participants processed words referring to one of nine objects related to three semantic themes: shelter, manipulation, and eating. They found greater changes in neural activation levels for the nine words when presented in a proto-sentence than in isolation, suggesting sentential context enhanced the context-relevant meaning of the words. This finding, combined with that of Sass et al. (2010), suggests that the context provided by even very basic sentence frames changes word processing (see also Connell & Lynott, 2009).

A recent study suggests that the decontextualised nature of WATs may not only make them dissimilar to language or discourse tasks, but may also influence access to the underlying semantic knowledge needed to produce a linguistic, taxonomic, or conceptual associate. Through analyses of fMRI data obtained from 86 participants watching a variety of movies, Kewenig et al. (2022) showed that the brain areas involved in processing abstract and concrete words, as extracted from discourse, differ according to whether contextual information, provided by visual referents, is present (high context) or not (low context). In high context conditions, activation patterns for abstract words are similar to those for concrete ones, whereas in low context conditions concrete words look like abstract ones in terms of neural activity. Importantly, their results showed differences in activation patterns for abstract and concrete concepts when only language was considered. These results provide further support for the context-dependent, dynamic nature of word meanings and concepts.

The studies discussed above support the explanation provided here for the predominance of conceptual associates in the WA experiments reported in this thesis. In languaging, words are integrated into a sentence and become attended to as part of that specific linguistic structure. However, when a task involves processing single words, without co-textual structure, as on WATs, cue words tend to activate conceptual knowledge. This explanation for WA behaviour also parallels findings from Benor and Levy (2006), in their analysis of constraining factors on the ordering of frozen or fixed binomial word pairs (such as salt and pepper) and unfrozen ones (such as deer and trees). They considered twenty factors that might influence the order of binomial pairs, finding that iconic sequencing (i.e., temporal order) was the most important. However, the second most important factor was experiential (conceptual) knowledge (referred to as *perceptual markedness* in their paper). They concluded that such words "sometimes exist in a simple formal relationship determinable by linguistic properties. But more commonly they are in a complex relationship that can be perceived only through extralinguistic, real-world knowledge." (p. 237).

The claim that the predominance of conceptual associations, found in all four WATs reported in this thesis, indicates that WA is not a simple language-based task further implies that it is a cognitively demanding task. This conclusion is supported by the findings of de Groot (1989) from a continuous WAT. She gave participants 60 seconds to produce as many associates as possible to each of 160 cue words (across three different sessions). She found that participants produced, on average, 7-10 response words for each cue. De Groot posited that the low number of words produced might have been due to the lack of context. De Groot's findings support the claim that most conceptual and lexical knowledge is context dependent (Barsalou, 1982, 1983; Gabora et al., 2008; Rosch, 1999; see Chapter 2, Section 2.4.3). Likewise, a recent study into auditory recognition found that isolated words and connected speech are processed differently (Gaston et al., 2023).

Further support for this interpretation of the predominance of conceptual associates is provided by a study into task demands by Simpson et al. (2001). They measured cerebral blood flow while participants viewed or read a list of 40 high frequency concrete nouns, or generated verbs that described the action or use of the nouns. Based on the observed changes in blood flow in the medial prefrontal cortex they proposed this area of the brain is involved in the mediation of cognitive task performance and task-related anxiety. Generating verbs for nouns was correlated with an increase in response time, self-reported anxiety, and heart rate, compared with viewing or reading nouns aloud and with generating verbs to the same nouns after repeated practice. As noted by the authors, this suggests that tasks such as WATs induce performance anxiety due to the cognitive demand they impose on participants. The task used by Simpson et al. involved a highly restricted association, rather than a free association. However, given their use of high frequency concrete nouns, it seems possible that the results would generalise to WATs. If WA is a cognitively demanding task, this may explain why participants access deeper conceptual processing to a greater extent than shallow linguistic processing, contrary to Santos et al.'s (2011) prediction. As Fillmore (1976) observed "acts and judgements that require abstraction from context are cognitively more complex than the kinds of acts and judgements that occur naturally in context" (p. 24).

In view of this interpretation of the nature of WATs, it is interesting that in their explanation of their word priming study discussed above, Hess et al. (1995) posited that people seek coherency at the highest available level. Thus, comprehension is determined at the discourse level in naturalistic language use but is influenced at the sentence level in experiments involving sentences. In tasks involving isolated words, such as WATs, it is posited that participants search for words that are coherent with the cue word, based on all the information they have about that word. This explanation would also predict that, when providing multiple associations, as in the experiments conducted in this thesis, participants use previous response words to help them produce further coherent responses, a point discussed in the next section.

Related to context effects, a further explanation for the claim that providing WAs is a different and cognitively more demanding task compared to languaging is provided by predictive processing theories of language. Predictions have been shown to facilitate language comprehension (e.g., Brothers et al., 2015; Shain et al., 2020). There is substantial evidence that when people process sentences and texts, they are constantly predicting upcoming words, based on the preceding text, conceptual or semantic knowledge, and grammatical knowledge, including syntactic structure and verb properties (for reviews and theories, see Federmeier, 2007; Kaan, 2014; Pickering & Gambi, 2018). In line with the differential effects found in the studies discussed above, it may be that tasks involving single words, without co-text, such as WATs, are demanding because they do not facilitate predictive processing in the same way as words embedded in texts.

The above interpretation of the type of knowledge accessed on WATs could explain the substantial differences between WA and corpus data, as reviewed in Chapter 2 (Section 2.3.2). Although, as some authors have argued, there are similarities between WA and corpus data (Bel Enguix et al., 2014; Wettler & Rapp, 1993; Wettler et al., 2005), corpus data does not align closely with WA behaviour (Kang, 2018; Mollin, 2009) or with human judgements of words that go together (Dąbrowska, 2014). The distinction between linguistic and conceptual associates made in this thesis, based on Santos et al. (2011), suggests this discrepancy may arise because WAs reflect multimodal experiential knowledge as well as distributional language statistics (e.g., Barsalou et al., 2008; Vigliocco et al., 2009).

The prominence of experiential (conceptual) associations compared with linguistic (or taxonomic) associations may also explain why models of the mental lexicon trained on WA data outperform ones based on corpora (Steyvers & Tenenbaum, 2005; see Chapter 2, Section 2.3.3). The mental lexicon reflects both language use and non-linguistic experience. Further support for this claim is provided by a recent comparison of child-directed speech from the CHILDES database (MacWhinney, 2000) with child-oriented WAs produced by adults (Cox & Haebig, 2023). Cox and Haebig found that the WA data explained additional variance in models of early word learning beyond that accounted for by the corpus data. They speculated that this may be due to WA capturing multimodal (perceptual and affective) experience. Likewise, it has been shown that fundamental properties and contexts of words may be under-represented in language because they are too predictable and thus uninformative (Rohde & Rubio-Fernandez, 2022). In WA this predictable information, or conceptual knowledge of words, seems be prominent.

In view of the evidence discussed above, the results of applying a linguistic-conceptual coding system (based on Santos et al., 2011) to WA data can be interpreted as robust evidence that WA is not primarily a language task. To the extent that WA provides insight into the mental lexicon, the predominance of conceptual associates revealed by the experiments reported in this thesis indicates that connections with the referents of words play a key role in its structure. This finding counters the fundamental purpose of the paradigmatic-syntagmatic coding system, which was proposed by Osgood et al. (1954) specifically to describe WA data on a purely linguistic basis (discussed in Section 8.6).

8.3 Multiple Responses in Word Association and the Mental Lexicon

To my knowledge, this thesis has provided the most detailed analysis to date of the type of response produced in each output position on multiple response WATs. Previous studies have collected two (Nelson et al., 2000) or three associates from L1 speakers (De Deyne et al., 2013, 2018) or two or more associates from L2 speakers (Kruse et al., 1987; Nissen & Henriksen, 2006; Schmitt, 1998b). However, few studies have reported patterns in response type across output positions. In line with the findings reported in this thesis, De Deyne et al. (2013) found that compound word completions (i.e., “bag: hand”, here coded as a type of collocation) and taxonomic associates were predominantly given as first responses, whereas conceptual associates (thematic relations and cue word properties) were more frequent across second and third responses. Their calculation was based on the expected frequency of each response word as a second or third response given the frequency with which it was given as a first response. Similarly, Spätgens & Schoonen (2020) examined first and multiple responses by bilingual and monolingual children, using their own coding system. They found that the proportion of taxonomic relations and features (cue word properties) decreased in later responses, whereas situational responses increased. Finally, in a comparison of multiple associates provided to verbs and nouns with corpus data over a wide collocation span, Schulte im Walde et al. (2008) found that first responses showed greater coincidence with corpus collocations, although this was not a large difference. The findings of the WATs conducted here suggest that synonym and antonym responses are also more likely to be produced as a first response. Moreover, the analyses revealed that these trends may emerge from the greater probability of producing a conceptual associate after a linguistic associate rather than producing a linguistic associate after a conceptual one. These results also support the notion that synonym associates reflect linguistic processing (Santos et al., 2011) and the claim of LASS theory (Barsalou et al., 2008) that conceptual processing is deeper than linguistic processing due to the activation of situated simulations.

Novel analyses conducted in this thesis explored the relationship between earlier and later response types. The findings suggested that on multiple response WATs the type of association produced in the first response to a cue influences subsequent response types. As discussed below, this supports the claim that WA is informative about retrieval processes rather than the storage of word knowledge (Santos et al., 2011). The analyses reported in Chapter 5 (experiment 3) revealed that when L1 speakers provided multiple associates, the second associate tended to have the same general and even specific relation to the cue word as the first associate provided. Importantly, the associate type of the first response was often a stronger predictor of the associate type of the second response than the cue word, despite the cue words having been chosen to influence response types, based on the associations identified by the primary responses in the USF norms (Nelson et al., 1998). This indicates a self-priming effect, rather than a spurious effect of the predicted association between the cues and response words. It was also not the result of chaining, wherein the second response is an associate of the first response word rather than the cue word. The few responses which showed a clear relation to a previously given response but not to the cue word were coded as other or erratic responses. Thus, this effect is interpreted here in line with previous studies, discussed below, that have described clustering effects in retrieving words in response to a prompt.

Troyer et al. (1997) proposed that people search semantic memory by accessing stored clusters and then switching to new clusters in a search process. They asked participants to produce as many words as possible within the category of animals, finding that both the tendency to produce clusters of items and to switch between clusters were associated with the total number of items produced (for example, a participant may produce various types of pets followed by different birds). They suggested that the ability to switch between clusters was an important aspect of optimal search strategies. More recently, models to explain performance on such tasks have been proposed that emphasize either search processes across semantic memory (Hills et al., 2012, 2015) or the structure of semantic memory (Abbott et al., 2015).

Hills et al. (2012) tested two search models of semantic memory (semantic space), inspired by optimal foraging strategy in animals, which has been linked to Lévy flight patterns, which show clustered search paths with longer, sparser paths between clusters. The first model is in line with Troyer et al. (1997) and suggests people produce words from a semantic cluster until they struggle to find more words to add to the cluster, at which point they switch to a new cluster. The alternative model posits that people search memory for a word close in meaning to the previous word that they recalled, such that transitions between categories are fluid and dynamic. Analyses of the animal words produced by participants (within a larger category retrieval task) indicated that switches between clusters were better explained by the dynamic model than the static one. They found that retrieval times were significantly longer at transitions between clusters as defined by fluid rather than static categories. For example, there was no processing cost in retrieving the succession of words “dog-cat-lion”, as dogs and cats are both pets and cats and lions are both felines. However, transitioning between words with no association, such as from “lion” to “bird” incurred a switching cost. In addition, they found that the word produced immediately after a switch tended to be higher frequency than the previous word. They suggested that memory search over stored representations in semantic space is a type of Markov process, in that the next word produced is predicted only by the current word and not by previous words (Hills et al., 2015).

In a reanalysis of Hills et al. (2012), Abbott et al. (2015) modelled clustering and switching effects as a random walk process over a semantic network, which they based on the USF norms (Nelson et al., 1998) for WA. The cue word was taken to be the start node and associations were modelled as a random walk over adjoining nodes. They achieved similar results, suggesting that this was due to the structure of the network. In other words, clustering and switching effects could arise due to online processing (Hills et al., 2012, 2015) or the structure of the semantic network (Abbott et al.), as a model of word storage in the mental lexicon. Although both Hills et al. and Abbott et al. limited their models to category members (animals), given the latter model was based on WA data, it is posited that similar processes may explain the patterning of response types found in analysing the multiple response WA data collected in this thesis.

The finding that response type, not just response word, may prime subsequent responses was surprising. However, this claim is supported by a WA study by McKoon and Ratcliff (1995). They created cue word lists in which the dominant response in the Minnesota norms (Jenkins, 1954) was of a particular type, such as synonym or antonym. Within these lists, they embedded cue words which elicited two high frequency associates of each type (for example, the cue “close” elicited high frequency responses of “far” and “near”). They found that respondents were more likely to produce the response that matched the dominant response type of the other cue words on the list. This suggests that words are not simply associated with other words in the mental lexicon, but higher order relations between words contribute additional structure. However, McKoon and Ratcliff only analysed responses to the embedded cue words. It is also noted that this was a probabilistic effect, ranging from a response probability for a primed superordinate response of only 12% (based on 10 test items) to a probability of 61% for primed adjective-noun pairs, such as “green-grass” (based on 6 test items).

Indirect support is provided by research showing that abstract relations between words can affect word processing. Bencini and Goldberg (2000) investigated how people make decisions about language. They gave participants sets of 16 sentences containing four different verbs in four different grammatical constructions. When asked to sort them into four meaning-related groups, participants were equally likely to sort the sentences by construction type as by verb. Similarly, in a lexical decision task, Johnson and Goldberg (2013) found that nonsense words in congruent constructions primed words that typically occurred in that construction (e.g., the sentence “He daxed her the norp” primed verbs of transfer such as “give” or “hand”). Although both studies explored grammatical constructions, not types of associations between words, they do suggest that priming effects may be possible at a more abstract level than that of word meanings and/or forms.

Moreover, a recent fMRI study by Zhang et al. (2020) found that abstract relations between words can be identified from neural activity patterns. They recorded neural activity while participants listened to stories, then used this data to model semantic representations and relations across cortical areas of the brain. They found that individual words were represented by overlapping networks distributed across various cortical areas, chiefly multimodal association areas. They also found that the type of relation between word pairs, such as part-whole or object-attribute, was associated with different activation patterns, independent of the categories the words belonged to (tools, emotions, etc.). This suggests that the type of association between two words modifies neural representations of the individual words.

The results of the analyses of multiple WA responses collected from L1 speakers (Chapters 4 and 5) suggest that both the cue and previous responses influence semantic search processes on this task. The cue word constrains the associations made, but the previous response word combines with the cue to create the context for the semantic search for the word and/or type of association produced as the second (and third) response. This is indicated by the self-priming of specific association types and by the tendency not to switch to linguistic processing after producing a conceptual associate. This suggests that clustering and switching effects in semantic search processes in WA may not just be a function of the specific words retrieved but may also reflect more abstract processing of how responses are related to the cue word. This interpretation implies that WA reflects word retrieval processes to a greater extent than the underlying semantic memory structure, in line with evidence discussed above for the rapid activation and integration of semantic meaning (e.g., Chwilla & Kolk, 2005). Although this view emphasises the dynamic nature of WA rather than the notion that WA provides insight into a stable mental lexicon, it may be that such processing effects contribute towards the probabilistic but not deterministic influence of cues on WA behaviour, which is what makes this task so informative.

These findings connect to the idea that in tasks involving isolated words, people exploit the reduced context provided (e.g., Hanks, 2013); in this case, the context is formed from the cue and their own previous response(s). Self-priming of response type may occur partly because WATs are not language-like tasks, which, it was argued in the previous section, accounts for the tendency of cue words to activate conceptual rather than linguistic meaning. In the production of multiple responses each response takes structure or context from the previous response as well as from the cue word. In so doing, WA becomes more like normal speech production, in which each word unfolds based on previous words, "laying down a path in talking"[10], as van Dijk (2016) aptly paraphrased from Varela et al. (1991; see Chapter 2, Section 2.5.1). The next two sections explore the implications of the predominance of conceptual associates for theories of semantic memory.

8.4 Conceptual, Taxonomic, and Linguistic Associations in the Mental Lexicon

As reviewed in Chapter 2 (Section 2.4), many theories of conceptual knowledge have emphasised categories or taxonomic relations and shared features (Murphy, 2002). The semantic network approaches discussed in the previous section

[10] This is the title of the article by van Dijk.

(Abbott et al., 2015; Hills et al., 2012, 2015) also assume taxonomic organisation of semantic memory. This assumption underlies the comparison between semantic (category) and associative (WA) relations between lexicalised concepts reviewed in Chapter 2 (Section 2.3). However, the WA data collected in this thesis suggests people access conceptual knowledge primarily through simulations of situations and to a lesser extent through simulation of entities and their properties. This implies that semantic memory is accessed and/or structured according to alternative or additional organising principles (e.g., Barsalou, 2008; De Deyne et al., 2015; Estes et al., 2011; Mirman et al., 2017; Savic et al., 2020). It is of note in this regard that the original system for categorising conceptual associations (Santos et al., 2011) did not include situational descriptor (thematic) associations. This omission was unexpected, given that this category was taken directly from LASS theory (Barsalou et al., 2008), on which the categorisation system was based. Thus, this category was introduced in the subsequent experiments both for selecting cues and coding responses.

Situational descriptors were found to be the dominant response type across all three original experiments reported in this thesis. Moreover, this was the only category for which every individual participant gave a substantial number of responses. To give a specific example, in experiment three (Chapter 5), the minimum proportion of situational descriptors given was 15%, whereas the minimum proportion of any other type of association was 2.5% (the minimum proportion of combined cue and evaluative cue properties was 3.3%). The mean value for all responses from individual participants was also much higher than for any other response category, at 27%, compared with a mean of 15.5% for forward collocations. These findings suggest that situational descriptor associations are central to how semantic meaning is structured and accessed in the mental lexicon.

As noted above, this finding challenges the assumption that conceptual knowledge is organised in hierarchical taxonomic structures (e.g., Collins & Quillian, 1969; Murphy, 2002; see Chapter 2, Section 2.3.3), which has been a dominant assumption not just in amodal theories but also prototype and exemplar theories of conceptual knowledge (see Chapter 2, Section 2.4). Nonetheless, as Lucas (2000) observed, various authors have challenged this view, suggesting, for example, that functional or perceptual properties may be more important. In a meta-analysis of priming studies, Lucas reported larger effect sizes in experiments in which the primes and targets had a functional relationship, such as instrument-action pairs, than in experiments involving coordinate or synonym word pairs. She also calculated that when primes were dominant word association responses of their targets as well as being categorically related, this doubled the priming effect size (increasing it to an approximate *d* value of 0.5). However, she did not calculate the effect size of primes that were related by WA but not by category relations, possibly because category relations were equated with semantic relations, whereas WA was seen as a language boost.

Interestingly, in a WAT utilising familiar and unfamiliar nouns, Chaffin (1997) showed that familiarity influenced whether responses were thematic or taxonomic. He found that familiar nouns tended to elicit thematic associations and properties, whereas unfamiliar nouns elicited taxonomic associations and synonyms. Chaffin (1997) speculated that taxonomic information is important for learning about new entities and words, causing an attentional shift from thematic relations and properties to definitional information. In other words, taxonomic information may be more abstract or language-like than thematic information grounded in experience of the concept. It is possible that when new concepts are learned, taxonomic information and synonym or antonym relations bootstrap conceptual knowledge in a similar way as language is thought to bootstrap number knowledge (Carey, 2009; Pitt et al., 2022).

There is evidence suggesting language is important for learning categories. For example, Zettersten and Lupyan (2020) conducted a series of experiments in which participants learned categories based on colour or shape discriminations, which were either easy or difficult to name using English colour words. They found that while feature discriminability facilitated learning and accuracy, the effect of nameability was larger on both accuracy and learning rate. Zettersten and Lupyan suggested that the naming advantage may arise because when words are available to name the stimuli they act as “priors,” which make it easier to form hypotheses about category structure.

Alternatively, many authors have suggested that coordinate knowledge is grounded in shared features (e.g., Collins & Loftus, 1975; Rosch & Mervis, 1975). Although perceptual similarity does not seem to underlie coordinate knowledge (Mirman & Magnuson, 2009), such knowledge may reflect multimodal simulation (e.g., Kalénine et al., 2009). Evidence from priming studies also supports the relevant importance of shared features over taxonomic relations (e.g., Cree et al., 1999). In this thesis, features (properties) were viewed as conceptual associates, under the assumption that they are (subconsciously) simulated. Thus, the extent to which taxonomic associates reflect linguistic and/or conceptual processing remains an open question (Santos et al., 2011).

The key question for the WA coding system developed in this thesis, based on Santos et al. (2011) is whether the distinction between linguistic, conceptual, and taxonomic associates is psychologically valid. It is argued that the response patterns across multiple associates, discussed in the previous section, provide supporting evidence in this regard. Previous studies further support this claim. A growing body of research has examined the distinction between taxonomic and thematic associations in structuring conceptual representations, as reviewed by Mirman et al. (2017; see also Estes et al., 2011). Although as these authors note, some studies use word pairs that conflate both types of knowledge or use WAs as a proxy for thematic knowledge, there is evidence indicating that taxonomic and thematic knowledge are stored or processed differently, as discussed next.

As reviewed in Chapter 2 (Section 2.4.4), early research suggested that age and/or education influenced the use of taxonomic and thematic relations, with children and illiterate adults relying on thematic relations (e.g., Luria, 1976; Petrey, 1977). Based on WA data, Nelson (1977) suggested that young children rely on real-world situations or conceptual knowledge, but as they get older, their knowledge is structured into relations grounded in taxonomy, meronomy, and synonymy. However, more recent research suggests that thematic relations do not get replaced by taxonomic ones in structuring conceptual knowledge (e.g., Savic et al., 2020; Sloutsky & Deng, 2019). For example, in a large picture naming study with 86 patients with post-stroke aphasia, Schwartz et al. (2011) found that all participants made more taxonomic than thematic mistakes, although mistakes were correlated across both types. Importantly, each mistake type was associated with lesions in different brain areas, indicating differences in the processing of the two types of knowledge. Blackett et al. (2022) also found that taxonomic and thematic mistakes reflect damage in different areas of the brain (see also Chapter 2, Section 2.4.6).

Likewise, in a behavioural study, Mirman and Graziano (2012) found that the tendency to attend to images of taxonomic or thematic distractors in a passive eye-tracking task was correlated with the tendency to choose a taxonomically or thematically related object in a fixed choice task (e.g., if presented with an image of a dog, deciding whether an image of a pig or a bone goes best with that image). This shows that individuals differ in the extent to which they use taxonomic and thematic associations, an effect they showed was not attributable to age or education level. The findings from the experiments conducted in this thesis also suggest that associations between words that are grounded in situational knowledge remain more accessible to adults, rather than a shift towards taxonomic relations.

An alternative explanation is that there is no processing advantage in retrieving a taxonomic association compared to a thematic association. This interpretation is supported by Estes et al. (2011), who reviewed several studies suggesting that taxonomic and thematic knowledge tend to be automatically co-activated, even on tasks requiring only one of these types of knowledge. In an fMRI study, Rabini et al. (2021) investigated category effects on processing simple sentences. They composed three-word sentences in which nouns were either from the same or different categories (for example, "Dogs chase hares" compared with "Hunters shoot deers"). Different areas of the brain showed relatively higher activation for different categories of nouns, and the activation patterns for sentences involving single categories were predictive of those for dual categories. However, behavioural response (reaction times) on a meaningfulness judgement task showed no difference between single and dual category sentences. Moreover, in an ERP picture-word priming study, Savic et al. (2017) found neural evidence that thematic relations were easier to integrate than taxonomic ones, again with no behavioural effects. These studies suggest that thematic associates should be as easily accessible as taxonomic associates.

The studies discussed above, in combination with the findings reported in this thesis, suggest that the primacy attributed to taxonomic categories in the organisation of conceptual knowledge may be a result of the type of tasks that have been used in research into conceptual knowledge, which are often based on category relations or shared features (see also Estes et al., 2011). As Mirman et al. (2011) noted, based on computer modelling, taxonomic knowledge may be more important in identification (e.g., picture naming), whereas thematic knowledge may be more important in processing relations. In WATs, which place no constraints on responses, thematic relations would appear to be the prevalent type of knowledge accessed. This suggests that knowledge of concepts that are familiar due to direct experience with the socio-environment is primarily organized in accordance with that experience, grounded in situational or thematic relations, and that taxonomic relations may play a less central role.

However, the main assumption behind the coding system developed in this thesis, derived from LASS theory (Barsalou et al., 2008), is that linguistic and conceptual associations can be distinguished as they reflect two different processes. In line with other dual approaches to cognition (e.g., Louwerse & Jeuniaux, 2010; Lynott & Connell, 2010; Paivio, 1986; Sadoski & Paivio, 2013; Vigliocco et al., 2009; see Chapter 2, Sections 2.4.5 and 2.6), LASS theory assumes that semantic knowledge reflects language use and situated simulation based on previous experience. As noted in Chapter 3 (Section 3.1.1), no predictions were made by Santos et al. (2011) regarding taxonomic associations.

Less research has considered linguistic knowledge in relation to taxonomic and thematic knowledge, and this tends to be operationalised in terms of collocation. For example, through a series of priming experiments, Jones and Golonka (2012) found evidence for processing differences underlying integrative (collocational), thematic, and taxonomic relations. Although they deliberately avoided word pairs that were highly associated by WA according to the USF norms (Nelson et al., 1998), their findings are still relevant here. They found similar priming effects for all three types of relation, but that priming was affected by lexical cooccurrence for integrative primes, and measures of similarity for thematic and taxonomic primes. Maki and Buchanan (2008) used multidimensional scaling, hierarchical clustering, and exploratory factor analysis to analyse pairs of words that were either collocations, dominant cue-response word pairs in the USF norms (Nelson et al., 1998), or words and their most commonly given defining feature. They found that all three types of word pairs differed from each other in all three analyses. These studies support the validity of the distinction made in this thesis, following Santos et al. (2011), between linguistic, conceptual, and taxonomic associations.

A recent fMRI study into the neural representation of concepts investigated the extent to which concepts were structured by taxonomic, experiential, and distributional (linguistic) information (Fernandino et al., 2022). They tested the correlation of six models (two models created from each type of information) with neural activity across various regions of the brain involved in processing semantic information, including association areas. Participants were presented with words and asked to rate their familiarity with each lexicalised concept (frequency of experience of that concept), for a range of object, event, and abstract concepts. One experiential model significantly outperformed all other models across all areas of the brain. This model was based on human ratings of lexical concepts according to 48 experiential dimensions, including sensorimotor, valence, valuation, and causal dimensions. In addition, the other experiential model, which included only eight dimensions, produced the next best performance. This study provides strong support for the claim made based on the WATs conducted here, that embodied conceptual information, in particular situational descriptors, may contribute more to word knowledge than either linguistic or taxonomic relations. However, it remains possible that the tendency to access situational knowledge or thematic associations is partly due to the nature of WATs, which require responding to isolated words, which, as argued above (Section 8.2), is very different from natural language use (see also Aitchison, 2003).

It is also noted that a recent theory of embodied language, based partly on LASS Theory, argues that multimodal simulations are recruited as needed in processing language, in a task-dependent manner. In their Language-induced activity in modality-specific brain structures (LIAMBS) theory, Cayol and Nazir (2020) posit that simulation is not necessary for language comprehension but rather for understanding the situations that language references. They argue that linguistic processing takes advantage of all pre-existing abilities for processing situations to help humans act optimally on their environment. In this way, simulations are utilised insofar as they support comprehension of the intended meaning of an utterance, but so are other processes for understanding, such as statistical learning of the distributional semantics of language use.

Importantly for the coding system developed in this thesis, based on Santos et al. (2011), the studies discussed in this section demonstrate that linguistic, conceptual, and taxonomic knowledge are dissociable constructs. This supports the claim that the distinction made between linguistic, conceptual, and taxonomic associations is psychologically valid, the chief purpose for proposing an alternative coding system to the language-based coding system proposed by Osgood et al. (1954) and redefined and extended by Fitzpatrick (2006, 2007, 2009). It is argued that this new coding system taps into the strength of WA: WATs can access any type of relation in the mental lexicon (De Deyne et al., 2018). Following Santos et al., it is assumed these relations reflect linguistic, conceptual, and taxonomic knowledge rather than paradigmatic and syntagmatic relations in language (Osgood et al., 1954).

8.5 Comparisons of Different WA Coding Systems

Throughout all the experiments conducted in this thesis the main aims were to explore whether a coding system derived from cognitive science theory could yield new insight into WA and whether this may, in turn, afford new insight into the mental lexicon. In the previous section I suggested that the results indicate that the distinction between linguistic and conceptual associates is psychologically valid. This distinction captured the type of semantic knowledge accessed by participants in the WATs conducted, revealing that, for many cue words, conceptual (thematic) knowledge is more available than linguistic (and possibly taxonomic) knowledge. In addition, analysis of multiple associates to each cue suggested that once situated simulations are activated, there is a reduced probability of accessing linguistic knowledge, whereas the reverse trend does not seem to hold. At the same time, the associate type of the first response may influence later response types, independent of the influence of the cue word. These findings were interpreted in view of the decontextualised nature of WATs and as support for theories suggesting language is grounded in multimodal conceptual knowledge, specifically LASS theory (Barsalou et al., 2008).

Despite these positive findings from the application of the linguistic-conceptual coding system developed in this thesis (based on Santos et al., 2011), it remained possible that other WA coding systems could reveal similar or more informative patterns across multiple responses to each cue. Only a few studies have examined the words or types of association given across multiple output positions (De Deyne et al., 2013, Nelson et al., 2000; Schulte im Walde et al., 2008). To my knowledge, no study has investigated this aspect of WA in depth, as in this thesis. Therefore, to assess whether the new linguistic-conceptual coding system could offer new insight into WA and the mental lexicon, it was compared with the paradigmatic-syntagmatic coding system (Osgood et al., 1954) and the meaning-, position-, and form- based coding system (Fitzpatrick, 2006, 2007, 2009) on L1 speaker (Chapter 6) and L2 learner (Chapter 7) WA data. The findings from these comparisons and their implications are discussed in this and the next section.

One feature of the new coding system is that collocations, a type of syntagmatic associate, and synonyms and antonyms, considered paradigmatic associates, are combined under the assumption that both reflect knowledge gained through exposure to language. Although syntagmatic associates are viewed as a reflection of contingencies in language (e.g., Osgood et al, 1954; Wingfield & Connell, 2022), synonym and antonyms are presumed to reflect substitution. Thus, in the paradigmatic-syntagmatic coding system these associations are assumed to reflect opposing linguistic processes, whereas in the linguistic-conceptual coding system they are assumed to reflect the same underlying processing, in contrast to conceptual processing through situated simulation. This is in line with dual coding theory, which also posits that synonyms and antonyms are processed through language (e.g., Sadoski & Paivio, 2013, p. 62).

One pertinent line of evidence in this regard is research which suggests that antonyms may be learned through contingencies in language (i.e., in the same manner as collocations) rather than by substitution. In a series of experiments Charles and Miller (1989) found that participants could discriminate between antonym pairs, particularly if the entire sentence was provided. Justeson and Katz (1991) explored Charles and Miller's (1989) claims further. They looked at 57 antonym pairs in the Brown corpus, which is tagged for part of speech (see Francis and Kučera, 1982) and a larger, untagged corpus (the APHB corpus, 25 million words, from the American Publishing House for the Blind). They found that most pairs co-occurred within a sentence, often at a statistically significant level (i.e., more frequently than would be expected by chance). Over half the co-occurring antonym pairs were used in direction conjunction (e.g., "hot and cold") or in parallel syntactic structures, a pattern that did not exist with randomly co-occurring adjectives. Both authors concluded that antonyms are not substitutable in context and suggested instead that such pairs may be learnt through co-occurrence in texts. For WA, these two studies imply that if participants produce antonym responses through language processing, they are recalling such words based on syntagmatic rather than paradigmatic use. Although this argument may not extend to synonym production on WATs, it indicates that the division between paradigmatic and syntagmatic associations is problematic.

Furthermore, the syntagmatic-paradigmatic shift in young children´s WA responses was attributed to their lack of synonym and antonym knowledge (Ervin, 1961), not a lack of collocational knowledge. Yet, Nelson (1977) and Petrey (1977) both argued that children's WA data reflected their experience of the world (see Chapter 2, Section 2.2.1). Likewise, Fitzpatrick (2007) found that L2 learners produced fewer defining synonyms than L1 speakers. These findings point to the possibility that synonyms and antonyms are learned through language. Conceptual or thematic associations, learned through experience of the world (perhaps with the aid of linguistic labels, see Section 8.8) are more accessible to language learners than synonyms and antonyms. This difference is captured by using a linguistic, conceptual, and taxonomic coding system, providing an explanation of the differences between younger children's or L2 learners' WA responses and those of older children or adults.

In addition, applying the paradigmatic-syntagmatic coding system (Osgood et al., 1954) revealed that L1 speakers produced predominantly syntagmatic associates to the high frequency cue words used (see Chapter 6, Section 6.2). This was unexpected, given that paradigmatic relations (such as taxonomic or category relations) are assumed to structure the adult mental lexicon (e.g., Burke & Peters, 1986; Deese, 1965; Riegel & Riegel, 1964; Zareva, 2007). However, other studies have reported a similar result (e.g., Hirsh & Tree, 2001; Nissen & Henriksen, 2006). Given that most responses were coded as conceptual, meaning-based associates when Fitzpatrick's (2007) coding system was used, this suggests that paradigmatic associations may not capture the semantic meaning of words to a greater extent than syntagmatic ones. It is possible that Osgood et al.'s definition of paradigmatic associations as words that could substitute one another in speech has led to the tendency to assume words of the same word class reflect paradigmatic associations (e.g., Ervin, 1961; Piranian, 1983; see Fitzpatrick & Thwaites, 2020 for discussion of this issue). However, Osgood et al. (1954) stated that syntagmatic associations "can include both words of the same as well as of different form classes" (p. 116). In line with their discussion of paradigmatic and syntagmatic associations (see Chapter 2, Section 2.1.2), in this thesis, cues and responses of the same word class which referred to inanimate and animate entities or contrasted time or place with an entity or thing were coded as syntagmatic, as well as collocation noun pairs (see Chapter 6, Section 6.2.1). This decision could explain the high proportion of associations coded as syntagmatic. It also suggests that the paradigmatic-syntagmatic coding system proposed by Osgood et al. may not have been consistently applied across subsequent studies in the manner that they intended.

Importantly, the tendency not to produce a linguistic association after a conceptual one, interpreted as evidence for the deeper processing involved in simulation (Barsalou et al., 2008; Santos et al., 2011), was not mirrored by a similar trend in producing paradigmatic and syntagmatic associates. This suggests that the paradigmatic-syntagmatic coding system is not capturing a processing difference that reflects different types of knowledge. Moreover, a linguistic, conceptual, and taxonomic coding system does not impose the ideology that one type of association reflects better semantic knowledge than another. This contrasts with the assumption that paradigmatic associations are superior to syntagmatic ones (e.g., Burke & Peters, 1986; Ervin, 1961; Sandgren et al., 2021).

When the specific response categories were compared to those introduced by Fitzpatrick (2007), the differences were much less pronounced. This was due to the considerable similarity between the two coding systems. However, as Fitzpatrick (2007) divided position-based responses into different types of collocation (including compound nouns) and classified all other types of response as meaning-based (apart from the few form-based and erratic responses), most responses to the high frequency cue words used in experiment 3 were coded as meaning-based (Chapter 6). Moreover, within this general category, only one category, conceptual associates, captured most of the WA data. In contrast, the four conceptual categories in the new coding system enabled further discrimination of responses.

Of note, Fitzpatrick (2006) employed 17 specific categories, including quality and context associations, similar to the object property and situational descriptor categories used in this thesis. However, responses in these categories were infrequent in that study, leading Fitzpatrick to discontinue their use in later studies (e.g., Fitzpatrick, 2007, 2009). This is presumably due to the cue words used in Fitzpatrick (2006), which were all from the academic word list (Coxhead, 2000) and included no concrete nouns. The majority of associates produced in Fitzpatrick's (2006) study were forward collocates or defining synonyms of the cue word. Given that academic words are, by definition, words that occur in academic discourse to a greater extent than they are used in daily conversation to refer to real-world entities, they are more likely to be learned through language rather than grounded in multimodal experience. This may explain why these two types of association dominated responses, further indicating that they are both linguistic relations. It also suggests that the optimal coding system for WA data may, in part, depend on the choice of cue words.

8.6 Word Association in an L2

As a final assessment of the new coding system, L2 speaker associations were collected and then coded using three different coding systems (Chapter 7). As with the L1 speaker WA data analysed in Chapter 6, the new linguistic-conceptual-taxonomic coding system developed in this thesis, based on Santos et al. (2011), was compared with the paradigmatic-syntagmatic system (Osgood et al., 1954) and the meaning-, position-, and form-based coding system (Fitzpatrick, 2006, 2007, 2009). As expected, L2 learners produced more erratic or other associations, especially in their second and third responses. More importantly, when L2 speaker associations were analysed using the linguistic-conceptual coding system the results suggested that learners rely on conceptual knowledge to compensate for their difficulties in retrieving linguistic associations. The L2 learners produced significantly fewer collocations and synonyms than L1 speakers to the same cues, but more similar proportions of conceptual associates.

It is posited that one reason why a syntagmatic-paradigmatic shift has not reliably been confirmed in L2 learners (Fitzpatrick & Thwaites, 2020) is that syntagmatic associates include collocations. It is speculated that studies that have found this shift may have included many cue words that tend to elicit synonyms and antonyms from L1 speakers (such as the Kent-Rosanoff words, see Chapter 2, Sections 2.1.1 and 2.2.2). As noted by Fitzpatrick (2006), the largest difference between L1 speaker and L2 learner WA behaviour is that the latter group produce significantly fewer collocation associates (although see Zareva & Wolter, 2012). Similar differences in the proportion of collocations and synonyms were also apparent in the specific categories used by Fitzpatrick (2007), but importantly, in that system, collocations and synonyms are split into position- and meaning-based categories, respectively, whereas they are both categorised as linguistic in the coding system used in this thesis. This suggests that the general categories of linguistic and conceptual associations may provide a more harmonious interpretation of the differences between L1 and L2 speakers than the paradigmatic-syntagmatic or meaning-, position-, and form-based distinctions. Indeed, the idea of a shift from one type of response to another does not seem to capture the underlying developmental changes in learners' lexicons. It would seem instead that increasing linguistic knowledge leads to greater opportunities or ease in accessing these kinds of links on

WATs, with a concomitant decrease in the reliance on conceptual knowledge. This implies that constructing WATs solely from words that have been found to elicit mostly linguistic associates from adult participants responding in their L1 may provide a more useful tool for assessing developing lexicons.

The L2 speakers also produced more semantic field associates (assumed to capture more abstract semantic knowledge) and slightly more taxonomic associates compared to L1 speakers, which it is speculated may reflect their experience of the foreign language being limited to the classroom and thus having fewer experiential connections compared with the L1. A similar argument was put forward by Li et al. (2011), to explain their finding that high school students produced a lower proportion of thematic responses in their L2 (English) compared to their L1 (Chinese), but similar proportions of taxonomic associates. However, in the experiment reported in Chapter 7, L2 learners produced predominantly situational descriptor (thematic) associates, a difference which may be due to their relatively higher English proficiency level. In addition, the production of evaluative properties by L2 learners is tentatively posited to reflect a compensatory strategy (not necessarily conscious) for difficulties in retrieving semantic associations.

Further support for classifying collocation, synonym, and antonym associates as linguistic was provided by the correlation between the number of linguistic associates produced by each participant and their score on the VST (Nation & Beglar, 2007). Although this correlation was quite weak ($\tau_b = 0.18$) it was consistent across participants, with no outliers. Importantly, the relation was stronger and more reliable than that between VST scores and paradigmatic or position-based associates. This finding was also somewhat unexpected as English proficiency level was similar across participants and the VST is a receptive vocabulary test, whereas WA involves both receptive and productive vocabulary knowledge, two key differences that would be expected to reduce any correlations between the two measures (e.g., Edmonds et al., 2022). In this regard it is of note that VST scores were not significantly correlated with the number of clang or other associates produced, although such associates are typically produced more frequently in L2 than L1 (e.g., Nissen & Henriksen, 2006). Likewise, in a restricted WAT, in which participants were required to provide multiple responses to nouns (such as a synonym, a property, and a coordinate or member of the same set), proficiency was correlated more with the overall number of responses than with the types of responses produced (Riegel et al., 1967).

Thus, these results suggest that as well as a qualitative difference in the exact words produced as associations by L1 and L2 speakers, there is also a quantitative difference in the number of linguistic associations, in line with Zareva (2007). It is posited that if WATs are constructed from cue words that have been found to elicit a range of linguistic associates (collocates, synonyms, and antonyms) from L1 speakers, WA behaviour may be more strongly related with L2 proficiency. One potential benefit of coding responses and determining proportions of linguistic associates, in contrast to judging L2 word associations by their stereotypy or similarity to L1 speaker norms (e.g., Schmitt, 1998b; Wolter, 2002), is that this does not impose a monolingual standard which may be inappropriate (e.g., Meara, 1983; Zareva, 2010). In addition, individual participants do not always score highly on stereotypy in their L1 (Kruse et al., 1987), suggesting that this may be a biased benchmark.

The relative difficulty in retrieving linguistic associates in an L2 was further evidenced by examining the relationship between multiple responses. Whereas L1 speakers tended to provide two responses of the same general or specific type, in the L2 speakers this pattern was only present for conceptual associates. This suggests that only these types of association led to the clustering effects in terms of which the L1 data were interpreted (see Section 8.3), indicating that linguistic connections between words may be less accessible or less well formed in this group of L2 learners.

Importantly, the comparison of the three coding systems for L2 WA data indicates that differences in syntagmatic and paradigmatic associations by L2 learners of different ability levels may be obscured by difficulties in accessing specific relations within both types of associations, as previously observed by Fitzpatrick (2009). The division of these specific types into two opposed general categories may have reduced the potential of coding WA responses as a method for understanding the development of linguistic knowledge in L2 learners. In contrast, the linguistic-conceptual coding system contains the inherent assumption that linguistic associations provide direct evidence of language knowledge, whereas conceptual associations provide only indirect evidence of word knowledge (semantics).

The finding that conceptual associates are more frequently produced by both L1 and L2 speakers is challenging to the appropriateness of using WATs to assess the development of the mental lexicon. For example, Meara (1983) argued that conceptually mediated responses were uninformative about the development of the L2 as they were presumed to reflect translation. Indeed, the evidence presented in Chapter 7 suggests that cue words which elicit dominant conceptual associates from L1 speakers do not discriminate between L2 learners in terms of their L2 lexical knowledge. Thus, the importance of cue word selection in L2 WA research is reiterated here (Fitzpatrick, 2006, 2007; Thwaites, 2018; Wolter, 2002). However, it is suggested that features such as word class and frequency may be less important than types of association given in the primary responses to the cue words, as indicated in norms lists such as the USF norms (Nelson et al., 1998) or the multilingual SWOW (De Deyne et al., 2018). Selecting cue words which elicit dominant linguistic responses from L1 speakers for use in WATs with L2 speakers may yield more informative results, as suggested by the positive correlation between linguistic responses and VST scores.

Notwithstanding, there is now strong evidence that the L1 is always activated when processing the L2 (for a review, see Brysbaert & Duyck, 2010), including neural evidence showing bilingual speakers activate both languages in tasks only requiring use of the L2 (Thierry & Wu, 2007), even in tasks which do not elicit behavioural effects (van Heuven, 2008). Moreover, in an L1 WAT with trilinguals, van Hell and Dijkstra (2002) found that participants responded significantly more quickly when the cue words were cognates in their L2, although there was no effect for the weaker L3. Together, this research suggests that participants completing a WAT in an L2 will inevitably activate L1 lexical knowledge. Thus, it may be inappropriate to argue that conceptually mediated responses are less valid than responses that are assumed to be produced through language. Nonetheless, comparison with the paradigmatic-syntagmatic coding system (Osgood et al., 1954) and the meaning-, position-, and form-based system (Fitzpatrick, 2006, 2007, 2009) indicated that all four associate types classified as linguistic by Santos et al. (2011) and in the WAs conducted in this thesis (collocations, synonyms, and antonyms) collectively contributed towards the discrimination of L2 vocabulary processing and possibly knowledge.

8.7 Advantages of Coding Word Association Data

As discussed in Chapter 2 (Section 2.1), there are two basic approaches to analysing WA data: producing WA norms based on the frequency with which particular response words are given to cues or, as in this thesis, coding responses. Frequency counts of the actual words produced is clearly a less time-consuming approach and, as discussed in the next section, avoids certain pitfalls that arise when coding WA data. However, as argued previously, comparing the WAs of different populations to L1 adult norms may result in biases and/or yield little information, given substantial individual differences in responses. Yet, in L2 WA studies in which the type of association is coded, it is typical to use 100 cue words or fewer (e.g., Fitzpatrick, 2006, 2007, 2009; Kim, 2013; Nissen & Henriksen, 2006; Norrby & Håkansson, 2007; Sökmen, 1993; Wolter, 2002; Zareva & Wolter, 2012), which makes the choice of cue words a critical issue in the study design. As noted by Fitzpatrick (2006), the rationale for cue word choice is often not provided or lacks systematic criteria. Indeed, her use of cues from the academic word list (AWL; Coxhead, 2000) was an attempt to address this issue.

Following on from and extending the WA experiment reported in Santos et al. (2011) and replicated in Chapter 3, in the experiments conducted in this thesis, the cue words were carefully selected, based on the most frequently given responses in the USF norms (Nelson et al., 1998). The association of these response words to the cue was determined and formed the basis of the selection criteria. In all three experiments (Chapters 4, 5, and 7), an equal number of cue words were selected that elicited a dominant response of a specific linguistic or conceptual association type. Although, as anticipated, the dominant WAs produced by L1 speakers at a U.K. university in the 2010s were not identical to those produced by L1 speakers at a U.S. university in the 1990s, the substantial similarities indicate considerable stability in WA behaviour (see Appendix 9). The ability to predict the type of WAs that should be produced most frequently means that WA data can be tailored to address specific research questions. As noted above, cue words that elicit a range of linguistic cues from L1 speakers may be most appropriate for research questions regarding language knowledge or processing, such as assessing developing lexicons. In contrast, cue words that elicit dominant conceptual responses may be better suited to questions concerning thematic knowledge and possibly for use in semantic priming studies.

To the best of my knowledge, Santos et al. (2011) were the first authors to select cue words based on dominant response type. The experiments reported in this thesis suggest that this could be an informative and systematic approach in WA research. Importantly, this criterion, especially if cues that elicit one extremely dominant response are avoided, has a probabilistic but not an absolute influence on the response word or type. Thus, group level and individual level differences can be explored, such as comparing L2 speakers of different ability levels in the WAT language or comparing individuals who tend to produce mostly linguistic associates with those who produce mostly conceptual ones. In this thesis, it facilitated the interpretation of patterns in association types over multiple output positions.

8.8 Challenges of Coding Word Association Data

The main limitation regarding the research reported here is the validity of the WA coding system. It should be noted that this aspect of WA research constitutes a challenge for any WA coding system; it is not a new issue that arose with the linguistic-conceptual coding system developed in this thesis, based on Santos et al. (2011). The key problem is that determining how to categorise a response word is not always a straightforward decision (e.g., Fitzpatrick, 2006; Fitzpatrick & Thwaites, 2020; Meara, 1983).

In the case of the paradigmatic-syntagmatic system proposed by Osgood et al. (1954), some researchers (e.g., Ervin, 1961; Piranian, 1983) have simplified this decision by relying purely on word class. However, in many studies, noun-noun associations have been coded as syntagmatic when they form a compound noun or when the response is more likely to co-occur with the cue than to replace it in any meaningful way (e.g., Nissen & Henriksen, 2006; Wolter, 2001). The main challenge is how to determine the association type when the cue and response possess a paradigmatic and a syntagmatic link, such as “root: vegetable”). As noted in Chapter 2 (Section 2.1.2), Osgood et al. suggested that such responses should be coded as paradigmatic and syntagmatic, although this practice has not been applied using this system. As an illustration of this issue, Wettler et al. (2005) gave the example of “coffee-tea” as a paradigmatic associate and “coffee-drink” as a syntagmatic one. Yet “tea” is also one of the most frequent collocates of “coffee” in both the BNC and

COCA corpora. Likewise, the association "drink" could refer to the noun instead of the verb, in which case it would be the superordinate category of coffee and thus, a paradigmatic associate. Hutchison (2003) noted that many associations can be classified in more than one way; yet suggested that "baby: boy" is a collocation, without considering that it could also reflect a taxonomic link.

In the WATs reported in this thesis some associations, such as "root: vegetable", were also ambiguous when employing the linguistic-conceptual distinction. In such instances, the categorisation strategy reported in Santos et al. (2011), whose WAT was replicated in Chapter 3, was used. Their solution was to code any ambiguous relation as linguistic, based on the assumption that WATs are primarily a linguistic task and so participants are more likely to be accessing language knowledge in producing associates than conceptual knowledge. Thus, for example, the response "call" to the cue "telephone" was coded as a forward collocation, despite the evident conceptual connection. However, also following Santos et al., properties were coded as conceptual even when they could form a collocation (e.g., "root: vegetable" or "green: grass").

The most appropriate way to deal with this coding issue may depend on the research aims. If the focus is on understanding individual WA behaviour, a potential solution would be to ask the respondents to explain ambiguous associations, a strategy explored by Fitzpatrick (2007) and in a recent online study by Liu et al., (2022). However, given that in the experiments conducted in this thesis, each respondent typically produced over 100 associations, it would have been impractical to interview them about all ambiguous associations and improbable for the respondents to be able to introspect accurately. More importantly, the focus of the experiments conducted in this thesis was to understand the mental lexicon. Idiosyncratic associations may be uninformative in this regard (Nelson et al., 2000). For the purpose of this thesis, a more appropriate solution would be to code ambiguous responses as dual link associates, as pioneered in Fitzpatrick & Izura (2011; see also Fitzpatrick et al., 2015). However, with four linguistic and four conceptual categories, this would have resulted in 64 possible combinations, which would have been too unwieldly to analyse and have led to sparse data issues. Despite this, it would be informative to include and compare a dual category for linguistic and conceptual associates with a dual category for paradigmatic and syntagmatic associates.

In this regard, one advantage of the linguistic-conceptual distinction over the paradigmatic-syntagmatic distinction is that these two processing systems are posited to be activated in parallel (Barsalou et al., 2008). This allows for interaction (e.g., Evans, 2009; Sadoski & Paivio, 2013; Santos et al., 2011), such that the production of each association could reflect the degree of activation of language and conceptual processing. Categorisation could be considered as a cline rather than a rigid dichotomy. Santos et al. (2011) described it as a probabilistic distinction (p.91). Conceptual and linguistic associations can be distinguished on the assumption that conceptual links activate (subconscious) simulations, whereas responses coded as language-based do not require simulation (collocations, synonyms, and antonyms). However, language-based associates may still prompt simulation, at least some of the time and in some people. On the other hand, some conceptual associates suggest co-construction through language and conceptual knowledge (e.g., Evans, 2009, 2016; Lupyan & Lewis, 2019; see Chapter 2, Section 2.6.2). As proposed by Vigliocco et al. (2009) in their discussion of abstract and concrete concepts, although an absolute distinction between linguistic and conceptual knowledge or processing may not exist, the relative contribution of both types of knowledge can vary as a function of the associations being produced. This in line with Rubin (2022), who argued, in a discussion of different types of memory, that scientific description necessitates categorisation despite the continuous nature of the phenomena being described. WA data is no exception.

Although some concepts are learned prelinguistically (e.g., Clerkin & Smith, 2022; Mandler, 2007), it is also apparent that language can support concept learning (e.g., Connell & Lynott, 2013; Dove et al., 2022; Pavlenko, 2014; Vincent-Lamarre et al., 2016; Whorf, 2011). The challenges that emerged in this thesis with the specific semantic field category may reflect the role of language in structuring conceptual knowledge. This type of conceptual associate was developed from the semantic field cue word category employed in Santos et al. (2011), which was presented as a conceptual category. However, in their study, semantic field cues were predominantly words that, in the replication (Chapter 3) elicited responses from the same, restricted lexical set (e.g., planets or months). Santos et al. appear to have coded such responses as taxonomic associations. The cues selected by Santos et al. represent quite a narrow view of semantic field compared with previous authors, who have viewed it as a conceptual association realised through language (for a historical review, see Nerlich & Clarke, 2000). In many accounts, semantic field is approximately synonymous with the mental lexicon, intended to capture the relations between words that structure how language represents conceptual knowledge (Kittay & Lehrer, 1992). Indeed, Barsalou (1992) described semantic fields as the lexicalisation of conceptual fields, with an emphasis on dynamic frames rather than features and taxonomies (p. 63). In other words, in this view, all meaningful WAs could be described as semantic field relations.

In this thesis, semantic field was operationalised as a conceptual association that is not grounded in perceptual or motor experience, in other words, as an abstracted semantic relation. This was in part due to the observed lack of an appropriate category for such associations in analysing the WA data collected in the replication experiment (Chapter 3). However, some authors have suggested that abstract concepts can be grounded by emotion and interoception (Vigliocco et al., 2014) or social interaction (Borghi et al., 2019). Associations such as "religion" or "god" to the cue "church" may reflect situated simulations of interoceptive experience, including emotion, suggesting that they could be interpreted as thematic associations or situational descriptors (Troyer & McRae, 2022). However, semantic field associations were more likely to be followed by a linguistic association compared with the other three types of conceptual associates. Such response words were also more similar to linguistic associates in terms of sensorimotor strength ratings in the LSN (Lynott et al., 2020). It is speculated that semantic field associations may reflect conceptual knowledge that is partly grounded by language (e.g., Borghi et al., 2019; Dove et al., 2022).

Likewise, it is possible that taxonomic relations can be learned purely through language or purely through experience of the world, with superordinate categories most likely to be learned through explicit instruction, via language (Renoult et al., 2012). In contrast, coordinate category relations are more likely to be able to be learned through direct experience, without relying on linguistic input (Jardak & Byers-Heinlein, 2019), or with language playing a later role in acquisition (Nelson, 1988). This may explain why taxonomic associations are both similar to and distinct from linguistic and conceptual associations. Such associations may frequently involve accessing both types of knowledge. Overall, the challenges that arose in implementing the linguistic-conceptual coding system can be interpreted as reflecting the challenge of disentangling language and conceptual knowledge (e.g., Barsalou et al., 2008; see Chapter 2, Section 2.6).

This argument suggests it may be difficult, and potentially impossible, to elucidate the contribution of linguistic and conceptual experiences to our knowledge of words and phrases (Morgan & Levy, 2016). Discourse is typically simultaneously situated within, and constructive of, real-world experience. This issue was acknowledged by Barsalou et al. (2008) in their article introducing LASS theory, who stated that "frequencies and correlations in perceived situations are mirrored in frequencies and correlations of words used to describe them" (p. 252). Distributional (language) and representational (conceptual) networks are not isolated, but interconnected, with both contributing information to the interpretation of any individual word (e.g., Evans 2009; Vigliocco et al., 2009). In a recent review, Warren and Dickey (2021) highlighted the lack of empirical studies and theoretical frameworks regarding the interaction of linguistic and world (conceptual) knowledge.

In addition to theories that language and conceptual knowledge interact to create meaning (e.g., Borghi et al., 2019; Evans, 2009; Humboldt, 1836/1999; Whorf, 2011), there is also the notion that language may be simulated to some extent, rather than being purely symbolic. For example, Kussmaul (1877) posited that onomatopoeic sounds and words form an essential starting point in the acquisition of language, both evolutionarily and in child development. More recently, this idea has been developed as the sound symbolism bootstrapping hypothesis by Imai and Kita (2014). Further research into sound symbolism and language has also suggested that linguistic symbols or words are not always completely arbitrary (e.g., Westbury et al., 2018; Winter & Perlman, 2021). Likewise, in the motor theory of speech (Liberman & Mattingly, 1985), language production and perception are viewed as complementary aspects of a single system (see also Galantucci et al., 2006; Pickering & Gambi, 2018). Borghi et al. (2019) proposed that abstract concepts are grounded in social and linguistic interactions. They reviewed evidence that due to this grounding in language, words that refer to abstract concepts activate areas of the brain that are involved in speech, in processing auditory input and oral motor actions.

From an enactive cognition perspective, Cisek (1999) argued that "*the function of the brain is to exert control over the organism's state within its environment*" (p. 9). In other words, cognition does not mediate between action and perception, but is part of an action-perception control loop (see also Buzsáki, 2019; Seth, 2021). Importantly, such control is dependent on the existence of regularities in the environment and relations between entities, both of which are features of language. If language is also part of a control loop, interacting with perception and action, it may also embody meaning. Likewise, Evans (2009, 2016) argues that although multimodal conceptual representations must be evolutionarily prior to linguistic representations, the language system contributes unique semantic content, which can suffice to create new knowledge of the world, albeit in a more schematic form. From this perspective, language may also play a role in people's illusions regarding the extent of their conceptual knowledge of objects and natural kinds (Johnson-Laird, 1983; Rozenblit & Keil, 2002; see Chapter 2, Section 2.4.2). The combination of being able to perceive and name entities may lead people to believe they understand them, even when their conceptual knowledge is limited due to lack of direct experience.

The evidence and theories discussed above illustrate the challenges of distinguishing between the contribution of linguistic and conceptual processing to WA. Some associates may reflect one type of knowledge more than the other, but this may vary between individuals and even within an individual over time (e.g., Connell & Lynott, 2014). Nonetheless, the comparison of responses with the Lancaster Sensorimotor Norms (LSN; Lynott et al., 2020) in Chapter 6 provided tentative support for the validity of the assumption that linguistic and conceptual associates are, to some extent, separable. Overall, conceptual response words were rated higher in sensorimotor strength than linguistic ones, suggesting situated simulation may play a greater role in producing conceptual than linguistic associates. Importantly, when the same response words were coded as paradigmatic and syntagmatic associates, there was no appreciable difference between the two response types. This implies that classifying responses as linguistic and conceptual associations captures information beyond that revealed by classifying all associations as language-based semantic associations. This indicates that the contribution of linguistic and conceptual processing to WA responses can at least be probabilistically distinguished, as claimed by Santos et al. (2011), even if these processes are not completely isolable.

8.9 Limitations of Study Designs

The main limitations of the experiments reported in this thesis are related to the challenges of coding WA data discussed above. However, four important methodological points are noted here.

First, the pen-and-paper format in which all the WA experiments, except for the replication experiment, were delivered created space limitations which were not ideal for collecting multiple responses. In order to recruit volunteer participants and keep the task delivery manageable, each cue word was presented in a table, with three spaces to the right for participants to write their response words. This enabled all WATs to be printed on one piece of A4 paper. However, compared to presenting each cue word individually, three times, it may have increased the tendency for response words to be influenced by previous responses. Digital presentation, which would allow each cue word to be presented individually and each response word to disappear from the screen before further responses were made, may have alleviated this possible issue.

Second, the main aim of this thesis was to test and refine the linguistic-conceptual coding system, to assess its utility for coding word association data. Thus, the cue words were chosen based on the association that was determined between cue words and the dominant response to them in the USF norms (Nelson et al., 1998). As a result, the comparison with the meaning-, position-, form-based system developed by Fitzpatrick (2007), should be interpreted with caution, given that the cue words did not provide a balanced sample of the categories used within this system.

Third, but most importantly, only the replication experiment included cue words that were chosen because they elicited a taxonomic associate as the dominant response in the USF norms (Nelson et al., 1998). In the further three WA experiments reported here, no words were chosen that elicited a dominant associate that was taxonomic. Therefore, the importance of taxonomic or categorical associations within the mental lexicon is underestimated in this thesis. For example, Gruenenfelder et al. (2016) found that hybrid graph-theoretic models which combined contextual models (taxonomic and thematic relations) and associative models (text-based co-occurrence) were best able to model word association data. This suggests including all three types of cues may elicit more representative WA data. In a recent study, Cox and Haebig (2022) elicited free WAs and child-oriented

WAs to 672 cues words from adult participants, then compared both sets of responses to the CHILDES language database (MacWhinney, 2000). They found that semantic networks based on the WA datasets showed greater taxonomic structure than a network based on CHILDES, indicating that taxonomic relations are more prominent in WA than corpus data.

However, there is evidence to suggest that predominance of conceptual associates would not be changed by the addition of taxonomic cues. In an experiment in which all words were selected from the USF norms, Crutch and Jackson (2011) found a small but significant negative correlation between ratings of associativity and similarity, indicating that people do not judge associations between words based on taxonomic relations. In a series of picture matching tasks, Savic et al. (2020) found that adults were more accurate and faster to find thematic than taxonomic matches to a target, and tended to rate thematic matches as more similar to targets than taxonomic ones. Property generation tasks also tend to elicit significantly more situational descriptors than taxonomic relations (Lenci et al., 2013; Wu & Barsalou, 2009). Lenci et al. (2013) produced property norms for concrete and abstract nouns from sighted and congenitally blind participants. For both concrete and abstract nouns, both groups produced mostly situational descriptors, followed by features or properties, with fewer taxonomic relations (mostly superordinates). These findings suggest that conceptual associates should remain dominant even if taxonomic cues were included in the WAT. Nevertheless, such a WAT, including equal proportions of linguistic, conceptual, and taxonomic associates, is a clear next step.

A final but crucial issue is the replicability of the experiments reported here. The relative stability of word associations within individuals has been confirmed, despite generational differences (Fitzpatrick et al., 2015). Many of the responses produced in the WATs conducted in this thesis conformed to the predicted type of association to the cue word, based on a norms list created over 20 years ago with a student population in a different country (Nelson et al., 1998). A comparison of the three dominant responses in the USF norms (Nelson et al., 1998) with the three dominant responses to each cue used in experiment three (Chapter 5) showed that approximately 50% of response words were the same. This demonstrates that WAs, as least for high frequency cues, is relatively stable across populations. The similarities between the WA data collected in this thesis and that collected in the USF norms also indicate that collecting multiple associates to each cue (as done here) may have a limited or no effect on the first response produced to each cue. This further supports the claim that it may be more informative to collect multiple associates to each cue (De Deyne & Storms, 2008; De Deyne et al., 2018). However, beyond similarities in the actual words produced, the question remains as to whether another researcher implementing the coding system described in this thesis would make similar decisions. The only evidence in this respect is the similar proportion of linguistic, conceptual, and taxonomic associates produced in the replication experiment (Chapter 3) to in the original study by Santos et al. (2011).

8.10 Future Directions

The research reported in this thesis suggests that categorising responses given on word association tasks as linguistic and conceptual associates is psychologically valid and may provide greater insight into processing word meanings than the paradigmatic-syntagmatic division. However, in focusing on these two types of relations, the role of taxonomic relations in the mental lexicon was not explored. Although the primacy of taxonomic relations in the organisation of conceptual knowledge has been challenged (e.g., Estes et al., 2011), this type of knowledge should be investigated in greater depth in further WA studies. The reason for excluding cue words that elicit a dominant taxonomic association was because, as noted by Santos et al. (2011), it is as yet unclear whether taxonomic knowledge is grounded in sensorimotor or linguistic experience, or both. In addition, given that taxonomic associates, by definition, can only be given to cue words referring to concrete nouns, a WAT comparing linguistic, conceptual, and taxonomic associates would be restricted to concrete noun cues to avoid confounding this factor. Nevertheless, such a comparison may provide insight into the relative role of linguistic and conceptual processing in producing taxonomic associations.

The repeated finding that participants in WATs are more likely to switch from a linguistic associate to a conceptual associate than from a conceptual to a linguistic one suggests WATs may be able to offer further evidence on the nature of taxonomic knowledge. If taxonomic relations are partly structured through language, such associates should be equally likely to be produced before or after a linguistic associate, but occur less frequently after a conceptual associate. However, if taxonomy is a fundamental aspect of conceptual knowledge, taxonomic relations should be evenly produced before or after conceptual associates. Furthermore, if conceptual knowledge is grounded in taxonomic relations to a greater extent than thematic relations or situational descriptors, including cue words that are predicted to elicit taxonomic associates should lead to participants producing more taxonomic associates overall than situational descriptors (but see De Deyne et al., 2013).

Within the responses coded as conceptual associates, the large number of situational descriptors given by both L1 speakers across all three experiments and by L2 learners in the fourth experiment suggests that it could be informative to divide this category further. One possibility might be to divide situational descriptors into those that refer to social and physical features of the situation (e.g., "beach: sunshine" and those that refer to events and actions within it (e.g., "beach: sunbathe"). This could be seen as aligning situational descriptors with Schank and Abelson´s (1977) notion of scripts. It would also emphasize a link to object and cue properties, as situational descriptors could in this way be viewed as situational properties and actions/events. It may be interesting to find out which type of situational association is more frequently produced and whether this reveals further individual differences.

Although this aspect was beyond the focus of this thesis, the linguistic-conceptual coding system revealed large individual variation, with a few participants who produced mostly linguistically associated responses and many participants who produced mostly conceptual associates. Such individual differences in WA behaviour were highlighted by Fitzpatrick (2007), yet have still not been extensively explored. If the distinction between linguistic and conceptual associations reflects differences in semantic processing, individual differences could relate to language or imagery skills or preferences (e.g., Paivio, 1986). It would be interesting to investigate whether the proportion of linguistic associates produced on a WAT correlates with language aptitude or strength of imagery tests. For example, Dąbrowska (2014) found that L1 speakers' knowledge of collocations was strongly correlated with vocabulary knowledge and reading experience. In a later study, she found evidence that L1 attainment is related to language aptitude (Dąbrowska, 2018). Moreover, given the discussion regarding the role of association in priming studies (Hutchison, 2003; Lucas, 2000, see Chapter 2, Section 2.3.1), it is speculated that the type of association individuals tend to produce on WATs may correlate with individual differences in priming effects, in line with individual differences in the tendency to access thematic or taxonomic relations (Mirman & Graziano, 2012). Recent evidence suggests such differences may even reflect differences in the underlying concepts that words reference (Marti et al., 2023).

In addition, in the final experiment, WA data was collected from L2 learners with similar English proficiency and vocabulary knowledge. It is important to extend data collection to other groups of learners, with higher proficiency levels and different L1 backgrounds. This would provide a much more robust examination of whether linguistic associations are predictive of L2 vocabulary knowledge and/or proficiency level. If cue words that tend to elicit linguistic associates from L1 speakers are able to differentiate between L2 learners of varying proficiency this could open up new lines of research. For example, it may be interesting to find out whether a cue word list composed solely of linguistic cues can prime L2 learners to produce more linguistic associates and whether any such effect is modulated by proficiency. Further research in this area may be able to lead to the production of a WA test that is more appropriate for assessing the developing mental lexicon of L2 learners. Alternatively, it may suggest that WA should be considered a task to explore semantic processing but may be less suitable as a test of the size or structure of the mental lexicon.

8.11 Conclusion

The main finding in this thesis is that WA reflects conceptual processing to a greater extent than linguistic processing. This suggests it is not primarily a language-based task (e.g., Osgood et al., 1954; Santos et al., 2011). This interpretation is based on analysis of multiple associates provided by both L1 and L2 speakers of English, using three different coding systems. The comparison between coding systems indicated that a coding system that distinguishes between linguistic and conceptual associations may be more informative about the nature of WA and the semantic knowledge accessed by this task than the paradigmatic-syntagmatic coding system proposed by Osgood et al. (1954) based on “purely linguistic criteria” (p. 115). Further, the experiments reported in this thesis confirmed the relevance of the detailed understanding attained by combining (but not replacing) such general categories with finer-grained distinctions, as pioneered by Fitzpatrick (2006, 2007, 2009). Thus, the main theoretical and methodological contribution of this thesis is the development of a linguistic-conceptual coding system based on the system proposed by Santos et al. (2011; see also Simmons et al., 2008) and informed by LASS theory (Barsalou et al., 2008). This new system for analysing WA data has the potential to lead to new insights about the mental lexicon and its development. It may also provide a new approach for contrasting WA data with corpus-based approaches to understanding semantic relations between words and evaluating the differences between WA and other semantic knowledge tasks, such as property generation or categorisation tasks.

In addition, the grounding of conceptual knowledge in multimodal representations was supported, suggesting that such thematic or situational knowledge may be more foundational in the relations people construct between words and concepts than taxonomic or categorical relations (e.g., De Deyne et al., 2015; Estes et al., 2011; Savic et al., 2020). However, it is noted that the importance of this latter knowledge in identifying instances of concepts is not assessed by WATs.

The research in this thesis also offers new insights into the effect of cue word properties on response behaviour and strongly suggests that cue words should be carefully chosen according to the aims of the study. For tasks investigating semantic knowledge, such as priming effects between cues and their primary associates, cue words that predominantly elicit conceptual associates, or specifically situational descriptors, may elicit stronger effects. In contrast, for studies investigating the development of word knowledge, for example in L2 learners as in the final experiment reported in this thesis, selecting cues predicted to elicit linguistic associates may be more informative. Interpreting WA behaviour in such groups as an increasing accessibility of linguistic associations rather than a syntagmatic-paradigmatic shift may offer a productive way forward for future WA research.

Although coding WA data can be challenging, the abstracted structure of association types provides insights into semantic processing beyond those obtained when only the individual response words are considered. It is argued here that word association data is so complex to analyse, despite being so simple to obtain, because it assesses both in-the-moment processing and long-term representations of both linguistic and conceptual knowledge. Each individual response is determined by in-the-moment processing yet the collection of responses from many participants reveals the underlying stability of long-term relations between words and the concepts they represent within a population. In particular, the relative accessibility of conceptual knowledge compared with linguistic knowledge challenges the assumption within cognitive science that WA solely reflects the distribution of words in language (see Kumar, 2021; Troyer & McRae, 2022). WATs are equally, and perhaps even more importantly, a useful tool for understanding relations that are constructed through multimodal conceptual knowledge. This provides an explanation for the lack of correspondence between such tasks and corpus-based measures of lexical relations.

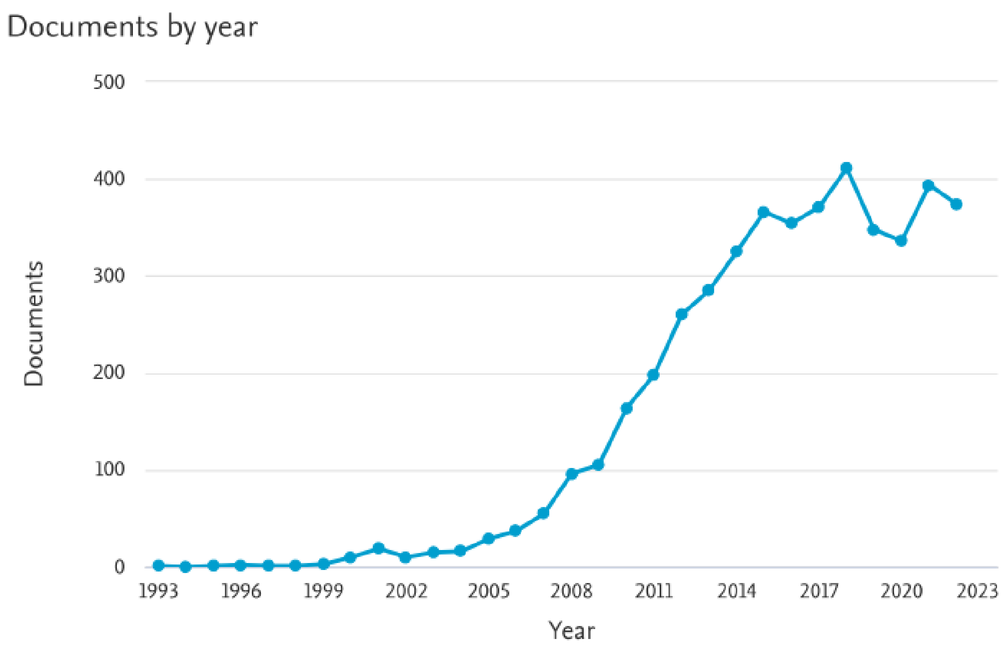
When WA responses were coded as linguistic, conceptual (situated simulation or thematic), and taxonomic relations, the general categories corresponded to detected differences in neural processing (e.g., Mirman et al., 2017; Vigliocco et al., 2009). In addition, an explanation was provided for the observed differences between WA and corpus data: differences arise as many WA responses are produced through accessing conceptual knowledge rather than linguistic knowledge. These three types of knowledge are assumed to interact and individuals are assumed to differ as to which type of knowledge they tend to access on WATs. However, it is posited in this thesis that the strength of WATs is that they provide insight into how all three types of knowledge shape the mental lexicon, thus accessing "the *right kind* of meaning" (De Deyne et al., 2018, p. 1000). In contrast to Aitchison (2003), who asserted that due to the multiple types of association people form between words, WATs "cannot tell us about the probable structure of the human word-web" (p. 85), it is argued here, following De Deyne et al., that this is exactly why coding WA data reveals the probabilistic structure of the mental lexicon. The mental lexicon is complex, dynamic, and partly individualistic, but these aspects are why, as Goldberg (2019) illustrated, humans can exploit old meanings and explore new ones according to the constraints and potentials of each new context.

Finally, although in this thesis the linguistic-conceptual coding system has been presented as a new alternative to the traditional paradigmatic-syntagmatic system, it may also be seen as motivating a return to the roots of WA. In the first ever reported WA study, Galton (1879) noted that, across multiple responses to each cue word, he most frequently produced associations he classified as *histrionic*, which he described as "an imaginary mental theatre" (p. 159). These would appear to be associates grounded in situated simulation, coded as situational descriptors in this thesis. Galton contrasted such responses with verbal associations, which were slightly more likely to be produced first, and associations based on multimodal mental imagery, “the chime of imagined bells, the shiver of remembered cold” (p. 159), which were more often produced second. These latter associations resemble the object and cue property categories used in this thesis. A coding system for word associations in which responses are assumed to reflect grounded or embodied conceptual processing as well as language processing, is perhaps then, not a new approach, but merely a forgotten one. The research reported in this thesis suggests this approach warrants further investigation.

Appendices

Appendix 1 – Scopus Search

Scopus search for titles, abstracts, and key words containing any of the four phrases embodied/grounded + cognition/language. (Total as of December 2022: 4587 articles.)



Appendix 2 – Cue Words by Concept Type

Forward collocation:

taxi, golf, golden, federal, financial, fashion, self, extension

Backward collocation:

muffin, league, cane, station, doll, lamp, lens, disc

Synonym:

teacher, guess, calculate, breakable, throw, thief, secluded, excellent

Antonym:

heavy, guilty, past, exit, accept, bumpy, divorce, good

Taxonomic category:

car, bee, cello, tiger, lobster, cashew, chimpanzee, parsley

Semantic field:

winter, September, north, Jupiter, Monday, monthly, gallon, Europe

Stereotypical object property:

wings, stubborn, smelly, slow, horns, slimy, sweet, green

Brand (not used in the replication):

Crest, Tylenol, Nike, Budweiser, Speedstick, Marlboro, Microsoft, McDonalds

Appendix 3 – Replication Experiment Word List

Cue Word Lists			
1	**2**	**3**	**4**
although	she	because	him
up	when	who	out
September	monthly	station	green
slow	guilty	wings	lamp
bee	throw	guess	stubborn
teacher	Europe	fashion	Monday
bumpy	cane	parsley	extension
muffin	accept	Jupiter	heavy
taxi	sweet	thief	disc
north	golf	federal	divorce
excellent	tiger	good	winter
league	doll	gallon	car
exit	golden	lobster	breakable
cello	cashew	past	self
slimy	secluded	smelly	chimpanzee
financial	horns	lens	calculate

Note. Each cue word list was also created in reverse order, starting with the two practice words (also in reverse order).

Appendix 4 – Dominant Responses to Each Cue in the USF Norms

Cue Type	Cue	Target	Percent of Responses	Cue Type	Cue	Target	Percent of Responses
FC	extension	**cord**	35.4	BC	cane	**sugar**	27.3
FC	fashion	clothes	37.8	BC	disc	**compact**	23.4
FC	fashion	**statement**	10.8	BC	doll	house	23.1
FC	federal	**government**	28.1	BC	doll	**baby**	12.8
FC	financial	money	41.2	BC	lamp	light	76.9
FC	financial	**aid**	15.5	BC	lamp	**desk**	1.9
FC	golden	yellow	11.3	BC	league	**baseball**	55.1
FC	golden	**Graham**	3.5	BC	lens	glasses	55.3
FC	golf	**ball**	15.5	BC	lens	**contact**	18.2
FC	golf	**club**	15.5	BC	muffin	bread	20
FC	self	me	47.9	BC	muffin	**blueberry**	11.6
FC	self	**esteem**	2.8	BC	station	**train**	31.9
FC	taxi	**cab**	53.7				
SN	breakable	glass	38.8	AN	accept	take	16.7
SN	breakable	**fragile**	25.9	AN	accept	**reject**	10.7
SN	calculate	math	27.3	AN	bumpy	road	32.7
SN	calculate	**compute**	5.8	AN	bumpy	**smooth**	15
SN	excellent	**great**	27.9	AN	divorce	**marriage**	33.1
SN	guess	question	14.2	AN	exit	**enter**	38.8
SN	guess	**hypothesis**	3.1	AN	heavy	**light**	36.1
SN	secluded	**alone**	31.8	AN	good	**bad**	75.8
SN	teacher	student	19.2	AN	guilty	**innocent**	39.9
SN	teacher	**instructor**	7.5	AN	past	**present**	42.6
SN	teacher	**professor**	7.5				
SN	thief	steal	38.8				
SN	thief	**robber**	22.4				
SN	throw	ball	23.4				
SN	throw	catch	23.4				
SN	throw	**toss**	20				

Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym. Responses in bold font are the most frequently produced response of the cue category type in the USF norms (Nelson et al., 1998). Both responses are shown when two relevant responses were given by the same proportion of participants.

Cue Type	Cue	Target	Percent of Responses	Cue Type	Cue	Target	Percent of Responses
TX	bee	sting	36.2	SF	Europe	**Asia**	10.8
TX	bee	**insect**	5.5	SF	gallon	milk	29.7
TX	car	**auto**	13.3	SF	gallon	**quart**	8
TX	cashew	**nut**	74.5	SF	Jupiter	**planet**	47.4
TX	cello	**violin**	23.5	SF	Monday	**Tuesday**	35.3
TX	chimpanzee	**monkey**	68	SF	monthly	period	19.1
TX	lobster	**crab**	13.9	SF	monthly	**weekly**	17.8
TX	parsley	green	14.1	SF	north	**south**	77.2
TX	parsley	**seasoning**	6	SF	winter	**summer**	37.2
TX	tiger	**lion**	30.8	SF	September	NA	NA

Note. TX = taxonomic, SF = semantic field. Responses in bold font are the most frequently produced response of the cue category type in the USF norms (Nelson et al., 1998). September does not appear in these norms.

Cue Type	Cue	Target	Percent of Responses
OP	green	**grass**	25
OP	horn	**car**	20.7
OP	horn	**bull**	7.6
OP	slimy	**goo**	11.7
OP	slow	fast	52.7
OP	slow	**turtle**	11.5
OP	smell	taste	18.1
OP	smell	**perfume**	3.6
OP	stubborn	**mule**	17.6
OP	sweet	sour	37.2
OP	sweet	**candy**	16.2
OP	wings	**bird**	47.6

Note. OP = stereotypical object property. The cue “horn” was changed to “horns” and “smell” was changed to “smelly” in the experiment by Santos et al. (2011). Responses in bold font are the most frequently produced response of the cue category type in the USF norms (Nelson et al., 1998).

Appendix 5 – Cue Words by Recoded Concept Type

All recoded words are in bold font. Note that the cue words “horns” and “smelly” are not in the USF norms in these exact forms, so these two cue types were not recoded. In addition, the cue words “guess” and “stubborn” were excluded from the analysis in the replication.

Forward collocation (10 cues):
taxi, golf, federal, extension, **doll**, **lamp**, **bumpy**, **bee**, **gallon**, **monthly**
Backward collocation (4 cues):
league, cane, station, disc
Synonym (4 cues):
secluded, excellent, **self**, **accept**
Antonym (9 cues):
heavy, guilty, past, exit, divorce, good, **throw, slow, sweet**
Taxonomic category (9 cues):
car, cello, tiger, lobster, cashew, chimpanzee, **golden**, **muffin**, **teacher**
Semantic field (6 cues):
winter, September, north, Jupiter, Monday, Europe
Stereotypical object property (8 cues; one excluded from replication):
wings, stubborn, smelly, horns, slimy, green, **lens**, **breakable**,
Conceptual, not object property (6 cues; one excluded from replication):
fashion, **financial**, **guess**, **calculate, thief, parsley**

Appendix 6 – Cue Words and Dominant Associates

Forward collocation

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
coral	reef	55	ocean	SD	21
dart	board	53	game	SF	25
drum	stick	39	band	OP	26
heart	beat	23	love	CP	19
law	order	18	police	SD	12
stomach	ache	40	food	SD	23

Backward collocation

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
address	home	41	phone	SF	20
bicycle	ride	34	wheel	CP	18
castle	sand	19	king	SD	15
criticism	constructive	11	mean	CP	10
memories	past	19	thoughts	SF	13
siren	police	39	loud	CP	21

Synonym

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
baby	child	33	cry	SD	18
fluffy	soft	38	cat	OP	30
infection	disease	50	sore	CP	15
spit	saliva	40	gross	CP	11
superstition	belief	17	witch	SD	13
vain	conceited	15	mirror	SD	11

Antonym

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
asleep	awake	52	dream	SD	20
deep	shallow	46	ocean	OP	16
dirty	clean	42	smell	CP	25
marriage	divorce	25	love	CP	15
sunset	sunrise	43	beach	SD	13
winter	summer	35	cold	CP	26

Object property

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
bang	gun	40	boom	SN	23
breakable	glass	57	fragile	SN	38
graceful	ballet	23	clumsy	AN	16
hairy	ape	14	bald	AN	13
tail	dog	36	end	FC	20
wiggle	worm	24	shake	SN	18

Cue property

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
beggar	poor	51	bum	SN	19
dew	wet	23	morning	BC	23
ice-cream	cold	38	cone	FC	21
lecture	boring	21	talk	SN	20
stove	hot	47	oven	SN	46
taboo	bad	20	forbidden	SN	12

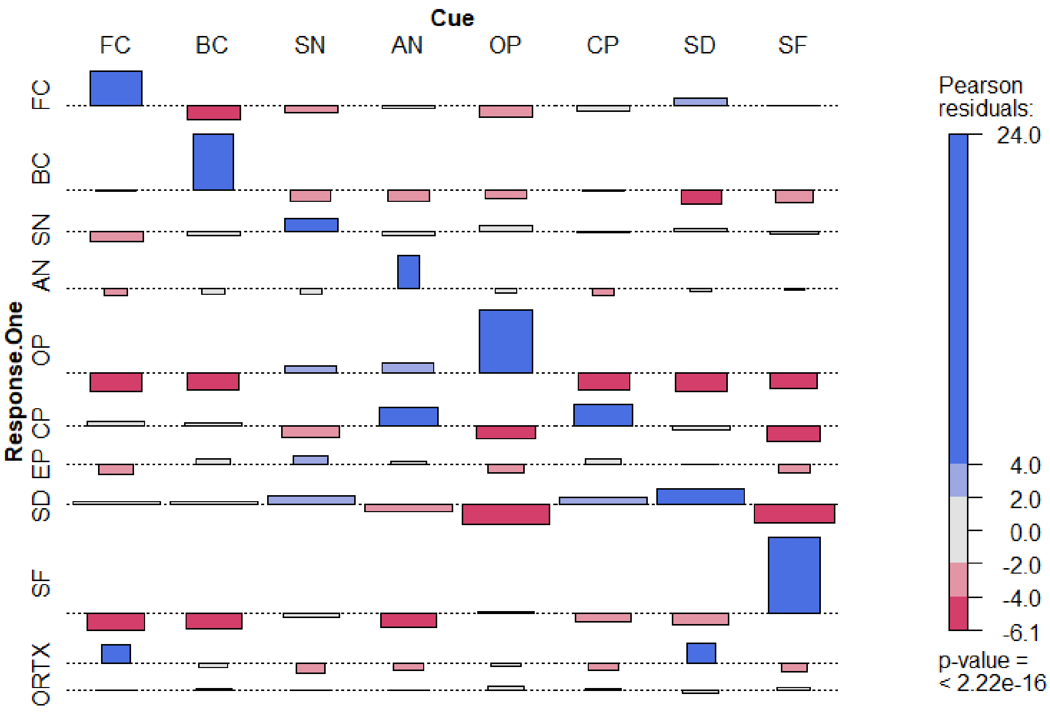
Situational descriptor

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
cheese	mouse	29	cracker	FC	28
exercise	sweat	17	jog	TX	11
hearing	ear	26	aid	FC	21
massage	relax	26	rub	SN	25
mirage	desert	26	image	SN	18
thief	steal	64	robber	SN	37

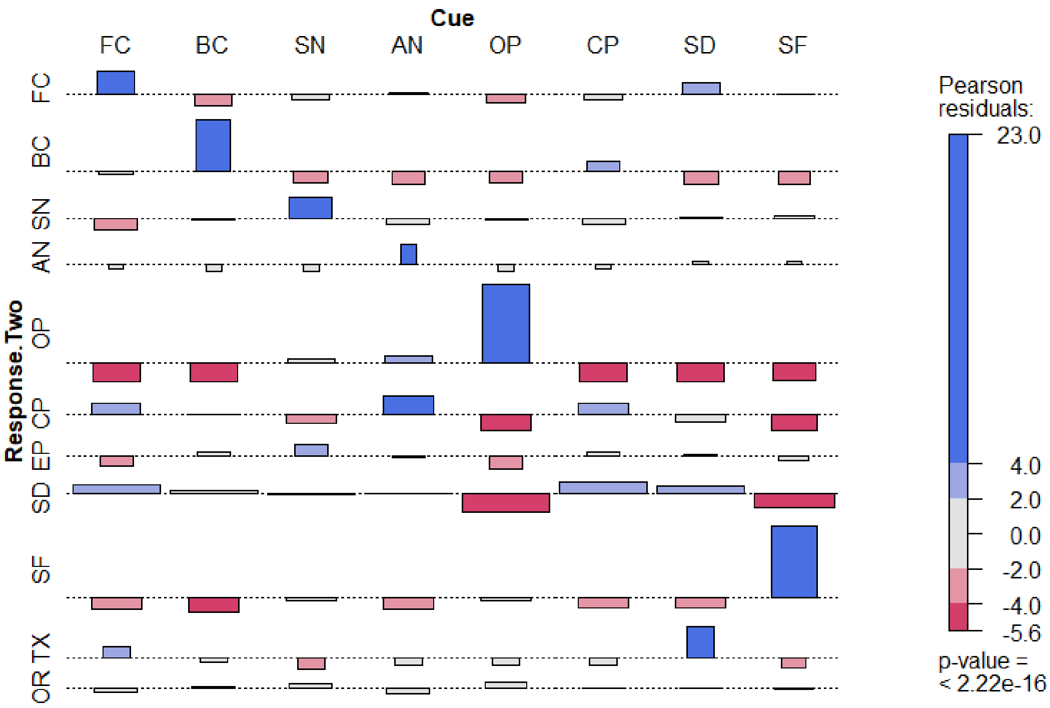
Semantic field

Cue word	Primary response	Percent respondents	Secondary response	Secondary response type	Percent respondents
active	sports	13	busy	SN	12
extinct	dinosaur	48	gone	SN	35
fashion	clothes	58	statement	FC	16
financial	money	61	aid	FC	23
hungry	food	37	starve	SN	33
inertia	force	12	movement	AN	12

Appendix 7 – Association Between Cue and Response One and Two



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = object property, CP = cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response.



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = object property, CP = cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = taxonomic, OR = other response.

Appendix 8 – Cue Word Lists

Cue Word Lists

1 (side 1)	**1 (side 2)**	**2 (side 1)**	**2 (side 2)**
child	friend	time	wish
programme	ticket	beach	event
nature	end	damage	computer
turn	toilet	station	bar
blood	foreign	for	social
church	serious	mistake	kiss
telephone	weekend	arrive	healthy
fire	sign	page	fish
name	restaurant	class	help
fat	tell	total	model
welcome	scene	trouble	soldier
swim	danger	lose	have
mirror	work	think	hospital
wings	cloud	member	important
measure	government	education	habit
local	line	interesting	root
real	boat	body	dream
belt	family	tourist	grow
elephant	enjoy	burn	crime
chance	field	quiet	bag

Note. Each cue word list was also created in reverse order.

Appendix 9 – Primary Response Compared with the USF norms

The tables below compare the three primary responses in the USF norms (Nelson et al., 1998) with the primary responses produced in the experiment reported in Chapter 5, for the first responses given to each cue and all three responses give to each cue. Each table shows the ten cue words of each cue type. Primary response words in bold font were also primary responses in the USF norms, those in italics were primary responses within first responses only and across all three responses.

Forward collocation

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
computer	program disc keyboard	*laptop screen* game	*laptop* **keyboard** *screen*
for	against you to	***against you*** *me*	***against you*** *me*
foreign	language country different	***language country*** student	***language country*** culture
grow	tall up plant	***plant up tall***	***plant tall up***
help	me wanted aid	***me aid*** *need*	***aid me*** *need*
social	security behavior people	*friends* club *media*	*friends media* family
telephone	call number talk	***call*** *mobile ring*	***call*** *mobile ring*
turn	off around on	***around*** *left* rotate	*left* ***around*** right
welcome	home mat hello	***home hello mat***	***home hello mat***
work	hard play job	***hard job*** *money*	***hard*** *money* ***job***

Backward collocation

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
bag	lunch paper sack	*hand purse shopping*	*hand purse shopping*
event	sport happening main	*party* birthday *festival*	*party festival* fun
field	football grass corn	***grass*** *green* **football**	***grass*** *green* cows
habit	bad smoking break	***bad smoking*** *routine*	***bad smoking*** *routine*
line	straight clothes fish	***straight*** *dot* fishing	*straight* draw *dot*
member	club group ship	***club*** *team* ***group***	***club group*** *team*
program	television computer guide	***television*** *tv* **computer**	*tv* ***television*** theatre
sign	stop name write	*language road post*	*road language post*
station	train gas radio	***train*** *bus police*	***train*** *bus police*
ticket	concert stub plane	*train* ***concert*** *bus*	*train* ***concert*** *bus*

Synonym

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
chance	luck opportunity risk	***luck opportunity*** rapper	***opportunity luck*** gamble
child	baby kid adult	***baby*** *young small*	*young* ***baby*** *small*
enjoy	fun like pleasure	***fun*** *happy life*	***fun*** *happy life*
have	own hold not	***hold*** *got* **not**	**own** ***hold*** *got*
important	urgent unimportant priority	*vital family* meeting	*family vital* crucial
local	near close city	*shop pub* town	*shop pub* home
mistake	wrong error correct	***error wrong*** *accident*	***wrong error*** *accident*
tell	say talk speak	*secret* ***speak talk***	*secret* ***speak talk***
total	sum whole all	***sum all*** number	***sum*** amount ***all***
wish	hope dream star	***dream star hope***	***dream hope star***

Antonym

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
arrive	leave depart late	train *time* ***late***	***late*** destination *time*
end	begin finish start	***start*** stop *beginning*	**finish** *beginning* ***start***
fat	skinny thin cat	***thin*** big weight	***thin*** food obese
friend	enemy pal buddy	*best love fun*	*best love fun*
healthy	sick good happy	*food* fit gym	*food* exercise diet
interesting	boring funny amusing	*fun fact* person	book *fact fun*
lose	win find weight	***win*** keys *lost*	***win*** game *lost*
quiet	loud noisy silent	***loud*** *library silence*	*library* **loud** *silence*
real	fake imaginary life	***fake life*** *authentic*	***fake life*** *authentic*
serious	funny joke silly	stern *face important*	*important face* work

Object property

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
belt	pants buckle loop	*trousers* ***buckle*** jeans	*trousers* ***buckle*** leather
burn	fire hurt hot	***fire hot*** sun	***fire hot*** **hurt**
name	person number mine	*first identity surname*	*identity surname first*
page	book paper turn	***book paper*** *turn*	***book paper*** number
root	tree plant canal	***plant tree*** *vegetable*	***plant tree*** *vegetable*
scene	play movie picture	*film* drama ***play***	***play*** *film* **movie**
soldier	army war gun	***war army*** strong	***war army*** fight
think	brain mind thought	***thought brain*** hard	***thought brain*** **mind**
toilet	bathroom seat flush	***bathroom*** loo paper	***bathroom*** sink **flush**
wings	bird fly chicken	***bird fly*** *angel*	***bird*** *angel* ***fly***

Cue property

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
beach	sand sun ocean	***sand sun*** *sea*	***sand sun*** *sea*
blood	red life donor	***red*** bank clot	***red*** death body
body	muscle arm fat	image **fat** mind	human legs **arms**
crime	bad jail criminal	*police prison* ***criminal***	*police prison* ***criminal***
danger	scary fear fire	*red warning* death	*red warning* run
elephant	big trunk mouse	*animal* **big** ***trunk***	*animal* grey ***trunk***
fire	hot water burn	***hot burn*** *danger*	***hot burn*** *danger*
model	beautiful pretty airplane	super **beautiful** *catwalk*	fashion clothes *catwalk*
trouble	bad problem shooting	***bad*** *maker naughty*	*naughty* ***bad*** *maker*
weekend	fun party relax	***fun*** *Saturday* ***relax***	***fun relax*** *Saturday*

Situational descriptor

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
bar	drink beer stool	***drink*** *alcohol* ***beer***	***drink*** *alcohol* ***beer***
boat	water sail ship	*sea* ***ship sail***	*sea* ***sail ship***
class	school student teacher	***school*** *people* university	***school*** **teacher** *people*
cloud	sky rain white	***sky rain*** water	***rain sky*** **white**
fish	water swim sea	***sea*** *food* ***water***	***sea*** *food* ***water***
hospital	sick nurse white	***nurse*** *bed doctor*	*doctor* ***nurse*** *bed*
measure	cup ruler length	*tape* ***ruler*** *weight*	***ruler*** *tape weight*
restaurant	food eat dinner	***food*** cafe ***eat***	***food*** drink ***eat***
swim	water dive pool	***water pool*** *sea*	***water pool*** *sea*
tourist	camera vacation visitor	*holiday* ***camera*** *travel*	*holiday* ***camera*** *travel*

Semantic field

Cue word	3 primary responses in USF norms	3 primary responses (first response only)	3 primary responses (all 3 responses)
church	God religion Sunday	***religion God*** pray	***religion God*** Jesus
damage	hurt broken break	***broken break*** control	***broken*** **hurt** ***break***
dream	sleep nightmare escape	***sleep*** big cloud	***sleep*** **nightmare** bed
education	school learn college	***school*** *university* important	***school*** *university* learn
family	love friend home	***friends love*** *mum*	***love friends*** *mum*
government	politics democracy law	***politics*** *parliament* ***law***	***politics law*** *parliament*
kiss	love hug lip	***love lips hug***	***love lips hug***
mirror	reflection image glass	***reflection image*** black	***reflection image*** **glass**
nature	tree forest birds	***trees*** *green* animal	***trees*** *green* outside
time	clock watch out	***clock watch*** *late*	***clock watch*** *late*

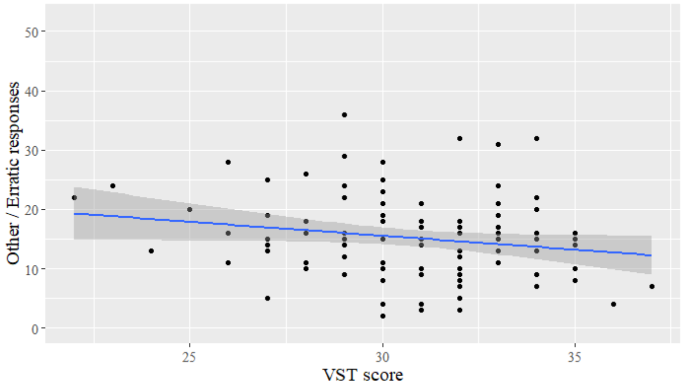
Appendix 10 – Cue Word Lists

Cue Word Lists			
1	**1 (cont.)**	**2**	**2 (cont.)**
station	fat	page	family
scene	think	welcome	chance
have	member	ticket	program
cloud	mistake	nature	church
social	fire	soldier	elephant
model	tourist	weekend	arrive
quiet	time	swim	wings
root	wish	real	class
government	blood	line	total
turn	bag	work	foreign
restaurant	end	body	lose
education	telephone	tell	hospital

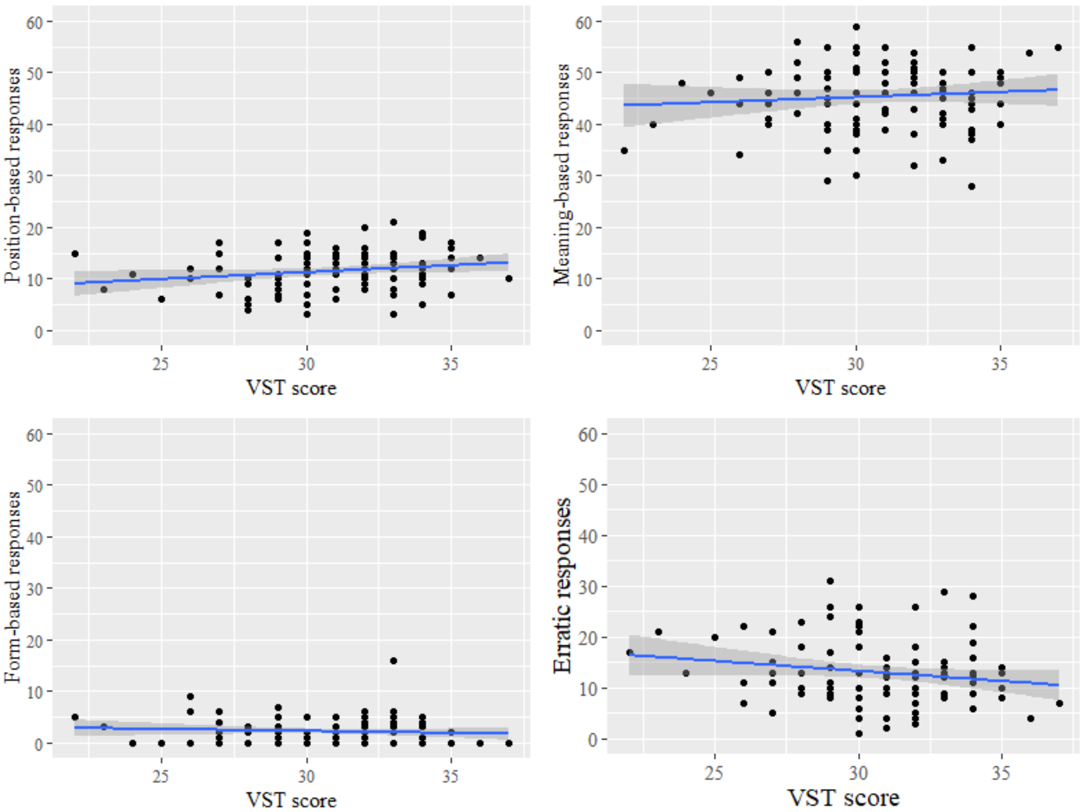
Note. cont. = continued. Each cue word list was presented in one list on one side of A4 paper. Each list was also created in reverse order.

Appendix 11 – Associations with VST Score

Association Between Other or Erratic Responses and VST Score



Association Between General Response Types in the Meaning-, position-, Form-based Coding System and VST Score



Notes. Due to the greater number of meaning-based responses, the y-axis in the scatterplots extends to 60 responses, rather than 50, as in the scatterplots shown in the main text.

Appendix 12 – Other Responses

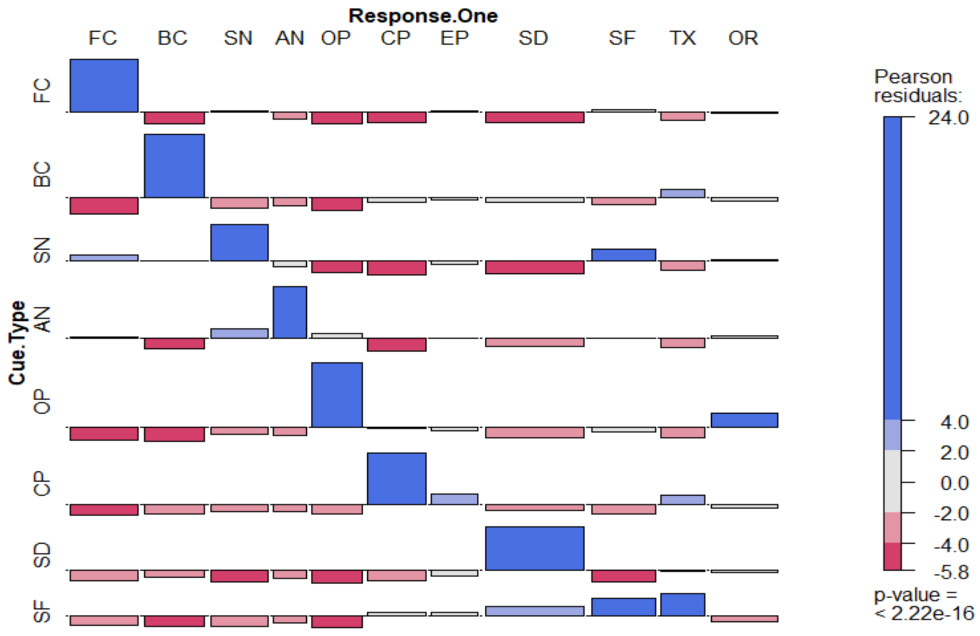
This table provides more detailed information about responses coded as other response within the conceptual-linguistic system. Linguistic Other = affix changes and responses based on form similarity; Conceptual Other = idiosyncratic and loosely connected responses; Other = chained responses, spelling mistakes, and associations given to a misinterpretation of the cue word (e.g., associates such as “way” and “car” to the cue word “root”, indicating it had been mistaken for “route”); instances in which no response was given are also reported separately.

Response Category	**Cue word concept type**								**Total**	
	Linguistic				**Conceptual**					
	FC	BC	SN	AN	OP	CP	SD	SF	No.	Prop.
Response one										
LO	11	3	11	7	7	1	10	2	52	0.021
CO	14	19	20	18	23	23	9	13	139	0.056
OR	3	2	0	10	24	2	10	0	51	0.020
NR	3	2	5	3	10	1	2	5	31	0.012
Total	31	26	36	38	64	27	31	20	273	0.109
All three responses										
LO	21	7	26	22	17	1	17	3	114	0.015
CO	88	88	121	102	116	78	50	70	713	0.095
OR	37	70	66	65	113	46	53	40	490	0.065
NR	26	28	51	38	55	18	14	26	256	0.034
Total	172	193	264	227	301	143	134	139	1573	0.209

Note. LO = linguistic other, CO = conceptual other, OR = other response, NR = no response, No. = number, Prop. = proportion.

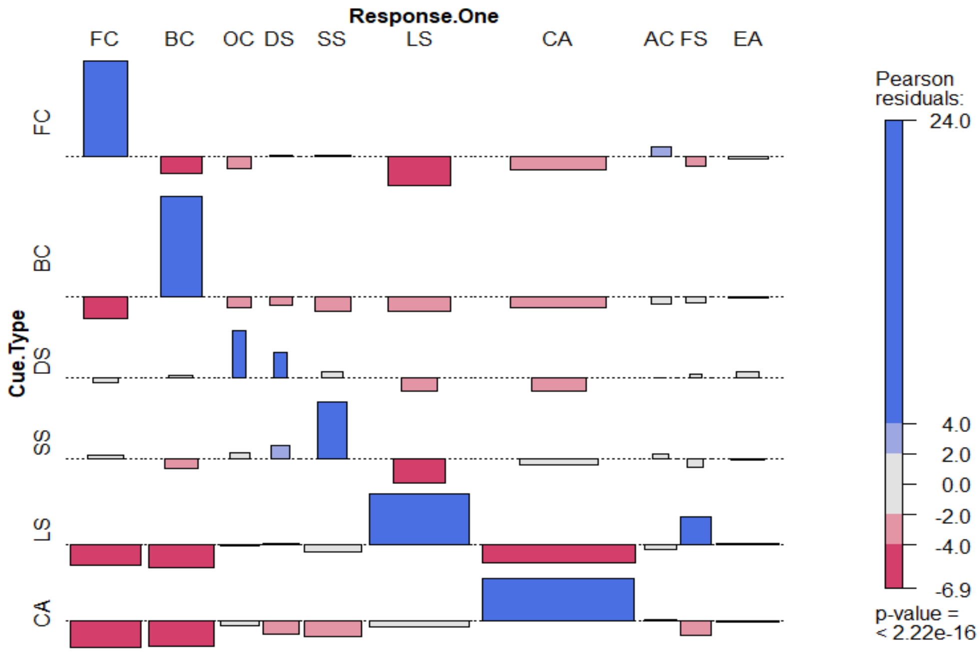
Appendix 13 - Association Between Cue and Response One

Association Between Cue and Response One Type (Linguistic-Conceptual)



Note. FC = forward collocation, BC = backward collocation, SN = synonym, AN = antonym, OP = stereotypical object property, CP = stereotypical cue property, EP = evaluative property, SD = situation descriptor, SF = semantic field, TX = Taxonomic, OR = Other.

Association Between Cue and Response One Type (Meaning, position, Form)



FC = forward collocation, BC = backward collocation, OC = other collocation; DS = defining synonym, SS = specific synonym, LS = lexical set, CA = conceptual association; AC = affix change, FS = form similarity; EA = erratic association.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review, 122*(3), 558–569. https://doi.org/10.1037/a0038693

Abrevaya, S., Sedeño, L., Fitipaldi, S., Pineda, D., Lopera, F., Buritica, O., Villegas, A., Bustamante, C., Gomez, D., Trujillo, N., Pautassi, R., Ibáñez, A., & García, A. M. (2017). The road less traveled: Alternative pathways for action-verb processing in Parkinson's disease. *Journal of Alzheimer's Disease*, *55*(4), 1429–1435. https://doi.org/10.3233/JAD-160737

Afiki, Y., Bar, M. (2020). Our need for associative coherence. *Humanities and Social Sciences Communications, 7*, Article 80. https://doi.org/10.1057/s41599-020-00577-w

Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition*, *69*(2), 135–178. https://doi.org/10.1016/s0010-0277(98)00063-8

Aiello, E. N., Grosso, M., Di Liberto, A., Andriulo, A., Buscone, S., Caracciolo, C., Ottobrini, M. & Luzzatti, C. (2022). Disembodying language: Actionality does not account for verb processing deficits in Parkinson's disease. *Journal of Neurolinguistics*, *61*, Article 101040. https://doi.org/10.1016/j.jneuroling.2021.101040

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon*. (3rd ed.). Blackwell Publishing Ltd.

Anderson, A. J., Bruni, E., Bordignon, U., Poesio, M., & Baroni, M. (2013). Of words, eyes and brains: Correlating image-based distributional semantic models with neural representations of concepts. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing* (pp. 1960–1970), Seattle, WA. http://www.aclweb.org/anthology/D13-1202

Anderson, R. C, Pichert, J. W, Goetz, E. T, Schallert, D. L, Stevens, K. V, & Trollip, S. R. (1976). Instantiation of general terms. *Journal of Verbal Learning & Verbal Behavior, 15*(6), 667–679. https://doi.org/10.1016/0022-5371(76)90059-1

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review, 116*(3), 463–498. https://doi.org/10.1037/a0016261

Argiris, G., Budai, R., Maieron, M., Ius, T., Skrap, M., & Tomasino, B. (2020). Neurosurgical lesions to sensorimotor cortex do not impair action verb processing. *Scientific Reports, 10,* Article 523. https://doi.org/10.1038/s41598-019-57361-3

Asquith, S. (2022). An investigation into the roles of guessing and partial knowledge in the Vocabulary Size Test. *TESL-EJ, 26*(3), 1–31. https://doi.org/10.55593/ej.26103a15

Baayen, R. H., Chuang, Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*. https://doi.org/10.1155/2019/4895891

Baror, S., & Bar, M. (2016). Associative activation and its relation to exploration and exploitation in the brain. *Psychological science*, *27*(6), 776–789. https://doi.org/10.1177/0956797616634487

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10*(1), 82–93. https://doi.org/10.3758/BF03197629

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*(3), 211–227. https://doi.org/10.3758/BF03196968

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer and E. F. Kittay (Eds.), Frames, fields and contrasts (pp. 21–74). Lawrence Erlbaum.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/s0140525x99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B, 364*, 1281–1289. https://doi.org/10.1098/rstb.2008.0319

Barsalou L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, *23*(4), 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Barsalou, L. W. (2021). Categories at the interface of cognition and action. In C. Mauri, I. Fiorentini, & E. Goria (Eds.), Building categories in interaction: Linguistic resources at work. John Benjamins. https://doi.org/10.1075/slcs.220

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). *Language and simulation in conceptual processing*. In M. De Vega, A.M. Glenberg, & A. C.

Graesser (Eds.), Symbols and embodiment: Debates on meaning and cognition (pp. 245-283). Oxford University Press.

Bauer, A. J., & Just, M.A. (2017). A brain-based account of "basic-level" concepts. *NeuroImage, 161,* 196-205. https://doi.org/10.1016/j.neuroimage.2017.08.049

Bauer, A. J., & Just, M.A. (2019). Neural Representations of Concept Knowledge. In G. I. de Zubicaray & N. O. Schiller (Eds), The Oxford Handbook of Neurolinguistics (pp. 519–547). Oxford Academic. https://doi.org/10.1093/oxfordhb/9780190672027.013.21

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279. https://doi.org/10.26686/wgtn.12560408.v1

Bedny, M., & Saxe, R. (2012). Insights into the origins of knowledge from the cognitive neuroscience of blindness. *Cognitive neuropsychology*, *29*(1-2), 56–84. https://doi.org/10.1080/02643294.2012.713342

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101–118. https://doi.org/10.1177/0265532209340194

Bel Enguix, G., Reinhard Rapp, R., & Zock,M. (2014). How well can a corpus-derived co-occurrence network simulate human associative behavior? In *Proceedings of the 5th Workshop on Cognitive Aspects of Computational Language Learning (CogACLL)*, 43–48. Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-0509

Bencini, G. M. L., & Goldberg, A. E. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, *43*(4), 640–651. https://doi.org/10.1006/jmla.2000.2757

Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, *82*(2), 233–278. http://www.jstor.org/stable/4490157

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3-4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blackett, D. S., Varkey, J., Wilmskoetter, J., Roth, R., Andrews, K., Busby, N., Gleichgerrcht, E., Desai, R. H., Riccardi, N., Basilakos, A., Johnson, L.P., Kristinsson, S., Johnson, L., Rorden, C., Spell, L. A., Fridriksson, J., & Bonilha, L. (2022). Neural network bases of thematic semantic processing in language production. *Cortex*, *156*, 126–143. https://doi.org/10.1016/j.cortex.2022.08.007

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of life reviews*, *29*, 120–153. https://doi.org/10.1016/j.plrev.2018.12.001

Bottini, R., Ferraro, S., Nigri, A., Cuccarini, V., Bruzzone, M. G., & Collignon, O. (2020). Brain regions involved in conceptual retrieval in sighted and blind people. *Journal of Cognitive Neuroscience*, *32*(6), 1009–1025. https://doi.org/10.1162/jocn_a_01538

Bottini, R., Morucci, P., D'Urso, A., Collignon, O., & Crepaldi, D. (2022). The concreteness advantage in lexical decision does not depend on perceptual simulations. *Journal of Experimental Psychology: General*, *151*(3), 731–738. https://doi.org/10.1037/xge0001090

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*(9), 1207–15. https://doi.org/10.1016/S0028-3932(00)00034-8

Bréal, M. (1897/1900). *Essai de semantique (science des significations)* / Semantics: Studies in the Science of Meaning. Librairie Hachette. https://archive.org/details/semanticsstudie00postgoog

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–49. https://doi.org/10.1016/j.cognition.2014.10.017

Brotsky, S. J., & Linton, M. L. (1967). The test-retest reliability of free association norms. *Psychonomic Science, 8,* 425–426. https://doi.org/10.3758/BF03332273

Brown, R., & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development, 31*(1), 1–14. https://doi.org/10.2307/1126377

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., …Dario Amodei. (2020). Language models are few-shot learners. In *Proceedings of the 34th*

International Conference on Neural Information Processing Systems (NIPS'20). Article 159, 1877–1901.

Browne, C., Culligan, B., & Phillips, J. (2013). The New General Service List. http://www.newgeneralservicelist.org

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(3), 359–371. https://doi.org/10.1017/S1366728909990344

Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of google books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, *2*, 27. https://doi.org/10.3389/fpsyg.2011.00027

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Burke, D. M., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology and Aging*, *1*(4), 283–292. https://doi.org/10.1037/0882-7974.1.4.283

Buzsáki, G. (2019). *The brain from inside out*. Oxford University Press. https://doi.org/10.1093/oso/9780190905385.001.0001

Capurro, R., & Hjørland, B. (2003). The concept of information. *Annual Review of Information Science and Technology*, *37*(1), 343–411. https://doi.org/10.1002/aris.1440370109

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

Cardona, J. F., Gershanik, O., Gelormini-Lezama, C., Houck, A. L., Cardona, S., Kargieman, L., Trujillo, N., Arévalo, A., Amoruso, L., Manes, F., & Ibáñez, A. (2013). Action-verb processing in Parkinson's disease: New pathways for motor-language coupling. *Brain Structure & Function*, *218*(6), 1355–1373. https://doi.org/10.1007/s00429-013-0510-1

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Cayol, Z., & Nazir, T. A. (2020). Why language processing recruits modality specific brain regions: It is not about understanding words, but about modelling situations. *Journal of Cognition*, *3*(1), Article 35. https://doi.org/10.5334/joc.124

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition, 25*(2), 203–226. https://doi.org/10.3758/BF03201113

Charles, W., & Miller, G. (1989). Contexts of antonymous adjectives. *Applied Psycholinguistics, 10*(3), 357–375. https://doi.org/10.1017/S0142716400008675

Chemero, A. (2011). *Radical embodied cognitive science*. MIT Press.

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29. https://aclanthology.org/J90-1003

Chwilla, D. J., & Kolk, H. H. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, *25*(3), 589–606. https://doi.org/10.1016/j.cogbrainres.2005.08.011

Cisek, P. (1999) Beyond the computer metaphor: Behaviour as interaction. *Journal of Consciousness Studies*, *6*(11-12), 125–42.

Clark, A. (2006). Material symbols. *Philosophical Psychology, 19*(3), 291–307. https://doi.org/10.1080/09515080600689872

Clark, H. H. (1970). Word associations and linguistic theory. In J. Lyons (Ed.), New Horizons in Linguistics. Penguin.

Clerkin, E. M., & Smith, L. B. (2022). Real-world statistics at two timescales and a mechanism for infant learning of object names. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(18), Article e2123239119. https://doi.org/10.1073/pnas.2123239119

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–9. https://doi.org/10.1037/0033-2909.112.1.155

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior, 8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Committee of Revising JACET Basic Words (Ed.). (2003). *JACET list of 8000 basic words*. Japan Association of College English Teachers.

Conklin, K., & Carrol, G. (2021). Words go together like ‘bread and butter’: The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*, *42*(3), 492–513. https://doi.org/10.1093/applin/amaa034

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience, 34*(10), 1308–1318. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic Bulletin & Review*, *16*(3), 573–577. https://doi.org/10.3758/PBR.16.3.573

Connell, L. & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–550. https://doi.org/10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. https://doi.org/10.1111/tops.12097

Connell, L., Lynott, D., & Banks, B. (2018). Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *373*(1752), Article 20170143. https://doi.org/10.1098/rstb.2017.0143

Conway C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, *112*, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032

Couchman, J. J., Boomer, J., Coutinho, M. V., & Smith, J. D. (2010). Carving nature at its joints using a knife called concepts. *The Behavioral and Brain Sciences*, *33*(2–3), 207–208. https://doi.org/10.1017/S0140525X10000336

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 129–147. https://doi.org/10.1037/0278-7393.31.1.129

Cox, C. R., & Haebig, E. (2023). Child-oriented word associations improve models of early word learning. *Behavior Research Methods*, *55*(1), 16–37. https://doi.org/10.3758/s13428-022-01790-y

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Cramer, P. (1968). *Word association*. Academic Press. https://archive.org/details/wordassociation0000cram/page/30/mode/2up

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science, 23*(3), 371–414. https://doi.org/10.1016/S0364-0213(99)00005-1

Cremer, M., Dingshoff, D., de Beer, M., & Schoonen, R. (2011). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism, 15*(2), 187–204. https://doi.org/10.1177/1367006910381189

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

Dąbrowska, E. (2014). Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon*, *9*(3), 401–418. https://doi.org/10.1075/ml.9.3.02dab

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition, 178*, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

Dąbrowska, E. (2019). Experience, aptitude, and individual differences in linguistic attainment: A comparison of native and nonnative speakers. *Language Learning, 69*, 72–100. https://doi.org/10.1111/lang.12323

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1-2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-x

Davies, M. (2004) *British National Corpus* (from Oxford University Press). https://www.english-corpora.org/bnc/.

Davies, M. (2008-) *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/.

Davis, C. P., Joergensen, G. H., Boddy, P., Dowling, C., & Yee, E. (2020). Making it harder to 'see' meaning: The more you see something, the more its conceptual representation is susceptible to visual interference. *Psychological Science, 31*(5), 505–517. https://doi.org/10.1177/0956797620910748

Davis, C.P., & Yee, E. (2021). Building semantic memory from embodied and distributional language experience. *WIREs Cognitive Science*, *12*(5), Article e1555. https://doi.org/10.1002/wcs.1555

Deacon, T. (1997). *The symbolic species: The co-evolution of language and the brain*. W. W. Norton & Co.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

De Deyne, S., Navarro, D. & Storms, G. (2013). Associative strength and semantic activation in the mental lexicon: Evidence from continued word associations. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*. https://escholarship.org/uc/item/1596128p

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231. https://doi.org/10.3758/BRM.40.1.213

De Deyne, S., Verheyen, S., Perfors, A., & Navarro, D. J. (2015). Evidence for widespread thematic structure in the mental lexicon. *Cognitive Science Society Proceedings*, 518–523.

De Deyne, S., Verheyen, S. & Storms, G. (2016). Structure and organization of the mental lexicon: A network approach derived from syntactic dependency relations and word associations. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, B. Job (Eds.), Towards a theoretical framework for analyzing complex linguistic networks. Understanding complex systems (pp. 47-79). Springer. https://doi.org/10.1007/978-3-662-47238-5_3

Deese, J. (1965). *The structure of associations in language and thought*. Johns Hopkins Press.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 824–845. https://doi.org/10.1037/0278-7393.15.5.824

de Mornay Davies, P. (1998). Automatic semantic priming: The contribution of lexical- and semantic-level processes. *European Journal of Cognitive Psychology, 10*(4), 389–412. https://doi.org/10.1080/713752286

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences, 26*(9), 751–766. https://doi.org/10.1016/j.tics.2022.06.010

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's *t*-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America, 116*(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

Derrah, R., & Rowe, D. (2015). Validating the Japanese bilingual version of the Vocabulary Size Test. *International Journal of Languages, Literature and Linguistics, 1*, 131–135. https://doi.org/10.7763/IJLLL.2015.V1.26

Dieciuc, M. A., and Folstein, J. R. (2019). Typicality: Stable structures and flexible functions. *Psychonomic Bulletin & Review, 26*, 491–505. https://doi.org/10.3758/s13423-018-1546-2

Dove G. (2011). On the need for Embodied and Dis-Embodied Cognition. *Frontiers in Psychology, 1*, Article 242. https://doi.org/10.3389/fpsyg.2010.00242

Dove G. (2020). More than a scaffold: Language is a neuroenhancement. *Cognitive Neuropsychology, 37*(5-6), 288–311. https://doi.org/10.1080/02643294.2019.1637338

Dove, G., Barca, L., Tummolini, L., & Borghi, A. M. (2022). Words have a weight: Language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research, 86*(8), 2451–2467. https://doi.org/10.1007/s00426-020-01438-6

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr, Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1-2), 145–177. https://doi.org/10.1016/j.cognition.2003.11.002

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Dymarska, A., Connell, L., & Banks, B. (2023). Weaker than you might imagine: Determining imageability effects on word recognition. *Journal of Memory and Language, 129*, Article 104398. https://doi.org/10.1016/j.jml.2022.104398

Edmonds, A., Clenton, J., & Elmetaher, H. (2022). Exploring the construct validity of tests used to assess L2 productive vocabulary knowledge. *System, 108*, Article 102855. https://doi.org/10.1016/j.system.2022.102855

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, *30*(2), 253–272. https://doi.org/10.1177/0265532212459028

Ellis, N. C. (2019). Essentials of a Theory of Language Cognition. *Modern Language Journal*, *103,* 39–60. https://doi.org/10.1111/modl.12532

Entwisle, D. R. (1966). *The word associations of young children*. Johns Hopkins Press.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *The American Journal of Psychology*, *74*(3), 361-372. https://doi.org/10.2307/1419742

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: The apprehension and consequences of thematic relations. In B. H. Ross (Ed.), The psychology of learning and motivation: Advances in research and theory (pp. 249–294). Elsevier Academic Press.

Evans, V. (2009). *How words mean: Lexical concepts, cognitive models and meaning construction*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199234660.001.0001

Evans, V. (2016). Design features for linguistically-mediated meaning construction: The relative roles of the linguistic and conceptual systems in subserving the ideational function of language. *Frontiers in Psychology*, *7,* Article 156. https://doi.org/10.3389/fpsyg.2016.00156

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Fenlon, J., Cormier, K., & Brentari, D. (2017). Sign language phonology. In S. J. Hannahs & A. Bosch (Eds.), The Routledge handbook of phonological theory (pp. 453–475). Routledge. https://doi.org/10.4324/9781315675428

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Fernandino, L., Tong, J. Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(6), Article e2108091119. https://doi.org/10.1073/pnas.2108091119

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences, 280*(1), 20–32. https://doi.org/10.1111/j.1749-6632.1976.tb25467.x

Fillmore, C. J. (1982). Frame semantics. In *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, 111–137. Hanshin. http://brenocon.com/Fillmore%201982_2up.pdf

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In J. R. Firth (Ed), Studies in linguistic analysis, pp. 1-32. Special volume of the Philological Society. Blackwell.

Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. In S. H. Foster-Cohen, M. Medved Krajnovic, & J. Mihaljevic Djigunovic (Eds.), EUROSLA Yearbook 6, (pp. 121–45). https://doi.org/10.1075/eurosla.6.09fit

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics,* 17(3), 319–31. https://doi.org/10.1111/j.1473-4192.2007.00172.x

Fitzpatrick, T. (2009). Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick & A. Barfield (Eds.), Lexical Processing in Second Language Learners (pp. 38–52). https://doi.org/10.21832/9781847691538-006

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition,* 33(3), 373–398. https://doi.org/10.1017/S0272263111000027

Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. J. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics, 36*(1), 23–50. https://doi.org/10.1093/applin/amt020

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237–274. https://doi.org/10.1017/S0261444820000105

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (2015). *Minds without meanings: An essay on the content of concepts*. MIT Press. https://doi.org/10.7551/mitpress/9780262027908.001.0001

Francis, W. & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Houghton Mifflin.

Friston, K. J. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gabora, L., Rosch, E., & Aerts, D. (2008). Toward an ecological theory of concepts. *Ecological Psychology, 20*(1), 84-116. https://doi.org/10.1080/10407410701766676

Gainotti, G. (2017). The differential contributions of conceptual representation format and language structure to levels of semantic abstraction capacity. *Neuropsychology Review*, *27*(2), 134–145. https://doi.org/10.1007/s11065-016-9339-8

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. https://doi.org/10.3758/bf03193857

Gallese, V., & Cuccio, V. (2018). The neural exploitation hypothesis and its implications for an embodied approach to language and cognition: Insights from the study of action verbs processing and motor disorders in Parkinson's disease. *Cortex*, *100*, 215–225. https://doi.org/10.1016/j.cortex.2018.01.010

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, *22*(3), 455–479. https://doi.org/10.1080/02643290442000310

Galton, F. (1879). Psychometric experiments. *Brain*, *2*, 149–162.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press. https://doi.org/10.7551/mitpress/9629.001.0001

Gaston, P., Brodbeck, C., Phillips, C., & Lau, E. (2023). Auditory word comprehension is less incremental in isolated words. *Neurobiology of Language*, *4*(1), 29–52. https://doi.org/10.1162/nol_a_00084

Gefter, A. (2015). The man who tried to redeem the world with logic. *Nautilus*, *21*. https://nautil.us/the-man-who-tried-to-redeem-the-world-with-logic-235253/

Geng, J., & Schnur, T. T. (2016). Role of features and categories in the organization of object knowledge: Evidence from adaptation fMRI. *Cortex*, *78*, 174–194. https://doi.org/10.1016/j.cortex.2016.01.006

Gibson, J. J. (1979/1986). *The ecological approach to visual perception*. Lawrence Erlbaum Associates.

Giraudoux, P. (2022). pgirmess: Spatial analysis and data mining for field ecologists. R package version 2.0.0. https://CRAN.R-project.org/package=pgirmess

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*(1-3), 61–105. https://doi.org/10.1016/0010-0277(92)90040-o

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–922. https://doi.org/10.1016/j.cortex.2011.04.010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. https://doi.org/10.3758/bf03196313

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press. https://doi.org/10.2307/j.ctvc772nn

Goldberg, A. E., & Lee, C. (2021). Accessibility and historical change: An emergent cluster led *uncles and aunts* to become *aunts and uncles*. *Frontiers in Psychology*, *12*, Article 662884. https://doi.org/10.3389/fpsyg.2021.662884

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 959–978. https://doi.org/10.3758/s13423-015-0860-1

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Avila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*(2), 906–912. https://doi.org/10.1016/j.neuroimage.2006.03.037

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, *166*(1), 94–126. https://doi.org/10.1075/itl.166.1.03fer

Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology*, *61*(3), 348–353. https://doi.org/10.1037/0021-9010.61.3.348

Gruenenfelder, T. M., Recchia, G., Rubin, T., & Jones, M. N. (2016). Graph-theoretic properties of networks based on word association norms: Implications for models of lexical semantic memory. *Cognitive Science*, *40*(6), 1460–1495. https://doi.org/10.1111/cogs.12299

Halliday, M. A. K. (2004). *The Language of Science: Volume 5 (Collected Works of M.A.K. Halliday).* Jonathan J. Webster (Ed.). Bloomsbury Publishing.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*(3), 355–384. https://doi.org/10.1080/15326900701326402

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., & Pietrini, P. (2016). How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage*, *135*, 232–242. https://doi.org/10.1016/j.neuroimage.2016.04.063

Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. MIT Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*(1-3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Harpaintner, M., Trumpp, N. M., & Kiefer, M. (2018). The semantic content of abstract concepts: a property listing study of 296 abstract words. *Frontiers in Psychology, 9*, Article 1748. https://doi.org/10.3389/fpsyg.2018.01748

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representations of action words in human motor and premotor cortex. *Neuron, 41,* 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Healy, K. (2019). *Data visualization: A practical introduction*. Princeton University Press.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 411–422. https://doi.org/10.1037//0278-7393.20.2.411

Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), Vocabulary and writing in a first and second language: Processes and development (pp. 22–66). https://doi.org/10.1057/9780230593404

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124*(1), 62–82. https://doi.org/10.1037/0096-3445.124.1.62

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review, 119*(2), 431–440. https://doi.org/10.1037/a0027373

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science, 7*(3), 513–534. https://doi.org/10.1111/tops.12151

Hirsh, K. W., & Tree, J. J. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics, 14*(1), 1–44. https://doi.org/10.1016/S0911-6044(00)00002-6

Hoedemaker, R. S., & Gordon, P. C. (2017). The onset and time course of semantic priming during rapid recognition of visual words. *Journal of Experimental Psychology: Human Perception and Performance, 43*(5), 881–902. https://doi.org/10.1037/xhp0000377

Humboldt, W. (1836/1999*). On language: On the diversity of human language construction and its influence on the mental development of the human species*. Cambridge University Press.

Humphries, M. (2021). *The spike: An epic journey through the brain in 2.1 seconds*. Princeton University Press.

Hurley, S. (2005). The shared circuits hypothesis: A unified functional architecture for control, imitation, and simulation. In S. Hurley & N. Chater (Eds.), Perspectives on imitation: From neuroscience to social science - Volume 1: Mechanisms of Imitation and Imitation in Animals (pp.177–194). MIT Press. https://doi.org/10.7551/mitpress/5330.003.0008

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10*, 785–813. https://doi.org/10.3758/BF03196544

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., Yap, M. J., Bengson, J. J., Niemeyer, D., & Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods, 45*(4), 1099–1114. https://doi.org/10.3758/s13428-012-0304-z

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron, 76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*(1651), Article 20130298. https://doi.org/10.1098/rstb.2013.0298

James, T. W., & Gauthier, I. (2003). Auditory and action semantic features activate sensory-specific perceptual brain regions. *Current Biology*, *13*(20), 1792–1796. https://doi.org/10.1016/j.cub.2003.09.039

Jardak, A., & Byers-Heinlein, K. (2019). Labels or concepts? The development of semantic networks in bilingual two-year-olds. *Child Development*, *90*(2), e212–e229. https://doi.org/10.1111/cdev.13050

Jenkins, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), Norms of word association (pp. 1–38). Academic Press.

Jenkins, J. J., Mink, W. D., & Russell, W. A. (1958). Associative clustering as a function of verbal association strength. *Psychological Reports*, *4*(1), 127–136. https://doi.org/10.2466/pr0.1958.4.g.127

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

Johnson, M. (1990). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Johnson, M. A., & Goldberg, A. E. (2013) Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes*, *28*(10). 1439–1452. https://doi.org/10.1080/01690965.2012.717632

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge University Press.

Jones, L. L., & Golonka, S. (2012). Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Frontiers in Human Neuroscience, 6,* Article 205. https://doi.org/10.3389/fnhum.2012.00205

Jones, L. L., Wurm, L. H., Calcaterra, R. D., & Ofen, N. (2017). Integrative priming of compositional and locative relations. *Frontiers in Psychology*, *8*, Article 359. https://doi.org/10.3389/fpsyg.2017.00359

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131*(5), 684–712. https://doi.org/10.1037/0033-2909.131.5.684

Just, M. A., Cherkassky, V. L, Aryal, S, & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE, 5*(1), Article e8622. https://doi.org/10.1371/journal.pone.0008622

Just, M. A., Wang, J., & Cherkassky, V. L. (2017). Neural representations of the concepts in simple sentences: Concept activation prediction and context effects. *NeuroImage, 157*, 511–520. https://doi.org/10.1016/j.neuroimage.2017.06.033

Justeson, J. S., & Katz. S. M. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics, 17*(1), 1–20. https://aclanthology.org/J91-1001

Kaan, E. (2014) Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism, 4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., & Baciu, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: A behavioral and fMRI study. *NeuroImage*, *44*(3), 1152–1162. https://doi.org/10.1016/j.neuroimage.2008.09.043

Kang, B. M. (2018). Collocation and word association: Comparing collocation measuring methods. *International Journal of Corpus Linguistics*, *23*(1), 85–113. https://doi.org/10.1075/ijcl.15116.kan

Keil, F. C., & Kominsky, J. F. (2015). Grounding concepts. In E. Margolis, & S. Laurence (Eds.), The conceptual mind: New directions in the study of concepts (pp. 677–692). MIT Press.

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: Event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research, 10*(1-2), 67–75. https://doi.org/10.1016/S0926-6410(00)00023-9

Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Psychiatry*, *67*(1), 37–96. https://doi.org/10.1037/13767-000 https://gutenberg.org/ebooks/9172

Kewenig, V., Vigliocco, G., & Skipper, J. (2022). When abstract becomes concrete: Naturalistic encoding of concepts in the brain. bioRxiv. https://doi.org/10.1101/2022.09.08.506944

Kiefer, M., & Harpaintner, M. (2020). Varieties of abstract concepts and their grounding in perception or action. *Open Psychology, 2*(1), 119–137. https://doi.org/10.1515/psych-2020-0104

Kim, M. S. (2013). The mental lexicon of low-proficiency Korean heritage learners. *Heritage Language Journal, 10*(1), 17–35. https://doi.org/10.46538/hlj.10.1.2

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith, (Eds.), The computer and literary studies. Edinburgh University Press.

Kittay, E. F., & Lehrer A. (1992). Introduction. In: A. Lehrer and E. F. Kittay (Eds.), Frames, fields and contrasts (pp. 1–20). Lawrence Erlbaum.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112*(3), 500–526. https://doi.org/10.1037/0033-2909.112.3.500

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179088.001.0001

Kruse, H., Pankhurst, J., & Smith, M. S. (1987). A multiple word association probe in second language acquisition research. *Studies in Second Language Acquisition, 9*(2), 141–154. https://doi.org/10.1017/S0272263100000449

Kuhnke, P., Kiefer, M., & Hartwigsen, G. (2020). Task-dependent recruitment of modality-specific and multimodal regions during conceptual processing. *Cerebral cortex*, *30*(7), 3938–3959. https://doi.org/10.1093/cercor/bhaa010

Kuhnke, P., Kiefer, M., & Hartwigsen, G. (2021). Task-dependent functional and effective connectivity during conceptual processing. *Cerebral cortex, 31*(7), 3475–3493. https://doi.org/10.1093/cercor/bhab026

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Kumar, A. A., Steyvers, M., & Balota, D. A. (2022). A critical review of network-based and distributional approaches to semantic memory structure and processes. *Topics in Cognitive Science, 14*(1), 54–77. https://doi.org/10.1111/tops.12548

Kussmaul, A. (1877). Disturbances of speech. In H. von Ziemssen (Ed.) & J.A. McCreery (Trans.), Cyclopedia of the practice of medicine, volume XIV: Diseases of the nervous system and disturbances of speech (pp. 581–875). https://archive.org/details/cyclopaediaofpra14ziemuoft

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539–550. https://doi.org/10.3758/bf03196991

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* The University of Chicago Press.

Lake, B. M., & Murphy, G. L. (2023). Word meaning in minds and machines. *Psychological Review, 130*(2), 401–431. https://doi.org/10.1037/rev0000297

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*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior: The Hixon Symposium (p. 112–146). Wiley.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners of English. *Language Learning, 61*, 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: a set of semantic feature norms from the congenitally blind. *Behavior Research Methods, 45*(4), 1218–1233. https://doi.org/10.3758/s13428-013-0323-4

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins. https://doi.org/10.1075/z.195

Lewis, G. A., Poeppel, D., & Murphy, G. L. (2017). *Neural mechanisms of conceptual relations*. New York University, ProQuest Dissertations Publishing. https://search.proquest.com/docview/1879433419

Li, D., Zhang, X., & Wang, G. (2011). Senior Chinese high school students' awareness of thematic and taxonomic relations in L1 and L2. *Bilingualism: Language and Cognition, 14*(4), 444–457. https://doi.org/10.1017/S1366728910000416

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *130*(1), 3–28. https://doi.org/10.1037/0096-3445.130.1.3

Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. Routledge. https://doi.org/10.4324/9780203342763

Liu, C., Cohn, T., De Deyne, S., & Frermann, L. (2022). WAX: A new dataset for word association eXplanations. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers),* 106–120. https://aclanthology.org/2022.aacl-main.9

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3,* 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M. (2018). Knowing the meaning of a word by the linguistic and perceptual company it keeps. *Topics in Cognitive Science, 10*(3), 573–589. https://doi.org/10.1111/tops.12349

Louwerse, M. M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science, 35*(2), 381–398. https://doi.org/10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2015). Effect size matters: The role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience, 30*(4), 430–447. https://doi.org/10.1080/23273798.2014.981552

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition, 114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin & Review, 7*(4), 618–630. https://doi.org/10.3758/bf03212999

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In J. D. Moore & J. F. Lehman (Ed.), Proceedings of the 17th annual meeting of the Cognitive Science Society (pp. 660–665). Erlbaum.

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience, 34*(10), 1319–1337. https://doi.org/10.1080/23273798.2017.1404114

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. M. Lopez-Morillas & L. Solotaroff (Trans.). Harvard University Press.

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology, 1,* Article 212. https://doi.org/10.3389/fpsyg.2010.00212

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Ma, X. (2013). Evocation: Analyzing and propagating a semantic link based on free word association. *Language Resources and Evaluation, 47*(3), 819–837. https://doi.org/10.1007/s10579-013-9219-2

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs* (3rd ed.). Lawrence Erlbaum.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, 102*(1-3), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Maki, W. S., & Buchanan, E. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review, 15*(3), 598–603. https://doi.org/10.3758/pbr.15.3.598

Mandelbaum, E., Dunham, Y., Feiman, R., Firestone, C., Green, E. J., Harris, D., Kibbe, M. M., Kurdi, B., Mylopoulos, M., Shepherd, J., Wellwood, A., Porot, N., & Quilty-Dunn, J. (2022). Problems and mysteries of the many languages of thought. *Cognitive Science, 46*(12), Article 13225. https://doi.org/10.1111/cogs.13225

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92,* 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Mandler, J. M. (2007). *The foundations of mind: Origins of conceptual thought*. Oxford University Press.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in infancy. *Psychological Science, 21,* 908–913. https://doi.org/10.1177/0956797610373371

Mareschal, D., Quinn, P. C., & Lea, S. E. G. (Eds.). (2010). *The making of human concepts*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199549221.001.0001

Margolis, E., & Laurence, S. (2015). *The conceptual mind: New directions in the study of concepts*. MIT Press.

Marti, L., Shengyi Wu, S., Piantadosi, S. T., & Kidd, C. (2023). Latent diversity in human concepts. *Open Mind: Discoveries in Cognitive Science, 7*, 79–92. https://doi.org/10.1162/opmi_a_00072

Martin, A. (1998). The organization of semantic knowledge and the origin of words in the brain. In N. G. Jablonski, & L. C. Aiello (Eds.), The origins and diversification of language (pp. 69–88). California Academy of Sciences.

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*, *23*, 979–990 https://doi.org/10.3758/s13423-015-0842-3

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*(5233), 102–105. https://doi.org/10.1126/science.270.5233.102

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*(6566), 649–652. https://doi.org/10.1038/379649a0

Maturana, H. (1970). *Biology of cognition*. Urbana: Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois. http://reflexus.org/wp-content/uploads/BoC.pdf

McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133. https://doi.org/10.1007/BF02478259

McEvoy, C. L., & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *The American Journal of Psychology, 95*(4), 581–634. https://doi.org/10.2307/1422189

McGregor, S., K. Agres, M. Purver, and G. Wiggins (2015). From distributional semantics to conceptual spaces: A novel computational method for concept creation. *Journal of Artificial General Intelligence, 6*(1), 55–89. https://doi.org/10.1515/jagi-2015-0004

McKoon, G., & Ratcliff, R. (1995). Conceptual combinations and relational contexts in free association and in priming in lexical decision and naming. *Psychonomic Bulletin & Review, 2*(4), 527–533. https://doi.org/10.3758/BF03210988

McLean, S., Kramer, B., & Stewart, J. (2015). An empirical examination of the effect of guessing on vocabulary size test scores. *Vocabulary Learning and Instruction, 4,* 26–35. http://vli-journal.org/wp/vli-v04-2-2187- 2759/

McNorgan, C., Reid, J., & McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. *Cognition, 118*(2), 211–233. https://doi.org/10.1016/j.cognition.2010.10.017

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), The adolescent brain: Learning, reasoning, and decision making (pp. 39–66). American Psychological Association.

McRae, K., Nedjadrasul, D., Pau, R., Lo, B. P. H., & King, L. (2018). Abstract concepts and pictures of real-world situations activate one another. *Topics in Cognitive Science, 10,* 518–532. https://doi.org/10.1111/tops.12328

Meara, P. (1978). Learners' Word Associations in French. *Interlanguage Studies Bulletin, 3*(2), 192–211. http://www.jstor.org/stable/43136005

Meara, P. M. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular, 11*(2), 29–38.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Merleau-Ponty, M. (1945/2005). *Phenomenology of perception*. C. Smith (Trans.). Routledge.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meyer, A.S., Belke, E., Telling, A. L., & Humphreys, G.W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*(4), 710–716. https://doi.org/10.3758/bf03196826

Meyer, D, Zeileis, A, & Hornik, K. (2006). "The Strucplot framework: Visualizing multi-way contingency tables with vcd." *Journal of Statistical Software, 17*(3), 1–48. doi:10.18637/jss.v017.i03.

Meyer, D, Zeileis A, & Hornik, K. (2022). *vcd: Visualizing categorical data*. R package version 1.4–10.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235–244. https://doi.org/10.1093/ijl/3.4.235

Mirman, D., & Graziano, K. M. (2012). Individual differences in the strength of taxonomic versus thematic relations. *Journal of Experimental Psychology: General, 141*(4), 601–609. https://doi.org/10.1037/a0026451

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition, 37*(7), 1026–1039. https://doi.org/10.3758/MC.37.7.1026

Mirman, D., Landrigan, J.-F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. https://doi.org/10.1037/bul0000092

Mirman, D., Walker, G. M, & Graziano, K. M. (2011). A tale of two semantic systems: Taxonomic and thematic knowledge. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33. https://escholarship.org/uc/item/4wq375c2

Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory*, *5*(2), 175–200. https://doi.org/10.1515/CLLT.2009.008

Montero-Melis, G., Van Paridon, J., Ostarek, M., & Bylund, E. (2022). No evidence for embodiment: The motor system is not needed to keep action verbs in working memory. *Cortex, 150,* 108–125. https://doi.org/10.1016/j.cortex.2022.02.006

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J. L., Madden, J., Alifano, F., Bergen, B., Bloxsom, N. G., Bub, D. N., Cai, Z. G., Chartier, C. R., Chatterjee, A., Conwell, E., Cook, S. W., … Ziv-Crispel, N. (2022). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic Bulletin & Review*, *29*(2), 613–626. https://doi.org/10.3758/s13423-021-01927-8

Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, *157*, 384–402. https://doi.org/10.1016/j.cognition.2016.09.011

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 863–883. https://doi.org/10.1037/0278-7393.21.4.863

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Murphy G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23*(4), 1035–1042. https://doi.org/10.3758/s13423-015-0834-3

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13. https://jalt-publications.org/tlt/issues/2007-07_31.7

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition*, *41*(6), 797–819. https://doi.org/10.3758/s13421-013-0312-y

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition*, *28*(6), 887–899. https://doi.org/10.3758/bf03209337

Nelson, D. L., McEvoy, C. L., & Pointer, L. (2003). Spreading activation or spooky action at a distance? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 42–52.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation/

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers 36*, 402–407. https://doi.org/10.3758/BF03195588

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin, 84*(1), 93–116.

Nelson, K. (1988). Where do taxonomic categories come from? *Human Development*, *31*(1), 3–10. https://doi.org/10.1159/000273198

Nerlich, B., & Clarke, D. D. (2000). Semantic fields and frames: Historical explorations of the interface between language, action, and cognition. *Journal of Pragmatics, 32*(2), 125–150. https://doi.org/10.1016/S0378-2166(99)00042-9

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics, 16*(3), 389–407. https://doi.org/10.1111/j.1473-4192.2006.00124.x

Norrby, C. E., & Håkansson, G. (2007). Girl–lass or curl? World associations in second language learners. *Australian Review of Applied Linguistics*, *30*(2), 22–1. https://doi.org/10.2104/aral0722

Osgood, C. E., Sebeok, T. A. (Eds.), & Gardner, J. W., Carroll, J. B., Newmark, L. D., Ervin, S. M., Saporta, S., Greenberg, J. H., Walker, D. E., Jenkins, J. J., Wilson, K., & Lounsbury, F. G. (1954). Psycholinguistics: A survey of theory and research problems. *The Journal of Abnormal and Social Psychology, 49*(4, Pt.2), i–203. https://doi.org/10.1037/h0063655

Ostarek, M., & Huettig, F. (2017). A task-dependent causal role for low-level visual processes in spoken word comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(8), 1215–1224. https://doi.org/10.1037/xlm0000375

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(4), 579–590. https://doi.org/10.1037/xlm0000318

Ovando-Tellez, M., Rohaut, B., George, N., Bieth, T., Hugueville, L., Ibrahim, Y., Courbet, O., Naccache, L., Levy, R., Garcin, B., & Volle, E. (2022). Does adding *beer* to *coffee* enhance the activation of *drinks*? An ERP study of semantic category priming. *Cognitive Neuroscience*, *13*(2), 61–76. https://doi.org/10.1080/17588928.2021.1940117

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Palmiero, M., Piccardi, L., Giancola, M., Nori, R., D'Amico, S., & Olivetti Belardinelli, M. (2019). The format of mental imagery: From a critical review to an integrated embodied representation approach. *Cognitive Processing*, *20*(3), 277–289. https://doi.org/10.1007/s10339-019-00908-z

Papeo, L., & Hochmann, J. R. (2012). A cross-talk between brain-damage patients and infants on action and language. *Neuropsychologia*, *50*(7), 1222–1234. https://doi.org/10.1016/j.neuropsychologia.2012.03.025

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Pavlenko, A. (2014). *The bilingual mind and what it tells us about language and thought*. Cambridge University Press. https://doi.org/10.1017/CBO9781139021456

Pearl, J. & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Pessoa, L. (2022). *The entangled brain: How perception, cognition, and emotion are woven together*. MIT Press. https://mitpress.mit.edu/9780262544603/the-entangled-brain/

Petrey, S. (1977). Word associations and the development of lexical memory. *Cognition, 5*(1), 57–71. https://doi.org/10.1016/0010-0277(77)90017-8

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pierce, J. R. (1961/1980). *An introduction to information theory: Symbols, signals and noise*. Dover Publications.

Piranian, D. (1983). Development of word associations in a second language. *Proceedings of the Ninth Annual Meeting of the Berkeley Linguistics Society*, 202–210. https://doi.org/10.3765/bls.v9i0.1991

Pitt, B., Gibson, E., & Piantadosi, S. T. (2022). Exact number concepts are limited to the verbal count range. *Psychological Science, 33*(3), 371–381. https://doi.org/10.1177/09567976211034502

Playfoot, D., Balint, T., Pandva, V., Parkes, A., Peters, M., & Richards, S. (2016). Are word association responses really the first words that come to mind? *Applied Linguistics, 39*(5), 607–624. https://doi.org/10.1093/applin/amw015

Postman, L., & Keppel, G. (Eds.). (1970). *Norms of word association*. Academic Press. https://doi.org/10.1016/C2013-0-11324-4

Pulvermüller, F., Assadollahi, R., & Elbert, T. (2001). Neuromagnetic evidence for early semantic access in word recognition. *The European Journal of Neuroscience*, *13*(1), 201–205. https://doi.org/10.1046/j.0953-816x.2000.01380.x

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–892. https://doi.org/10.1162/0898929054021111

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences, 3*(1), 111–169. https://doi.org/10.1017/S0140525X00002053

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, *7*(3), 113–118. https://doi.org/10.1016/s1364-6613(03)00003-2

Quillian, M. R. (1967). Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430. https://doi.org/10.1002/bs.3830120511 Reprinted in R. J. Brachman, & H. J. Levesque (Eds.), (1985). Readings in knowledge representation (pp. 97–118). M. Kaufmann Publishers. https://openlibrary.org/

Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, *12*(8), 459–476. https://dl.acm.org/doi/pdf/10.1145/363196.363214

R Core Team (2021). R: A language and environment for statistical ## computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Rabini, G., Ubaldi, S., & Fairhall, S. L. (2021). Combining concepts across categorical domains: A linking role of the precuneus. *Neurobiology of Language*, *2*(3), 354–371. https://doi.org/10.1162/nol_a_00039

Ram, K., & Wickham. H. (2018). wesanderson: A Wes Anderson Palette Generator. R package version 0.3.6. https://CRAN.R-project.org/package=wesanderson

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science, 24*(6), 1017–1023. https://doi.org/10.1177/0956797612460691

Rapp, R. 2002. The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *COLING 2002: The 19th International Conference on Computational Linguistics*. https://aclanthology.org/C02-1007.pdf

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences, 16*(11), 550–558. https://doi.org/10.1016/j.tics.2012.09.003

Rice, G. E., Lambon Ralph, M. A., & Hoffman, P. (2015). The roles of left versus right anterior temporal lobes in conceptual knowledge: An ALE meta-analysis of 97 functional neuroimaging studies. *Cerebral cortex*, *25*(11), 4374–4391. https://doi.org/10.1093/cercor/bhv024

Riegel, K. F., Ramsey, R. M., & Riegel, R. M. (1967). A comparison of the first and second languages of American and Spanish students. *Journal of Verbal Learning and Verbal Behavior, 6*(4), 536–544.

Riegel, K. F., & Riegel, R. M. (1964). Changes in associative behavior during later years of life: A cross-sectional analysis. *Vita Humana, 7*(1), 1–32. https://doi.org/10.1159/000270049

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior, 12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73)80056-8

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188–194. https://doi.org/10.1016/s0166-2236(98)01260-0

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language, 87*, 16–37. https://doi.org/10.1016/j.jml.2015.10.006

Roget, P. M. (1911). *Roget's Thesaurus of English words and phrases*. http://www.gutenberg.org/etext/10681

Rohde, H., & Rubio-Fernandez, P. (2022). Color interpretation is guided by informativity expectations, not by world knowledge about colors. *Journal of Memory and Language, 127*, Article 104371. https://doi.org/10.1016/j.jml.2022.104371

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.) Cognition and categorization (pp. 27–48). Lawrence Erlbaum.

Rosch, E. (1999). Reclaiming concepts. *Journal of Consciousness Studies, 6*(11-12), 61–77.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosenzweig, M. R. (1970). International Kent Rosanoff word association norms, emphasizing those of French male and female students and French workmen. In L. Postmen & G. Keppel (Eds.) Norms of word association (pp. 95-176). Academic Press.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521 562. https://doi.org/10.1016/S0364-0213(02)00078-2

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rubin, D. C. (2022). A conceptual space for episodic and semantic memory. *Memory & Cognition*, *50*(3), 464–477. https://doi.org/10.3758/s13421-021-01148-3

Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test*. University of Minnesota.

Sadoski, M., & Paivio, A. (2013). *Imagery and text: A dual coding theory of reading and writing* (2nd ed.). Routledge. https://doi.org/10.4324/9780203801932

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Sandgren, O., Salameh, E. K., Nettelbladt, U., Dahlgren-Sandberg, A., & Andersson, K. (2021). Using a word association task to investigate semantic depth in Swedish-speaking children with developmental language disorder. *Logopedics, Phoniatrics, Vocology*, *46*(3), 134–140. https://doi.org/10.1080/14015439.2020.1785001

Sandler, W. (2017). The challenge of sign language phonology. *Annual Review of Linguistics*, *3*(1), 43–63. https://doi.org/10.1146/annurev-linguistics-011516-034122

Santos, A., Chaigneau, S. E., Simmons, W. K. & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition, 3*(1), 83–119. https://doi.org/10.1515/langcog.2011.004

Sass, K., Heim, S., Sachs, O., Theede, K., Muehlhaus, J., Krach, S., & Kircher, T. (2010). Why the leash constrains the dog: The impact of semantic associations on sentence production. *Acta Neurobiologiae Experimentalis*, *70*(4), 435–453.

Saussure, F. (1916/1983). *Course in general linguistics. (trans. Roy Harris)*. Duckworth.

Savic, O., Savic, A. M., & Kovic, V. (2017). Comparing the temporal dynamics of thematic and taxonomic processing using event-related potentials. *PLoS ONE 12*(12), Article e0189362. https:// doi.org/10.1371/journal.pone.0189362

Savic, O., Thierry, G., & Kovic, V. (2020). Conceptual relation preference: A matter of strategy or one of salience? *Acta psychologica*, *204*, Article 103018. https://doi.org/10.1016/j.actpsy.2020.103018

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Erlbaum.

Schmitt, N., (1998a). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning 48*(2), 281–317. https://doi.org/10.1111/1467-9922.00042

Schmitt, N., (1998b). Quantifying word association responses: What is native-like? *System, 26*(3), 389–401. https://doi.org/10.1016/S0346-251X(98)00019-0

Schulte im Walde, S., Melinger, A., Roth, M., & Weber, A. (2008). An empirical characterisation of response types in German association norms. *Research on Language and Computation*, *6*(2), 205–238. https://doi.org/10.1007/s11168-008-9048-4

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., Mirman, D., & Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(20), Article 8520–8524. https://doi.org/10.1073/pnas.1014935108

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Seth, A. (2021). *Being you: A new science of consciousness*. Penguin Random House.

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*, Article 107307. https://doi.org/10.1016/j.neuropsychologia.2019.107307

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27* (3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharwood Smith, M. (2009). A MOGUL perspective on consciousness and input processing. [*Write up of a plenary presentation given at the 2009 conference on Second Language Processing and Parsing: State of the Science.*] https://www.cognitionframework.com/publications/

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, *49*(1), 222–231. https://doi.org/10.1016/j.cortex.2011.10.005

Sheng, L., & Lam, B. P. (2015). Slot-filler and taxonomic organization: The role of contextual experience and maternal education. *Journal of Educational and Developmental Psychology, 5*(1), 128–138. https://doi.org/10.5539/jedp.v5n1p128

Sheya, A., & Smith, L. (2019). Development weaves brains, bodies and environments into cognition. *Language, Cognition and Neuroscience, 34*(10),1266-1273. https://doi.org/10.1080/23273798.2018.1489065

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English, *ELT Journal, 62*(4), 339–348. https://doi.org/10.1093/elt/ccm091

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology, 102*(1-3), 106–119. https://doi.org/10.1016/j.jphysparis.2008.03.014

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex, 15*(10), 1602–1608. https://doi.org/10.1093/cercor/bhi038

Simpson, J. R., Jr, Snyder, A. Z., Gusnard, D. A., & Raichle, M. E. (2001). Emotion-induced changes in human medial prefrontal cortex: I. During cognitive task performance. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 683–687. https://doi.org/10.1073/pnas.98.2.683

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195183115.001.0001

Sloman, S., & Fernbach, P. (2017). *The knowledge illusion: The myth of individual thought and the power of collective wisdom*. Pan Macmillan.

Sloutsky, V. M., & Deng, W. (2019). Categories, Concepts, and Conceptual Development. *Language, Cognition and Neuroscience, 34*(10), 1284–1297. https://doi.org/10.1080/23273798.2017.1391398

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214–241. https://doi.org/10.1037/h0036351

Sökmen, A. J. (1993). Word association results: A window to the lexicons of ESL students. *JALT Journal*, *15*(2), 135–150. https://jalt-publications.org/jj/articles/2848-word-association-results-window-lexicons-esl-students

Sonbul, S., El-Dakhs, D., Conklin, K., & Carrol, G. (2022). “Bread and butter” or “butter and bread”? Nonnatives’ processing of novel lexical patterns in context. *Studies in Second Language Acquisition,* 1–23. https://doi.org/10.1017/S0272263122000237

Sonbul, S., El-Dakhs, D., & Masrai, A. (2022). Second language productive knowledge of collocations: Does knowledge of individual words matter? *Studies in Second Language Acquisition,* 1–23. https://doi.org/10.1017/S0272263122000341

Spätgens, T., & Schoonen, R. (2020). The structure of developing semantic networks: Evidence from single and multiple nominal word associations in young monolingual and bilingual readers. *Applied Psycholinguistics*, *41*(5), 1141-1169. https://doi.org/10.1017/S0142716420000430

Spivey M. J. (2023). Cognitive science progresses toward interactive frameworks. *Topics in Cognitive Science*. Advance online publication. https://doi.org/10.1111/tops.12645

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Suzuki-Parker, J., & Higginbotham, G. (2019). Does method of administration influence word association test responses? *Vocabulary Education Research Bulletin*, *8*(1), 11-16. https://jaltvocab.weebly.com/publications.html

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281. https://doi.org/10.1162/0898929053124965

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Thwaites, P. (2018). *Lexical and distributional influences on word association response generation.* [Unpublished PhD thesis]. Cardiff University. https://orca.cardiff.ac.uk/id/eprint/119182

Tomasino, B., Fink, G. R., Sparing, R., Dafotakis, M., & Weiss, P. H. (2008). Action verbs and the primary motor cortex: A comparative TMS study of silent reading, frequency judgments, and motor imagery. *Neuropsychologia*, *46*(7), 1915–1926. https://doi.org/10.1016/j.neuropsychologia.2008.01.015

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia 35*(10), 1319–1327. https://doi.org/10.1016/S0028-3932(97)00085-7

Trapp, S., Shenhav, A., Bitzer, S., & Bar, M. (2015). Human preferences are biased towards associative information. *Cognition & Emotion*, *29*(6), 1054–1068. https://doi.org/10.1080/02699931.2014.966064

Tresselt, M. E., & Mayzner, M. S. (1964). The Kent-Rosanoff word association: Word association norms as a function of age. *Psychonomic Science*, *1*(1), 65–66.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Troyer, M., & McRae, K. (2022). Thematic and other semantic relations central to abstract (and concrete) concepts. *Psychological research*, *86*(8), 2399–2416. https://doi.org/10.1007/s00426-021-01484-8

Truscott, J., & Smith, M. S. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, *7*(1), 1–20. https://doi.org/10.1017/S1366728904001178

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory (pp. 381–402). Academic Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84(*4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

University of Oxford. (1988). *EAT: the Edinburgh associative corpus*. M. Wilson, G. Kiss, & C. Armstrong. Oxford Text Archive. http://hdl.handle.net/20.500.12024/1251

van Dam, W. O., Brazil, I. A., Bekkering, H., & Rueschemeyer, S. A. (2014). Flexibility in embodied language processing: context effects in lexical access. *Topics in cognitive science*, *6*(3), 407–424. https://doi.org/10.1111/tops.12100

van Dijk, L. (2016). Laying down a path in talking. *Philosophical Psychology, 29*(7), 993–1003. https://doi.org/10.1080/09515089.2016.1213379

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. https://doi.org/10.1017/S1366728998000352

van Hell, J.G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789. https://doi.org/10.3758/BF03196335

van Heuven, W. J., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex*, *18*(11), 2706–2716. https://doi.org/10.1093/cercor/bhn030

Van Rensbergen, B., De Deyne, S., & Storms, G. (2016). Estimating affective word covariates using word association data. *Behavior research methods*, *48*(4), 1644–1652. https://doi.org/10.3758/s13428-015-0680-2

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* MIT Press.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex, 24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*(4), 422–488. https://doi.org/10.1016/j.cogpsych.2003.09.001

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition: An Interdisciplinary*

Journal of Language and Cognitive Science, 11(3), 403–430. https://doi.org/10.1017/langcog.2019.23

Vincent-Lamarre, P., Massé, A. B., Lopes, M., Lord, M., Marcotte, O., & Harnad, S. (2016). The latent structure of dictionaries. *Topics in Cognitive Science, 8*(3) 625–659. https://doi.org/10.1111/tops.12211

Vivas, L., Manoiloff, L., García, A. M., Lizarralde, F., & Vivas, J. (2019). Core semantic links or lexical associations: Assessing the nature of responses in word association tasks. *Journal of Psycholinguistic Research*, *48*(1), 243–256. https://doi.org/10.1007/s10936-018-9601-8

Von Neumann, J. (1958). *The computer and the brain*. Yale University Press.

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

Warren, T., & Dickey, M. W. (2021). The use of linguistic and world knowledge in language processing. *Language and Linguistics Compass, 15*(4), Article e12411. https://doi.org/10.1111/lnc3.12411

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge. Further fractionations and an attempted integration. *Brain*, *110*(5), 1273–1296. https://doi.org/10.1093/brain/110.5.1273

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–854. https://doi.org/10.1093/brain/107.3.829

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

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1), 79–95. https://doi.org/10.1017/S0272263108080042

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024–1037. https://doi.org/10.1162/08989290051137558

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language, 99,* 122–150. https://doi.org/10.1016/j.jml.2017.09.006

Wettler, M., & Rapp, R. (1993). Computation of word associations based on co-occurrences of words in large corpora. *Very large corpora: Academic and industrial perspectives*. https://aclanthology.org/W93-0310.pdf

Wettler, M., Rapp, R., & Sedlmeier, P. (2005). Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, *12*(2-3), 111–122. https://doi.org/10.1080/09296170500172403

Whorf, B. L. (2011). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. J. B. Carroll (Ed.) MIT Press.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wickham, H. et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18*(4), 303–324. https://doi.org/10.1191/0267658302sr203oa

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/bf03196322

Wingfield, C., & Connell, L. (2022). Understanding the role of linguistic distributional knowledge in cognition. *Language, Cognition and Neuroscience*, *37*(10), 1220–1270. https://doi.org/10.1080/23273798.2022.2069278

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge. https://doi.org/10.4324/9781315165547

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, *6*(1). Article 79. https://doi.org/10.5334/gjgl.1646

Wittgenstein, L. (1953/2009). *Philosophical investigations* (4th ed.; G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans). Wiley-Blackwell.

Wojcik, E. H., & Kandhadai, P. (2020). Paradigmatic associations and individual variability in early lexical-semantic networks: Evidence from a free association task. *Developmental Psychology, 56*(1), 53–69. https://doi.org/10.1037/dev0000844

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*(1), 41–69. https://doi.org/10.1017/S0272263101001024

Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, *30*(3), 315–329. https://doi.org/10.1016/S0346-251X(02)00017-9

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition, 40*(2), 395–416. https://doi.org/10.1017/S0272263117000237

WordNet. (2010). Princeton University "About WordNet." *Princeton University*. https://wordnet.princeton.edu/

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

Wurm, M.F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications*, *10*, 289. https://doi.org/10.1038/s41467-018-08084-y

Zareva, A. (2007). Structure of the second language mental lexicon: How does it compare to native speakers' lexical organization? *Second Language Research*, *23*(2), 123–153. https://doi.org/10.1177/0267658307076543

Zareva, A. (2010). Multicompetence and L2 users' associative links: Being unlike nativelike. *International Journal of Applied Linguistics*, *20*, 2–22. https://doi.org/10.1111/j.1473-4192.2009.00226.x

Zareva, A., & Wolter, B. (2012). The "promise" of three methods of word association analysis to L2 lexical research. *Second Language Research*, *28*(1), 41–67. https://doi.org/10.1177/0267658311423452

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition, 196,* Article 104135. https://doi.org/10.1016/j.cognition.2019.104135

Zhang, Y., Han, K., Worth, R., & Liu, Z. (2020). Connecting concepts in the brain by mapping cortical representations of semantic relations. *Nature Communications*, *11*(1), Article 1877. https://doi.org/10.1038/s41467-020-15804-w

Zwaan, R. A. (2004). The Immersed Experiencer: Toward an Embodied Theory of Language Comprehension. In B. H. Ross (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol. 44, (pp. 35–62). Elsevier Science.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE, 7*(12), Article e51382. https://doi.org/10.1371/journal.pone.0051382

Zwaan, R., Stanfield, R., & Yaxley, R. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430