

Lexical and distributional influences on word
association response generation

Peter Thwaites

Centre for Language and Communication Research
School of English, Communication and Philosophy, Cardiff
University

Submitted to Cardiff University in partial fulfilment of the
degree of Doctor of Philosophy

2018

Acknowledgements

This thesis would not have been possible without the help and support of a great many people. Chief among these are my three supervisors. The first of these was Tess Fitzpatrick, who (displaying remarkable patience in the face of a barrage of rambling emails) welcomed me into the Lexical Studies program at Cardiff, and supervised the first half of this thesis with care, intelligence, and enthusiasm.

After Tess's departure from Cardiff, Alison Wray and Andreas Buerki stepped up to supervise the second half of the thesis. Both showed a remarkable knack for pinpointing the flaws in my theories, the weak links in my writing, and the deficiencies in my statistical analyses, and offered patience and enthusiasm in helping me to address these (the remaining flaws in this thesis are, of course, my own). I have learnt so much from all three of my supervisors, and for this I offer deep gratitude.

Beyond my supervision team, I am grateful for the support of the other members of the Lexical Studies programs at Cardiff and Swansea universities – especially Caroline Handley and Dale Brown – for the helpful and stimulating conversations and email exchanges. I would also like to thank the faculty at the Foreign Language Institute at Yeungnam University, and in the Department of English Education at Keimyung University, for their support. In particular, thanks are owed to Karl Agnew for taking time out from early parenthood to sit around drinking coffee and talking about semantics with me; and to Whit Altizer for dragging me off up mountains when PhD stress seemed to be getting the better of me.

Thanks are also due to a number of people who helped with various other aspects of the work presented below: Mike Scott, who adapted his Wordsmith software in order to make the relative entropy analysis in Chapter 8 possible; Daniel Ryan for helping me to make several types of connections; to Violetta Seretan and Eric Wehrli for their help with corpus research; and to Professor Sir Paul Curran for his lifelong support and guidance.

Finally, I could not have completed this project without the unending patience, love, and support of my family. To the hardworking staff at the Bank of Mum and Dad – thank you for everything. To Jade and Abby: I look forward to being able to be the husband and daddy that you deserve again, rather than the grumpy, haggard imposter you've lived with over the past couple of years.

Thesis summary

This thesis is the result of an attempt to investigate the determinants of word association responses. The aim of this work was to identify those properties of words – their frequency, grammatical class, and textual distribution, for example – which influence the generation of word association responses, and to align these effects with wider psycholinguistic views of the mental lexicon.

The experimental work in the early chapters focuses on grammatical influences on word association. In particular, it is demonstrated that both grammatical class and verb transitivity influence the type of response most likely to be selected by participants. The immediately following chapters ask why this would be so. The analysis of several models of word association suggests that the development of a clearer understanding of the way in which a word's textual distribution impacts upon associative response patterns may be an important stepping stone towards a coherent model of associative response generation.

In the later part of the thesis, a series of novel experiments is conducted comparing word association response patterns with corpus-derived data. This work in turn lays the foundation for the development of a new usage-based model of word association, which is shown, in the penultimate chapter, to be capable of explaining a wide range of research findings, including not only the grammatical class and transitivity-related findings described above, but also earlier findings relating to the influence of lexical variables on the structure of the associative network, and to the discovery of individual and age-related response patterns in word association.

Table of Contents

Extended Abstract.....	12
Chapter 1: Introduction.....	15
1.1 The promise of the word association task.....	15
1.2 Why is word association research struggling to deliver results?.....	17
1.3 Setting word association on a firmer foundation.....	23
Chapter 2: Revisiting Individual Word Association Profiles: A Replication of Fitzpatrick 2007	27
2.1. Introduction	27
2.1.1 Fitzpatrick, 2006.....	27
2.1.2 Fitzpatrick, 2007.....	29
2.1.3 Why select this study for replication?	31
2.2. Replication design.....	32
2.2.1 Participants	33
2.2.2 The task.....	33
2.2.3 Test administration	33
2.2.4 Categorization of responses	33
2.2.5 Preparing the data for analysis	33
2.3. Results.....	34
2.3.1 Do adult native speakers respond to cue words in a predictable, homogeneous way?	34
2.3.2 Do individual native speakers respond to cue words in a consistent way?	35
2.3.3 Did the replication participants respond to the cue words in similar ways to those of the original study?.....	37
2.3.4 Cue level variation in responses.	41
2.4. Discussion.....	44
2.4.1 Differences between groups.....	45
2.4.2 Subjectivity of coding.....	45
2.4.3 Cue-related response variation.....	46

2.5. Conclusions	47
Chapter 3: Cue word influence on word association behaviour.....	51
3.1. Introduction	51
3.2. How is word association measured?	52
3.3 The lexicon as a network	53
3.3.1 Distributional and semantic access in word association	59
3.4 Lexical variables	61
3.4.1 Grammatical class	61
3.4.2 Frequency.....	65
3.4.3 Age of acquisition.....	73
3.4.4 Concreteness and imageability.....	75
3.4.5 Affective variables.....	80
3.5 Discussion.....	82
Chapter 4: Do nouns and verbs elicit different types of word association response?	87
4.1 Introduction	87
4.1.1 Grammatical class selection.....	87
4.2 The current study.....	91
4.2.1 Participants	92
4.2.2 Cue selection.....	92
4.2.3 Materials	95
4.2.4 Dependent measures.....	95
4.2.5 Research questions	97
4.2.6 Removal of unsuitable cues	98
4.3 Results.....	98
4.3.1 Response stereotypy.....	98
4.3.2 Basic-level response type.....	99
4.3.3 Detailed response types.....	100

4.4 Discussion.....	103
4.4.1 Do noun and verb cues differ in terms of the stereotypy of the responses which they elicit, when key variables are controlled?	103
4.4.2 Do noun and verb cues differ in terms of the category of the responses that they elicit, when the same variables are controlled?.....	103
4.4.3 What can be learned from the analysis of responses to nouns and verbs using detailed levels of analysis?	105
4.5 Conclusion.....	108
Chapter 5: Does verb transitivity influence word association responses?.....	110
5.1 Introduction	110
5.2 Defining transitivity.....	110
5.3 Previous research on transitivity and lexical processing	112
5.4 The current study.....	117
5.4.1 Predictions	117
5.4.2 Method	118
5.4.3 Categorization of data.....	120
5.5 Results.....	122
5.5.1 Grammatical class of responses.....	122
5.5.2 Response type.....	123
5.5.3 Response distribution	126
5.6 Discussion and conclusion	126
Chapter 6: Theoretical explanations for transitivity effects.....	130
6.1 Introduction	130
6.2 Clark’s Generative-Transformational WA model.....	130
6.3 Usage-based theories	135
6.3.1 Contiguity-based WA models.....	136
6.3.2 Semantic models of word association	142
6.4 Discussion.....	152

Chapter 7: Combining co-occurrence data with word association response categorisations	154
7.1 Introduction	154
7.2 A composite usage-based WA model	155
7.3 Experiment 1a	158
7.3.1 Aims and predictions	158
7.3.2 Selection of data sources	160
7.3.3 Corpus search parameters	160
7.3.4 Selection of co-occurrence strength measure	161
7.3.5 Procedure	163
7.3.6 Results	163
7.4 Experiment 1b	164
7.4.1 Results	164
7.4.2 Discussion	165
7.5 Experiment 2	167
7.5.1 Aims and predictions	167
7.5.2 Method and results	167
7.5.3 Discussion	169
7.6 Conclusion	175
Chapter 8: What can usage-based statistics reveal about word association?	178
8.1 Introduction	178
8.1.1 Entrenchment, relative entropy, and word association	178
8.1.2 Statistical approaches to entrenchment and abstraction	181
8.1.3 Type/token distributions and word association	182
8.1.4 Testing the composite model using type/token distributions	185
8.2 Experiment	185
8.2.1 Hypotheses	185
8.2.2 Procedure	186

8.2.3 Results	187
8.3 Discussion.....	188
8.3.1 Comparing corpora and WA responses	188
8.4 Conclusion.....	196
Chapter 9: General Discussion	198
9.1 Revisiting some initial questions.....	198
9.2 Deeper investigations into word variable influence	199
9.3 Re-evaluating the composite model	202
9.4 A revised composite model.....	208
9.5 Advantages of the composite model	210
9.6 Some unanswered questions.....	211
9.6.1 Which semantic knowledge?	212
9.6.2 Two “litmus tests” for word association models	222
9.7 Implications for further research	225
9.7.1 Methodology and the rehabilitation of the WA method for psycholinguistic research....	226
Chapter 10: Conclusion	228
Bibliography	234
Appendix 1 Test materials for replication of Fitzpatrick 2007, Chapter 2.	255
Appendix 2: Cue word list for Nouns and Verbs study, Chapter 4.....	261
Appendix 3 Cue word list for Transitivity study, Chapter 5.....	266
Appendix 4: Categorization of responses in Chapter 5.....	271

List of figures and table

Table 2.1: Comparison of Fitzpatrick (2007) with the present replication	32
Table 2.2: Scheme for classification of responses (from Fitzpatrick, 2007: 325)	34
Table 2.3: Total number of responses of each type per participant in tests 1 and 2	35
Table 2.4: Correlations between mean sub-category scores in Fitzpatrick 2007 and the current replication	37
Table 2.5: Comparison of minimum, maximum, and mean number of responses given per participant in the current study and Fitzpatrick (2007)	39
Table 2.6: Comparison of minimum, maximum, and mean number of responses given per cue in the current study	40
Table 2.7: Response type distributions for a sample of cues, paired according to like frequency and grammatical class	42
Table 2.8: Comparison of minimum, maximum, and mean number of responses given per cue at the highest and lowest frequency bands.	42
Table 2.9: Comparison of minimum, maximum, and mean number of responses given per cue according to grammatical class	43
Table 3.1: Methods for measuring word association responses	55
Table 3.2: Percentage of syntagmatic responses for various grammatical classes in two large-scale studies	62
Table 4.1: Comparison and examples of Zipf values (from Van Heuven, Mandera, Keuleers, and Brysbaert (2014)	94
Table 4.2: Comparison of variable values for noun and verb cues	94
Table 4.3: Scheme for categorization of responses (adapted from Fitzpatrick, Playfoot, Wray, and Wright, 2015)	102
Table 4.4: All meaning-based primary responses to verbs, categorized by semantic relationship	106
Table 4.5: All meaning-based primary responses to verbs, in their original categorizations	106
Table 5.1: t-tests and Min/Max statistics for transitive and intransitive cues	119
Table 5.2: Response stereotypy means and t-tests	125
Table 5.3: Response stereotypy means and t-tests after removal of semantically narrow cues.	126
Table 7.1: Post-hoc Tukey HSD tests of mean difference in MI values for four types of associative relation	164
Table 7.2: Mean and standard deviation values for MI scores across 4 basic categories	166
Table 7.3: Correct classification (position-based vs non-position based) of cue-response pairs (%)	168

Table 7.4: Correct classification (position-based vs non-position based) of cue-response pairs by corpus condition (%)	170
Table 8.1: Associative and distributional properties of 5 cues with the largest difference between associative and distributional entropy	189
Table 8.2: Association responses and most frequent collocations for these 5 cues	190
Figure 2.1: Mean number of responses of each sub-type given by participants in the current replication and Fitzpatrick (2007)	38
Figure 2.2: Mean number of responses of each sub-type per cue in the current study	41
Figure 4.1: Total response counts of each basic type for noun and verb cues	99
Figure 4.2: Total response counts of each detailed meaning-based category to noun and verb cues	100
Figure 4.3: Total response counts of each detailed position-based category to noun and verb cues	101
Figure 5.1: Mean responses of four grammatical classes to transitive and intransitive cues. Note that all proper nouns were included in the “noun” column	122
Figure 5.2: Mean responses to transitive and intransitive cues categorised as cue-response and response-cue	123
Figure 7.1: The first iteration of the composite model of word association response generation	156
Figure 9.1: The second iteration of the composite model of word association response generation	206

Extended Abstract

This thesis asks why research using the word association (WA) method is generally failing to generate informative findings on the structure of the mental lexicon, in spite of widely-held belief in its potential. It seeks to show that this potential might best be realised through the use of an iterative approach to WA research design which focuses on critically examining the numerous untested assumptions and underexplored sources of variation in WA response patterns.

Initial support for this approach is provided by a replication of a study by Fitzpatrick (2007) which shows that prior assumptions regarding homogeneity of responses were incorrect: individual respondents have their own WA response profiles, distinct from any homogeneous group-level pattern (responses were coded on two levels: firstly at a basic meaning-, position-, and form-based level, and subsequently into more detailed sub-categories).

The replication study also suggested that a range of cue-level variables, including frequency, concreteness, age-of-acquisition, and, most notably, grammatical class, also appear to affect response patterns. A series of progressively-structured experiments explores this variation in more detail. This reveals a consistent and independent influence of cue grammatical class and (when the cue is a verb) transitivity on response patterns.

The nature of this latter finding highlights a tension in existing theories of word association between the influence of semantic knowledge, on one hand, and co-occurrence-based information, on the other, on response generation. A second set of experiments explores this tension by comparing WA data with corpus-derived lexical co-occurrence patterns. In order to facilitate these investigations, a new model of WA response generation, termed the *composite model*, is designed. The model allows hypotheses regarding the role of semantic, predictive, and form-based knowledge on WA response generation to be generated and tested.

These experiments reveal a complex influence of distributional information on WA response patterns. A key finding is that corpus-derived type/token distributions of verb cues and their noun complements significantly influence WA response patterns (that is, high type variation was positively correlated with the number of verb responses and response types given to a cue); but corpus data is generally unsuccessful in predicting primary WA responses (that is, it is not the most frequent collocates of a cue that tend to be generated as WA responses). This suggests that distributional influences on WA may be indirect: word co-occurrences influence the semantic network underlying the generation of WA responses, but do not influence responses directly. This in turn implies a distinction between the formation of associations, which is influenced by a diverse range of processes including the perception

of statistical co-occurrences and lexical similarities; and their generation in word association tasks, which involves selective access to semantic representations of cue words. These findings themselves provide evidence of the potential for carefully-designed word association research, supported by a theoretically justified model of lexical knowledge, to provide insights into the structure of the mental lexicon.

Chapter 1: Introduction

1.1 The promise of the word association task

There is a perception within the applied linguistic community that the word association (WA) method is failing to deliver on its research potential. Schmitt (2010), for example, has suggested that second language researchers are “still waiting for a breakthrough in methodology which can unlock [the] undoubted potential [of the word association task]”, and Wolter (2002), regarding WA-based L2 language proficiency testing research, asks whether there is “still hope” for the method. Away from the L2 studies, some researchers have argued that the psycholinguistic basis of word association is not well understood (D. L. Nelson, McEvoy, & Dennis, 2000), and, for this reason, others have cautioned against the use of WA norms as explanatory variables in psycholinguistic experiments (Hutchison, Balota, Cortese, & Watson, 2008; McRae, Khalkhali, & Hare, 2012). There is thus some anxiety regarding the status and validity of WA research in both first and additional language research. This thesis engages in depth with the potential for WA to live up to its promise, and as a first step this chapter will explore the anxiety around the task. Its aim is to set out some of the concerns currently surrounding word association experiments, and reflect on the most useful balance between using empirical studies to address methodological issues, on the one hand, and answer research questions on the other.

Firstly, however, it is worthwhile to ask why it is that the word association task continues to be perceived as being so rich in potential, given the anxiety discussed above. There are three ways to answer this question. They can be broadly termed intuitive, methodological, and empirical.

The intuitive potential of the word association method refers to its face validity. For psychologists, this means that the task appears to offer access into a fundamental aspect of human thought – the tendency for one idea to call others regularly to mind. De Deyne & Storms (2015), for example, refer to associations as “a language of thought”, while Deese (1962, 1966), in a broad overview of several hundred years of work on association, suggests that much early interest was motivated by reflection on this associative quality of thought and its assumed potential for revealing truths about the nature of mental structure. Studying word association was thus viewed as a way to study the mind itself. Deese’s own work was an example of this: he showed that word associations have an inherent organisational structure which could not emerge through simple exposure to language, but must be the result of cognitive work such as contrasting and classifying. He presented his position as an attack on the Lockean concept of the *tabula rasa* – the idea that the senses alone are enough to explain human knowledge. That Locke’s own views on thought were influenced by reflection on association

was not lost on Deese; through his analysis, word association emerges as a cornerstone of Western philosophy.

Word association is also intuitively appealing to psycholinguists. One reason for this is the apparent clarity with which it reveals links between words in the mental lexicon, and thus offers insights into the way in which the lexicon is organised. De Deyne & Storms (2015: 412), for example, have suggested that “word associations provide a privileged route” into the semantic structure of the lexicon, allowing its organisational features to be studied. In L2 studies, this apparently direct route into the lexicon has led some researchers to suggest that word association may be the key to understanding the differences between the first and second language lexical organisation (Sökmen, 1993; van Hell & De Groot, 1998; Wolter, 2001, 2006), while numerous attempts have been made at devising language proficiency tests using formats based on word association (Dronjic & Helms-Park, 2014; Meara & Fitzpatrick, 2000; Read, 1993, 1995; Schmitt, Ng, & Garras, 2011).

The second aspect of the promise of the word association task is its methodological qualities. The task is generally very easy to administer, with most studies reporting that participants take less than 20 minutes to complete a task comprising up to 100 items (Fitzpatrick, 2007; Namei, 2004). This ease of administration has recently led to online forms of the WA task being developed (De Deyne, Navarro, & Storms, 2012). This format is capable of generating huge amounts of data in multiple languages with little effort required on the part of the investigator, and is already demonstrating promising findings (De Deyne, Kenett, Anaki, Faust, & Navarro, 2016; Gruenenfelder, Recchia, Rubin, & Jones, 2016; Van Rensbergen, De Deyne, & Storms, 2016; Van Rensbergen, Storms, & De Deyne, 2015). Secondly, numerous methods have been developed for analysing this data. These include response categorizations (Ervin, 1961; Fitzpatrick, Playfoot, Wray, & Wright, 2015; Meara, 2009), association strength measures (D. L. Nelson et al., 2000), response time (de Groot, 1989; Fitzpatrick & Izura, 2011; Izura & Playfoot, 2012; van Hell & De Groot, 1998), and response stereotypy (de Groot, 1989; D. L. Nelson et al., 2000; and see Chapter 3). These multiple methods allow rich data to be acquired from an easily extracted set of associations. Finally, the simplicity of the basic word association task has allowed it to be adapted to address specific research questions in different fields of study. These are most notable in the field of psychology, in which tests have been devised to study creativity (Benedek & Neubauer, 2013; Eysenck, 1994; Gough, 1976), psychiatric disorders (Merten, 2010; Mohr, Graves, Gianotti, Pizzagalli, & Brugger, 2001; Silberman, Bentin, & Miikkulainen, 2007), and cross-cultural differences (Tanaka-Matsumi & Marsella, 1976); but also in fields including education (Bahar, Johnstone, & Sutcliffe, 1999; Gulacar, Sinan, Bowman, & Yildirim, 2014) and consumer science (Ares & Deliza, 2010; Guerrero et al., 2010; Roininen, Arvola, & Lähtenmäki, 2006).

The final reason for continued interest in the word association task is empirical. Put simply, the task continues to yield valuable insights, even as its fundamental nature is disputed. Even without looking at research using word association in the fields beyond applied linguistics, the empirical developments have been considerable. For example, work on the organisation of the lexicon, which began with Deese (1962a, 1962b, 1964, 1966) and continues in the work of numerous others working in both first and second language domains (e.g. De Deyne, Leopold, Navarro, Perfors, & Storms, 2012; Van Rensbergen, Storms, & De Deyne, 2015; Wilks, 2009; Wilks & Meara, 2002), has yielded considerable new knowledge on the principles by which lexical information becomes organised within the mind, and on the properties of lexical networks. To give one example of the potential of this research, Van Rensbergen et al. (2016) have demonstrated that large-scale WA-based models are capable of outperforming corpus-derived methods in the task of approximating human semantic similarity judgements. In other areas, word association norms play a role in understanding the nature of priming effects (Heyman, Hutchison, & Storms, 2015; Hutchison, 2003; Hutchison et al., 2008; McDonald & Lowe, 1998; Murphy & Hunt, 2013) and have supported the generation of new theories of language development (Andrews, Vinson, & Vigliocco, 2008; Santos, Chaigneau, Simmons, & Barsalou, 2011; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008; Steyvers & Tenenbaum, 2005). In second language studies, word association studies continue to be useful in exploring the differences between first and second language lexical knowledge (de Groot, 1989; Meara, 2009; Schoonen & Verhallen, 2008; van Hell & De Groot, 1998; Wolter, 2001; Zareva, 2007, 2011). There are also a small number of studies which have made use of the WA method to investigate dementia (e.g. Eustache, Cox, Brandt, Lechevalier, & Pons, 1990). This wealth of new knowledge demonstrates the empirical potential of the WA task.

1.2 Why is word association research struggling to deliver results?

Given these intuitive, methodological, and empirical benefits of the word association task, why does this sense of anxiety and unfulfilled promise persist? The answer to this question is that for each of the areas of promise outlined above, there is a corresponding tension which demands that empirical results be taken with caution.

Firstly, the intuitive sense that the word association paradigm offers a direct view into the workings of the mind has tended to run up against unexpected complexities. Indeed, it can be argued that the very intuitiveness of word association has itself tended to fuel a lack of criticality towards the processes underlying the task. Researchers have, for example, used results of associational studies (introspective and otherwise) to support pre-conceived theories of mind and association. As Warren (1916, 1921) points out, much of the early history of associationism was dominated by the view that

associations are learned through temporal contiguity – the co-occurrence of phenomena in experience. This was still the case in the 1960's, when Deese (1966) complained that many then-current WA studies appeared to do little more than reassert the classical rules of contiguity. One reason for the lack of criticality towards WA at that time was that a focus on temporal contiguity suited the predictions made by Behaviourism. The classical view that associations were formed through exposure to temporal contiguities between words and concepts, with little cognitive work attributed to the participant, was adopted by behaviourists as evidence for a simple stimulus-response-based view of psychology. It took Deese's own work, with its emphasis on exploring non-contiguous (paradigmatic) responses which suggested an underlying mental tendency towards the imposition of structure upon linguistic experience, to show that this line of thought was misguided. As Deese noted, however, it should not have taken so long: that it did so was in no small part down to researchers either ignoring or explaining away paradigmatic WA responses. Somewhat ironically, however, the years which have followed Deese's work have seen something of a pendulum effect, in that word associations are now more commonly taken to be entirely semantic in nature (De Deyne & Storms, 2015; Mirman & Magnuson, 2008; Mollin, 2009). Not enough studies have attempted to seek middle ground here, although several which do will be reviewed in the chapters which follow.

In other cases, there has been a reluctance among researchers to tackle underlying complexities within the paradigm. This is essentially the problem described by McRae et al. (2012), who argue that much confusion over the nature of word association could have been avoided if researchers had paid more attention to the difference between the way in which associations are *learned* and the processes which guide their *organisation* or *production*. Part of this problem is that, in their haste to explain other phenomena or to support their own models of language processing, researchers have often used associative norms as an explanatory variable in their studies, without delving more deeply into what these norms actually represent. Hutchison (Hutchison, 2003; Hutchison et al., 2008) notes that several types of relationship are reflected in associative responses, including those based on semantic factors and temporal contiguity. Thus the fact that two words are associated does not reveal the nature of that association. Hutchison, Balota, Cortese, & Watson (2008: 1055) suggest that:

it is important to remember that association norms are themselves dependent measures. Using one dependent measure to predict another measure is not necessarily "explaining" the target phenomenon of interest. Instead, one might more appropriately ask 'what drives the association itself?' In this case, the goal for psycholinguists is to remove association norms as explanatory constructs and

replace them with explanations for why words are likely to co-occur in the first place.

In other words, the face validity of the word association paradigm obscures significant complexity as to the processes underlying the task. Too few studies have been conducted with a view to understanding exactly what it is that word association says about the relationships between words, and what the best approaches are to revealing those relationships.

Much of this unexplored complexity in word association hides behind the numerous assumptions which have been made about the methodological aspects of the paradigm. As Meara (2009a) has noted, word association studies can quickly generate a large amount of data which can be collected and analysed in a number of ways. Researchers must make difficult choices regarding the selection of cue words, study participants, and procedures of data collection and analysis. Because of this variety of methodological possibilities, Meara (2009a) has argued that it has not always been clear what constitutes “good” word association data, or how this data should be analysed. Of his early work in the field, he noted that “the data word association studies generated was enormously rich, but it was difficult to know how to exploit this richness”. There has been a general reluctance to engage with this question of WAT methodology. Instead, researchers have often fallen back upon untested assumptions about the words, people, and methods they have used.

On the first of these aspects of WAT design, Meara (2009a) has suggested that “far too little consideration has been given to what words should be used as stimuli”. Several WA researchers of the 1960’s gave some attention to this problem by analysing response patterns to words of different grammatical classes and frequency (W. P. Brown, 1971; Deese, 1962b, 1964, 1966; Entwisle, 1966a; Entwisle, Forsyth, & Muuss, 1964). There remained, however, a tendency to overgeneralise findings. This tendency extended to Deese’s own work, wherein he asserted that his results on the structure of associations between concrete nouns would apply equally well to abstract nouns and verbs, despite the absence of evidence to support this assumption (1966: 141-2). Furthermore, very few studies of word association have looked at the influence of other lexical variables on the task (though see Chapter 3). Numerous variables have been demonstrated to influence other psycholinguistic tasks, such as lexical decision and word naming. These include distributional variables such as frequency, age of acquisition, and contextual and semantic diversity; and various affective and perceptual measures, such as concreteness (Balota, Yap, & Hutchison, 2007; M. M. Louwerse & Connell, 2011; Pexman, 2012; Warriner, Kuperman, & Brysbaert, 2013; Yap, Lim, & Pexman, 2015; Yap, Tan, Pexman, & Hargreaves, 2011). In word association, few studies have attempted to systematically investigate the influence of these variables. While Van Rensbergen, Storms, & De Deyne (2015) have noted their

importance in the structure of associative responses, the degree to which these findings can shed light on the nature of the WA task itself, or help to address not only Meara's concerns about the difficulty of selecting appropriate cue words, remains low.

In place of a critical examination of the influence of various cue word factors on word association, many researchers have, as both Meara (2009) and Fitzpatrick (2006: 124) have noted, turned to existing cue lists. Such lists have the advantage of offering comparability between studies, as well as giving the appearance of validity because they follow established precedents. Numerous researchers have, however, pointed out problems with this practice. Amongst the earliest critics was again Deese (1962a, p. 79), who attacked the widely-used Kent-Rosanoff list (Kent & Rosanoff, 1910) on the grounds that its cues were too uniformly high in frequency and too biased towards nouns and adjectives. Fitzpatrick (2006) echoed this sentiment, suggesting that the list is unlikely to capture a representative sample of response patterns. She also criticises the apparent lack of rationale for their selection, even in their original setting in psychological research. In spite of these problems, however, the list was still in use in an applied linguistic setting almost one hundred years after its original publication (Namei, 2004).

One of the most serious problems with preselected lists of cue words (as well as, paradoxically, a further reason for their continued use) stems from the use of response norms derived from them. Well-known examples of such norms lists include the Edinburgh Associative Thesaurus (EAT; Kiss, Armstrong, Milroy, & Piper, 1973), the South Florida norms (D. L. Nelson, McEvoy, & Schreiber, 2004), and the Postman and Keppel norms (Postman & Keppel, 1970). These lists of responses were published with the worthy intention that future researchers would have access to large sets of WA data without having to collect it themselves. To this end, they have been successful, as evidenced by their hundreds of citations. They have, however, also had some less positive effects. Firstly, Fitzpatrick (2007: 321) suggests that such norms lists have encouraged the belief that there is a "normal", homogeneous way for L1 users to respond to word association tests (WATs). As a result, the majority of WA studies to date have worked with group-by-group analyses and paid little attention to individual differences in word association patterns. In a series of studies, Fitzpatrick (2006, 2007, 2009; Fitzpatrick, Playfoot, Wray, & Wright, 2015) has demonstrated that this assumption of L1 homogeneity in WA is misleading: respondents in fact demonstrate sustained preferences in the way they respond to WA cues, both on an individual and a generational level. Secondly, as noted above, these norms lists have frequently been used by researchers as explanatory variables in psycholinguistic studies, wherein they have been taken as proxies for semantic variables such as relatedness (D. L. Nelson et al., 2000) or neighbourhood size (Mirman & Magnuson, 2008). Such an approach again reflects

assumptions about the generalizability of data collected from what are very often relatively homogeneous groups of respondents and/or cue words, as well as about the nature of association itself.

Another problem identified by Meara is that methods of data collection, treatment, and analysis have been inconsistent, and thus presented difficulties in the interpretation and comparison of research findings. One aspect of this problem is the way in which cue-response pairs are categorized. Since Jenkins (1954), the most popular scheme of categorization is the syntagmatic-paradigmatic system which dates back to Saussure (1916). Put simply, a paradigmatic response is one of the same grammatical class as the cue; it could replace the cue in a sentence. These responses are generally held to be semantic in nature. A syntagmatic response is one of a different grammatical class to the cue. It is likely to precede or follow the cue in a sentence, and is largely thought to stem from a knowledge of contiguity.

There are numerous problems with this scheme of categorization. The simplest of these is that it is difficult to use. Meara (2009: 22) notes the awkwardness of the scheme, while several researchers (Deese, 1966; Fitzpatrick, 2006, 2007; Petrey, 1977) suggest that it may obscure the true nature of some associations. Deese (1966) gives the examples of *grass-green* and *sky-blue* as "syntagmatic responses [which] are not so much representative of the sequential transitions in ordinary language as they are of the most basic and important cognitive characteristics of grass and sky" (1966: 109). Fitzpatrick (2007) further suggests that the syntagmatic/paradigmatic distinction may not be exact enough to provide informative data about individual respondents and their responses. Fitzpatrick's own work with a more detailed scheme supports this finding, while Higginbotham (2010: 383), contrasting the syntagmatic/paradigmatic scheme with that of Fitzpatrick, suggests that the use of the former categories obscured details which only became visible through the latter.

One potential solution to this problem is to develop new schemes for categorizing WA responses. As Fitzpatrick, Playfoot, Wray, & Wright (2015) note, however, this approach creates problems of its own, in that there has been an over-proliferation of new schemes. They suggest that researchers have frequently adopted novel schemes with no real grounding either in previous research or any established theory of lexical knowledge. Some of these categories have been designed to suit specific research questions or psycholinguistic hypotheses (e.g. Santos, Chaigneau, Simmons, & Barsalou, 2011); at other times researchers have added ad-hoc categories in order to explain difficult-to-categorize examples (e.g. Sökmen, 1993). This variation in methods means that these studies are quite difficult to compare side-by-side, impeding the generation of a baseline set of WA findings, and thus also the progress of a general understanding of the WA task.

One of the reasons for the popularity of the syntagmatic-paradigmatic distinction is empirical: this coding scheme has helped to reveal the now well-attested maturational shift from syntagmatic to paradigmatic responses in first language development (Entwisle et al., 1964; Ervin, 1961; K. Nelson, 1977). The discovery of this shift was a key driver of WA research in the 1960s and 70s, and current researchers continue to debate the psycholinguistic nature of the shift (Cronin, 2002; McGregor, Marian, & Sheng, 2006). Some researchers have argued that the shift is driven by the extent to which a word is known: familiar words tend to yield paradigmatic responses, while less familiar ones would be more likely to elicit syntagmatic ones (Postman & Keppel, 1970; Stolz & Tiffany, 1972; Wolter, 2001). Although not well supported by the available evidence, this hypothesis has nevertheless inspired research into a hypothesised corresponding shift in the associations of second language learners. While a small amount of evidence for this has been collected (Politzer, 1978; Söderman, 1993), other researchers (Fitzpatrick, 2006; Kruse, Pankhurst, & Smith, 1987; Nissen & Henriksen, 2006) have found evidence which conflicts with the viability of the hypothesis. Fitzpatrick (2007, 2009; Fitzpatrick, Playfoot, Wray, & Wright, 2015), in particular, attacks the assumption of a syntagmatic-paradigmatic shift in L2 response on the grounds that it masks considerable individual variation in response profiles. She further suggests that this research represents another instantiation of the tendency to assume within-group homogeneity of responses in word association.

Several other methodological challenges have confounded the apparent simplicity of the word association task. There has been disagreement over the number of responses to be collected to each cue word, with some researchers (De Deyne, Navarro, et al., 2012; Schmitt, 1998) arguing that multiple responses allow a greater degree of access into the lexicon of respondents, but others (D. L. Nelson et al., 2000) suggesting that this procedure negatively affects the reliability of associative datasets. De Groot (1989), meanwhile, uses multiple methods of WA elicitation and measurement in a study investigating models of semantic memory. These included individual and continuous associative tests, response time measurement, and analysis of response homogeneity. Other researchers have also used some or all of these methods (Fitzpatrick & Izura, 2011; Playfoot & Izura, 2013; Santos et al., 2011) but little progress has been made in the task of understanding and differentiating their psycholinguistic implications. The absence of any sustained effort to collect WA response time (RT) data, in particular, is unfortunate, because such a dataset would make the task of comparing WA data to the numerous other paradigms for which large-scale RT databases already exist (Balota et al., 2007; Balota, Yap, Hutchison, & Cortese, 2012; Keuleers, Lacey, Rastle, & Brysbaert, 2012; Pexman, 2012).

We have seen, then, that the intuitive nature of the word association paradigm may have led some researchers to assume that the data it provides offers an uninhibited view of the workings of the mind

or the structure of the lexicon. Subsequent research has shown that this is not the case – word association data is, in fact, complex, multi-layered, and often difficult to interpret. We have also seen that some of this complexity has been concealed by a tendency for research either to rely on untested assumptions about the nature of association and the psycholinguistic implications of methodological choices, the homogeneity of respondents, or the interchangeability of cue words; or instead to reinvent the wheel, developing novel modes of categorization and analysis without due attention to the matter of if and how such new methods build upon and explain earlier findings. This leaves WA research in a situation in which few generally accepted research practices exist, and the lack of critical engagement with methodological questions has left a dangerous precedent for disregarding previous findings.

1.3 Setting word association on a firmer foundation

What the above discussion makes clear is that there exists a tension between the need to understand what word association *is*, (a question which requires work with a methodological focus and might view WA data as a dependent variable), and the desire to investigate other, equally important and meaningful research questions *using* word association (work which takes a broader psycholinguistic orientation and treats WA as an independent variable). It should not be taken as a fault of the innovative research currently being conducted on the paradigm that a consistent body of research has not yet emerged, given that one of the benefits of the word association task is that it offers numerous methodological possibilities and can thus be adapted to a wide variety of problems. The challenge for WA researchers now is to preserve this flexibility so that the WA task might continue to be seen as useful for the investigation of a wide variety of phenomena (psycholinguistic and otherwise), while at the same time seeking to establish firm foundations with regard to how different samples of cue words, participants, and different methods of data collection, treatment, and analysis, combine to influence the results which researchers can expect to obtain. In this sense, it is desirable that word association research of the immediate future would seek to address fundamental methodological issues surrounding the paradigm, while at the same time tackling a range of psycholinguistic questions.

There is a growing body of work which sets a precedent for this type of research. Many of these studies will be examined in greater detail in later stages of this thesis; for the present, a few brief examples will be given. Firstly, the work of Prior and her colleagues (Prior & Bentin, 2003, 2008; Prior & Geffet, 2003) has explored the relationship between contiguity-based and semantic accounts of word association formation. In the most recent of their studies, Prior and Bentin (2008) looked at features of sentential contexts which led to the formation of associations between words. Their results suggested that sentences which required participants to integrate the meanings of words were

necessary to facilitate the formation of association. Put another way, these studies suggest that simple co-occurrence may not be enough to enable the development of an association, at least in the absence of multiple exposures to such co-occurrence. Semantic processing thus plays a facilitative role in association formation.

There are two key points to be made about this research. The first is that the source of association has been an important and contentious issue for several hundred years and has, as noted above, seen something of a pendulum effect since the 1960's. It is therefore important that Prior and Bentin chose to address this issue, and the study's finding that these two sources of association may not be as clearly separable as previously assumed ought to be the beginning of a more sustained effort to understand this area. The second point is that Prior's studies have attempted to separate the formation of associations from their production in WATs. This is again a key methodological point; since Bousfield (1953; Bousfield, Cohen, & Whitmarsh, 1958) and Deese, we have known that the mind acts upon linguistic input, organising and connecting words and concepts after initial perception. This research shows that a simple correspondence between input and associative output is unlikely. Thus it is important that more studies follow Prior's lead in attempting to disentangle the processes involved in the formation of association from its production.

Fitzpatrick (2006, 2007, 2009; Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015) is another researcher who has attempted to problematise important areas of WA research. Two assumptions, in particular, have been tested in her research. Firstly, Fitzpatrick took a principled approach to the selection of cue words, eliminating high frequency cues and highly concrete ones on the basis that cues of this type were thought likely to yield a narrow band of responses. Secondly, she investigated the suitability of the popular syntagmatic-paradigmatic categorization system, which has, as shown above, been criticised on the grounds of its lack of psycholinguistic validity, practicality, and detail. Building upon this existing system, Fitzpatrick devised a new system influenced by an analysis of responses in earlier studies, and given a theoretical grounding by Nation's (2001) framework of lexical knowledge. The new system comprised meaning-based, position-based, and form-based distinctions, with sub-categories offering further detail within each group. Finally, using this scheme, she challenged the supposed homogeneity of L1 response patterns which, as described above, appears to have been one of the side-effects of the use of preselected lists of associative cues and norms. Using a test-retest methodology, she found that individuals demonstrate stable preferences for certain types of response, as revealed through her new coding scheme. This result, which was first established using L1 participants (Fitzpatrick, 2007), has since been replicated with L2 respondents (Fitzpatrick, 2009) and with cue words of various grammatical classes (Higginbotham, 2014).

Prior's and Fitzpatrick's work offer a model for future WA research. The defining features of this model are that:

- it builds upon previous research rather than reinventing the wheel, thus respecting the need for continuity to balance criticality.
- it problematises that research by breaking it down into its constituent parts (e.g. formation vs. production of associations) and challenging untested assumptions about theory or method (e.g. group homogeneity; coding schemes).
- While it begins from this position of problematisation, it does not lose sight of wider psycholinguistic debates, such as about the nature of the lexicon or the differences between first and second language processing.
- In the case of Fitzpatrick's work, it takes an iterative approach, implementing gradual methodological refinements in each study. This allows readers to track the impact of methodological changes and thus builds confidence in the findings, as well as allowing the researcher to address different research questions (see de Groot (1989), for another example of the iterative approach applied to WA research).

The remainder of this thesis will attempt to follow this model. It will begin with a replication of Fitzpatrick's (2007) study, described above, before seeking to challenge some of the other assumptions which have limited our understanding of the word association task.

Chapter 2: Revisiting Individual Word Association Profiles: A Replication of Fitzpatrick 2007

2.1. Introduction

In the previous chapter, it was argued that several features of Fitzpatrick's (2006, 2007, 2009; Fitzpatrick *et al.*, 2015) series of studies made it an ideal model for future word association research. These features included its critical engagement with previously untested assumptions; its iterative, incremental approach to building upon existing research findings; and its commitment to informing both methodological concerns and wider psycholinguistic research topics. This thesis aims to build upon Fitzpatrick's body of work, firstly by presenting a replication of one of these studies (Fitzpatrick, 2007), and subsequently by further developing some of the issues which arise from the replication. In this section, the original study will be described in more detail and a detailed rationale will be given for its replication. Several implications emerging from this replication will then be discussed, and an agenda for the next stage of the thesis will be set.

Fitzpatrick's studies were motivated by her earlier work (Meara & Fitzpatrick, 2000) developing Lex30, a word association-based test of the productive vocabulary knowledge of second language learners. A central problem in testing lexical knowledge through word associations is understanding how L2 proficiency manifests itself in word association test (WAT) response type, homogeneity, and resemblance to L1 patterns (Dronjic & Helms-Park, 2014; Fitzpatrick, 2006; Meara, 2009; Qian, 2002; Qian & Schedl, 2004; Schmitt, 1998; Wolter, 2002). Fitzpatrick noted, however, that while numerous studies had revealed differences in the associations given in a first and second language (Meara, 1978, 1983; Riegel & Zivian, 1972; Söderman, 1993), there was significant disagreement on the nature of those differences. As noted in the previous chapter, for example, there has been much disagreement about the existence of a syntagmatic-paradigmatic shift in L2 WAT responses. Ultimately, these questions underpin the credibility of the word association format as a method for assessing L2 lexical knowledge.

2.1.1 Fitzpatrick, 2006

Responding to these challenges, Fitzpatrick (2006) argued that the efficacy of future tests depended on establishing a clearer understanding of L1 and L2 response patterns. She cited the lack of methodological consistency in earlier studies as one reason why this understanding was yet to emerge. She thus put forward a program of research aimed at removing "some of the obstacles to progress in this area by identifying, and then addressing, specific assumptions and weaknesses which have been problematic in previous studies" (Fitzpatrick, 2006: 123).

Although the first of these studies (Fitzpatrick, 2006) is not the one replicated below, it will nevertheless be discussed here in some detail because it is the source of many of the innovations employed in Fitzpatrick's later studies. Although each of these studies investigated unique research questions, Fitzpatrick's approach was essentially iterative, with each study building upon the strengths and addressing the weaknesses of the last. It is therefore important to understand her starting point.

Two of the problems which Fitzpatrick aimed to address in her 2006 study were among those discussed in the introduction. The first was cue selection. Building upon Meara's concerns (Meara, 2009) that this issue had not received enough attention in WA studies, Fitzpatrick identified the dominance of high frequency words and concrete nouns in previous studies as particular problems. She argued that such words "have a stronger 'influence' on the response word than others" and gave as examples the cue-response pairs *bread-butter* and *man-woman* (Fitzpatrick, 2007: 323). In these cases, the "influence" Fitzpatrick describes is that the cue strongly predicts the response. Fitzpatrick suggested that cues which elicit very homogeneous sets of responses are relatively unhelpful to the researcher because they do not allow differences between (e.g. L1 and L2) respondents to emerge: such cues "mask more subtle behaviour patterns" (Fitzpatrick, 2006: 127). Fitzpatrick addressed this problem by selecting cues for several of her studies (Fitzpatrick, 2006, 2007, 2009) from the Academic Word List (Coxhead, 2000). This list excludes words from beyond the 2000 most frequent in English, and thus allows WA researchers to avoid the problem of using too many high frequency cues. Fitzpatrick also states that highly concrete nouns were avoided, although no specific procedure for doing so was described in these studies.

The second methodological problem was the use of the syntagmatic-paradigmatic coding scheme. Fitzpatrick aimed to address two specific issues inherent in this scheme. Firstly, noting the practical and theoretical problems commented on by other authors and discussed in the previous chapter (Deese, 1966; Meara, 1983), she clarified the syntagmatic, paradigmatic, and clang (i.e. phonologically or orthographically similar) categories, recasting them as meaning-based, position-based, and form-based, as well as including an "erratic" category to capture responses which possessed no objectively recoverable relationship. Secondly, Fitzpatrick observed that the broad nature of the syntagmatic-paradigmatic scheme meant that it lacked the detail required to adequately describe every type of response. This again meant that subtle differences in response patterns might not be apparent. To allow more fine-grained distinctions to be identified, Fitzpatrick added sub-categories to each of these basic groups, allowing meaning-based responses, for example, to be further categorised as synonyms, members of the same lexical set, or as conceptual associations. In addition to adding detail to the

categorisation of responses, this process also brought the new scheme into closer agreement with theoretical views of the way words relate to one another (Aitchison, 2012; Nation, 2001).

The final methodological problem identified by Fitzpatrick (2006) concerned “the degree to which the subject’s stimulus-response link is communicated and understood” (*ibid.*: 123) – by which was meant the capacity of the researcher to understand the way in which the respondent had arrived at their response to a cue. Fitzpatrick noted that even with this improved method of coding responses, there would still be examples where the relationship between cue and response would be difficult to understand, or could feasibly be categorised in more than one way. Fitzpatrick’s solution here was to conduct post-test interviews with respondents in order to arrive at a precise categorisation of responses. This solution is somewhat problematic because it assumes that the selection of WAT responses is under participants’ conscious control; or at least that respondents are aware of the processes involved in arriving at their responses. It is, however, possible that these processes are entirely automatic in at least some cases. If so, post-test interviews would reveal little other than *a posteriori* justifications of responses, rather than the actual processes underlying response selection. The issue therefore raises challenging questions about the plausibility of accurately determining the processes by which responses are produced, and will be returned to in the discussion, below.

Fitzpatrick (2006) made several valuable discoveries. Firstly, in addition to supporting previous findings regarding differences between first and second language associations, such as the greater heterogeneity of L2 responses, Fitzpatrick’s new coding scheme revealed significant differences in the type of responses given. In particular, L2 respondents gave more conceptual associations, while L1 participants were more likely to produce defining synonyms. As Fitzpatrick pointed out (2006: 143), these responses would all have been coded as paradigmatic under the previous scheme, thus obscuring this distinction. Finally, although L1 responses were somewhat more homogeneous than L2 ones, the study nevertheless revealed considerable variation in the type of responses given by L1 participants. For example, Fitzpatrick (*ibid.*) reported that the number of position-based responses given by participants to their 60 cues ranged from 2 to 48. As such, the concept of a “nativelike” pattern of responses was not supported by the study.

2.1.2 Fitzpatrick, 2007

In her next study (2007), which will be replicated here, Fitzpatrick took this latter finding as a starting point. Given the significant variation in response patterns between participants in the L1 group, she asked whether individual respondents in fact demonstrated consistent individual response preferences. In practice, this would mean that some respondents would typically give, for example, meaning-based responses rather than position-based ones, while others might prefer synonym

responses over other types of response. Such a finding, she argued, would challenge the long-held but largely untested assumption that WAT participants responding in the same language respond to cues in generally homogeneous ways.

Fitzpatrick used a test-retest methodology to investigate this issue. Two WATs were given at an interval of one week. The cues used in these tests, selected from the Academic Word List (Coxhead, 2000), were not the same, but were matched for frequency and grammatical class. Participants responded in their L1 (English) only, in order to avoid the complication of comparing first and second language responses.¹ The two tests were then statistically compared in order to assess the homogeneity of the group of participants as a whole and the consistency of each participant's responses individually, across the two tests.

Fitzpatrick made two further refinements to her methodology. Firstly, the sub-categories of the meaning-, position-, and form-based coding scheme were updated through the removal or merging of several groups which had proven to capture little variation in response patterns. Examples included the derivational and inflectional affix groups, which were merged into a single "affix manipulation" category, and the "phrasal xy" and "phrasal yx" collocation categories, which became an "other collocation" group. This resulted in the refined 10 sub-category scheme which is reproduced in Table 2.1, below.

Secondly, rather than attempting to repeat the post-test interview format of the preceding study, Fitzpatrick opted simply to rely on her own codings of response types. This allowed a larger number of cues (100) to be used than in the preceding study (60), due to the removal of the significant time demands imposed by the need to interview each participant. This in turn gave a better opportunity for individual response preferences to emerge.

The findings of the study were striking. Fitzpatrick found that 22 of 30 participants' Test 1 response types correlated strongly and significantly with their own Test 2 ($r > .9$); the remaining participants' response patterns correlated less strongly but remained significant at the $p < .01$ in all but one case. This indicates respondents' individual consistency in response type between testing events. Furthermore, while each participant's Test 1 also correlated significantly ($p < .01$) with, on average, 8 other participants' Test 2, indicating some group-level homogeneity, the correlations between participants' Test 1 and their own Test 2 were significantly stronger than correlations between each participant's Test 1 and every other participant's Test 2. In order to demonstrate this, Fitzpatrick

¹ This comparison was conducted in Fitzpatrick, 2009, which found that participants' L1 response preferences are maintained in their L2, with the two sets of responses becoming more similar as a function of proficiency.

calculated Euclidean distances between each participant's Test 1 and Test 2, as well as between each participant's Test 1 and every other participant's Test 2. A t-test comparing these Euclidean distances revealed significantly closer distances for the within-subjects distances than the between-subjects ones ($p < .001$). This indicated that responses were more consistent at a within-subjects level than a between-subjects one. In other words, while there does appear to be a certain level of L1 group-level homogeneity in response types, it is weaker than the response profiles produced by individual participants.

2.1.3 Why select this study for replication?

Of the various WA studies which might be replicated, Fitzpatrick (2007) is a particularly strong candidate because it serves as a useful starting point for this thesis, in that it offers clear affordances for further research. One such opportunity derives from the fact that, while the study's main findings advanced the field of word association by providing a new perspective on individual- and group-level response patterns, the psycholinguistic underpinnings of these patterns remain largely unexplored. Research to date has not, for example, investigated factors which co-vary with particular response preferences. In addition, the methodological innovations which the study offers might be put to good use on investigating other poorly understood aspects of word association, such as the potential for Fitzpatrick's coding scheme to detect "subtle behaviour patterns" (Fitzpatrick, 2006: 127) might be helpful in analysing how L2 responses change over longer periods of time, or as a function of proficiency (Fitzpatrick, 2012).

Before stepping off into this new empirical territory, however, it is important to replicate the study in order to confirm its findings. This will allow the thesis to build upon firm foundations. This is not only a matter of confirming the individual response profiles which Fitzpatrick's work discovered. The high level of methodological innovation which Fitzpatrick's original study attempted is another factor requiring replication. Although, as noted above, Fitzpatrick attempted to balance this innovation with a respect for earlier research, the adoption of new procedures regarding cue selection, coding responses, and analysis of results represents a distinct break with established practice. Perhaps for this reason, few subsequent studies have adopted these innovations (these studies include Zareva, 2012; Zareva and Wolter, 2012; and Van Rensbergen, Storms and De Deyne, 2015; though see Boyum, 2016, for a study which did make use of some aspects of Fitzpatrick's approach). If future research, including that presented in the subsequent chapters of this thesis, are to do so, further evidence for the efficacy and reliability of Fitzpatrick's methods needs to be established.

The replication will be what Porte (2012) terms an "approximate" one, in that in that it is faithful to the original design in every way other than that it draws from a different population type, all of whom

were English language teachers. Other important aspects of the study, such as the cue words, procedure, and statistical methods, were not changed. Porte (2012: p8) suggests that the main reason for conducting such a replication is “to see if the results of the original study are generalisable, for example, to a new population, setting, or modality”. While there is already some evidence that Fitzpatrick’s findings can be extended to other contexts (e.g. Fitzpatrick, 2009, demonstrated similar findings with English L1 learners of Welsh, and Higginbotham (2010, 2014) with L1 Japanese learners of English), this replication uses an older and, in a professional sense, more homogeneous group of participants than previous ones. It is, as such, sufficiently different from the groups used in the above studies to provide an indication of the generalisability of Fitzpatrick’s findings and establish a secure grounding for continuing research.

2.2. Replication design

The principal purpose of this replication is to establish whether findings similar to those reported by Fitzpatrick emerge when conducting the same tests with a markedly different group of respondents. Fitzpatrick’s (2007: p323) original research questions were:

1. Do groups of adult L1 participants respond to cue words in a predictable, homogeneous way?
2. Do individual native speakers respond to cue words in a consistent way?

To facilitate a comparison with the original study, a third question was added to Fitzpatrick’s:

3. Did the replication participants produce WA response profiles like those of the original study?

A comparison of the methodology used in the two studies is set out in Table 2.1.

Table 2.1
Comparison of Fitzpatrick (2007) with the present replication.

Fitzpatrick 2007	Replication
30 participants	32 participants
English L1 speakers “working or studying in an academic environment” in Swansea, UK	English L1 teachers of English as a second language working at a university in Korea
Two 100 item tests, drawn from the Academic Word List, with cue words matched for frequency and word class	The same two tests, presented in the same order
Tests took place at a 1 week interval	Tests took place at a 1-2 week interval
Participants were verbally instructed to “please write down the first word you think of when you read each of the words listed”. No additional written explanation mentioned in Fitzpatrick, 2007.	Participants were given the same verbal instructions as Fitzpatrick, 2007. Test papers also included the written instructions, “Please write down the first word you think of when you read each of the words listed. There are no right or wrong answers.”

2.2.1 Participants

All of the 32 participants (21 male, 11 female) in the replication were employed as English as a Foreign Language (EFL) teachers in a university in Korea. All participants identified English as their first language. Their nationalities were American (20), Canadian (6), English (2), South African (1), Northern Irish (1), Australian (1), and New Zealander (1). Ages ranged between 24 and 64.

2.2.2 The task

Participants completed a 100-word WAT in each of two sessions. The words used in these tests were identical to those used in the original study. The tests were introduced with the words, "Please write down the first word you think of when you read each of the words listed. There are no right or wrong answers". The test documents are reproduced in Appendix 1.

2.2.3 Test administration

Due to the widely differing schedules of the participants in the replication, the first test had to be administered in several sessions consisting of between 1 and 8 participants each. These sessions were held at the participants' convenience over a one-week period. The maximum test time was 15 minutes in all cases. This duration was chosen because previous studies (e.g. Fitzpatrick 2006: p129, Zareva and Wolter, 2012: p49) have suggested that it is sufficient for participants to complete the test, while at the same time encouraging them to respond without thinking too deeply.

The second testing session included all participants in a single session, and was held one week after the end of the first test period. This meant that some participants had a gap of (at most) two weeks between the first and second tests, while others had a gap of one week. All participants gave informed consent to take part in the study.

2.2.4 Categorization of responses

Participants' responses were categorized according to the 10-point scheme developed in Fitzpatrick (2006) and refined in Fitzpatrick (2007). The scheme is reproduced in Table 2.2, below.

2.2.5 Preparing the data for analysis

Prior to the coding of responses, all participants' responses were randomized and anonymised to ensure that the coder (the author) did not sub-consciously impose a response profile upon participants' responses. This anonymised spreadsheet was used to code responses, rather than participants' original papers.

All responses were coded into basic categories and sub-categories according to the scheme set out in Table 2.2, below. The response and sub-category code were then entered into SPSS data analysis software for investigation.

Table 2.2
Scheme for classification of responses (from Fitzpatrick, 2007: 325)

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

2.3. Results

Table 2.3 shows how many responses of each type were given by each participant across the two tests. This data will be discussed below in view of each of the three research questions.

2.3.1 Do adult native speakers respond to cue words in a predictable, homogeneous way?

Fitzpatrick's first research question investigated the extent of homogeneity in response types within the L1 group of participants used in her study. Her aim in doing so was to test the assumption that L1 groups respond to WA cues in homogeneous ways. As in the original study, this replication found some striking differences in the way individuals responded, which did not support this assumption of homogeneity. This is illustrated by the very large differences between the minimum and maximum number of responses in each sub-category, and the high standard deviations for most sub-categories (see Table 2.5). In most cases, standard deviations were between 50 and 100% of the mean for each

category. Illustrative examples of this variation can also be seen throughout the dataset. For example, one participant gave only 3 consecutive xy collocations in Test 1, while another gave 56. Another participant offered no synonyms at all in Test 1, while one gave 46. Together, this data is not suggestive of a homogeneous group “norm” to which the majority of participants conformed. The results, then, support Fitzpatrick’s (2007: p326) conclusion that adult native speakers do not respond in homogeneous ways to the cues used in these studies.

2.3.2 Do individual native speakers respond to cue words in a consistent way?

Given that no group-level response pattern emerged, the next question is whether participants produced consistent patterns of responses across their own two tests. In order to investigate this, correlation analyses were run between each individual’s Tests 1 and 2, and between every participant’s Test 1 and every other participant’s Test 2, in order to investigate the level of consistency between the two testing events.

The results of the correlation analyses between all participants’ first and second tests showed that 31 of 32 participants’ Test 1 correlated significantly ($p < .01$) with their own Test 2. 67% of these correlations were very strong ($r > .9$). These results are very similar to those of Fitzpatrick (2007: 326), who found that 29 of 30 participant’s Tests 1 and 2 correlated at the $p < .01$ level, and that 73% of all correlations were very strong ($r > .9$). These results seem to suggest that individual native speakers do respond in consistent ways to matched WA tests.

Nevertheless, two findings point to a very general group-level homogeneity. Firstly, as in numerous earlier studies (Namei, 2004; Nissen & Henriksen, 2006; Wolter, 2001), position- and meaning-based responses strongly outnumbered form-based ones. Secondly, as in the original study, there are strong and significant correlations in the response patterns given in each participant’s Test 1 and some other participants’ Test 2. Fitzpatrick (2007) calculated the number of significant ($p < .01$; two-tailed) correlations of this type, finding that there were an average of around 8 between-subjects correlations. In the present study one-tailed correlations were calculated in order to offer a more sensitive measure of between-subjects correlation, since the probable direction of effect has already been established by the original study.

The results of this analysis were that each participant’s Test 1 correlated significantly ($p < .01$) with an average of 14 other participants’ Test 2. Given the one-tailed analysis used here, this suggests that the group used in the current replication showed a similar degree of group-level homogeneity in the way they responded to the cue words to that in Fitzpatrick’s study.

Table 2.3

Total number of responses of each type per participant in tests 1 and 2.

Participant	Defining Synonym		Specific Synonym		Lexical Set/ Contextual		Conceptual		XY collocation		YX collocation		Other collocation		Affix change		Form only		No link / Blank	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
1	3	5	31	28	26	20	24	31	7	3	5	5	0	3	1	2	0	0	3	3
2	2	6	36	37	33	23	18	22	5	4	2	2	1	3	1	2	2	1	0	0
3	5	8	49	45	16	20	19	22	5	2	2	1	2	1	2	0	0	0	0	1
4	0	0	5	4	8	8	15	32	39	32	25	10	5	10	0	1	0	0	3	3
5	0	3	15	13	14	22	25	25	31	22	8	8	3	5	1	0	0	0	3	2
6	2	3	15	14	13	14	23	23	30	30	13	7	2	3	0	1	1	1	1	4
7	2	5	22	15	22	21	14	21	21	16	13	13	2	7	0	0	0	0	4	2
8	2	5	14	15	18	16	31	39	15	9	12	7	1	3	1	4	0	0	6	2
9	2	2	6	6	15	14	23	35	29	20	21	8	2	5	0	1	1	4	1	5
10	3	7	38	29	27	30	19	21	3	4	3	4	1	4	2	0	1	0	3	1
11	2	2	21	21	25	27	32	35	9	6	8	2	1	6	1	0	0	0	1	1
12	2	5	18	13	15	14	27	29	20	18	12	12	4	2	1	3	0	1	1	3
13	0	0	0	3	9	5	21	25	41	35	21	20	6	8	0	1	0	0	2	3
14	0	2	4	8	9	8	33	38	34	23	10	10	2	2	0	1	0	2	8	6
15	0	0	3	4	19	12	11	23	43	41	15	8	2	6	1	1	2	1	4	4
16	1	0	2	5	9	5	10	28	56	43	17	11	4	5	0	0	0	0	1	3
17	0	4	4	11	17	15	17	21	33	31	20	7	1	6	1	2	1	1	6	2
18	1	0	9	6	24	12	29	36	17	18	9	9	4	6	0	1	1	2	6	10
19	2	1	7	7	15	3	16	35	40	31	12	9	1	3	0	2	1	3	6	6
20	0	1	5	0	9	9	34	34	34	27	10	13	4	5	8	1	1	1	3	9
21	1	2	5	9	14	24	47	37	20	16	6	5	3	3	1	1	1	0	2	3
22	3	4	22	19	23	15	11	16	29	27	8	9	1	2	2	6	0	0	1	2
23	1	6	25	29	26	35	19	13	14	8	5	5	1	1	4	2	4	1	1	0
24	2	5	26	31	25	21	25	28	9	5	6	6	1	1	0	2	0	0	6	1
25	0	4	16	19	12	15	41	25	10	13	13	10	1	2	0	5	1	3	6	4
26	3	7	21	23	22	17	25	35	14	6	7	4	4	2	4	4	0	1	0	1
27	2	4	17	12	35	24	23	30	9	20	11	4	1	3	1	1	0	0	1	2
28	2	2	26	11	17	23	21	18	19	29	12	5	2	8	1	3	0	0	0	1
29	3	4	36	41	29	27	22	20	3	2	4	1	0	1	1	1	2	0	0	2
30	3	6	22	25	23	24	26	31	10	6	12	3	1	2	0	2	0	0	3	1
31	1	1	5	4	16	4	17	21	41	46	10	11	4	4	2	2	0	0	4	7
32	3	5	15	36	22	28	17	10	26	8	12	4	2	2	2	5	0	1	1	1

Given that this group-level homogeneity does seem to exist both in this replication and its parent study, it is necessary to employ a more sensitive measure of correlation in order to establish whether the consistency shown by individuals was significantly greater than the general level of homogeneity in response type shown throughout the group. As such, following Fitzpatrick's (2007) procedure, Euclidean distances were calculated firstly between each participant's Test 1 and their own Test 2 ($n = 32$); and then between every participant's Test 1 and every other participant's Test 2 ($n = 991$). An independent samples t-test was then used to compare the mean proximities from the two data sets. The resulting t-value suggested that the within-subjects proximities were significantly closer than those between subjects ($t(43) = 13.226$, $p < .001$). The corresponding figure in Fitzpatrick 2007 was $t(45) = 17.563$, $p < .001$. This score suggests a slightly higher level of homogeneity between participants' Tests 1 and 2 in this replication than was found in the original study; possible interpretations of this are offered in the discussion. However, the degree of closeness within-subjects is still very significantly greater than at the group level. This suggests that the response preferences displayed by individual participants exerted a stronger effect than any form of group-level response pattern.

2.3.3 Did the replication participants respond to the cue words in similar ways to those of the original study?

The final research question concerns whether the responses produced by the two sets of participants (Fitzpatrick 2007 vs. the current replication) fell into similar sub-categories. In order to assess the degree of similarity between the two sets of data, Pearson correlations were calculated using the mean number of responses per participant in each sub-category for each of the 4 tests (i.e. two in the original study and two in the replication). The results of this analysis are presented in Table 2.4.

Table 2.4
Correlations between mean sub-category scores in Fitzpatrick 2007 and the current replication

	Thwaites T1	Thwaites T2	Fitzpatrick T1	Fitzpatrick T2
Thwaites T1				
Thwaites T2	.956**			
Fitzpatrick Test 1	.785**	.722*		
Fitzpatrick Test 2	.712*	.711*	.958**	

The correlation scores reveal, firstly, very strong, significant correlations between the response patterns in the within-study tests (i.e. Thwaites Tests 1 and 2, and Fitzpatrick Tests 1 and 2). This finding is not surprising, given the above evidence that participants responded to the two tests in consistent ways. Secondly, the cross-study correlations were weaker than those within each study, although they were all significant and relatively strong. This again suggests a certain amount of within-groups homogeneity. This homogeneity does not appear to be attributable simply to a shared L1, since all participants identified English as their first language. The descriptive data presented in Table 2.5, and in graphic form

Figure 2.1
 Mean number of responses of each sub-type given by participants in the current replication and Fitzpatrick (2007).

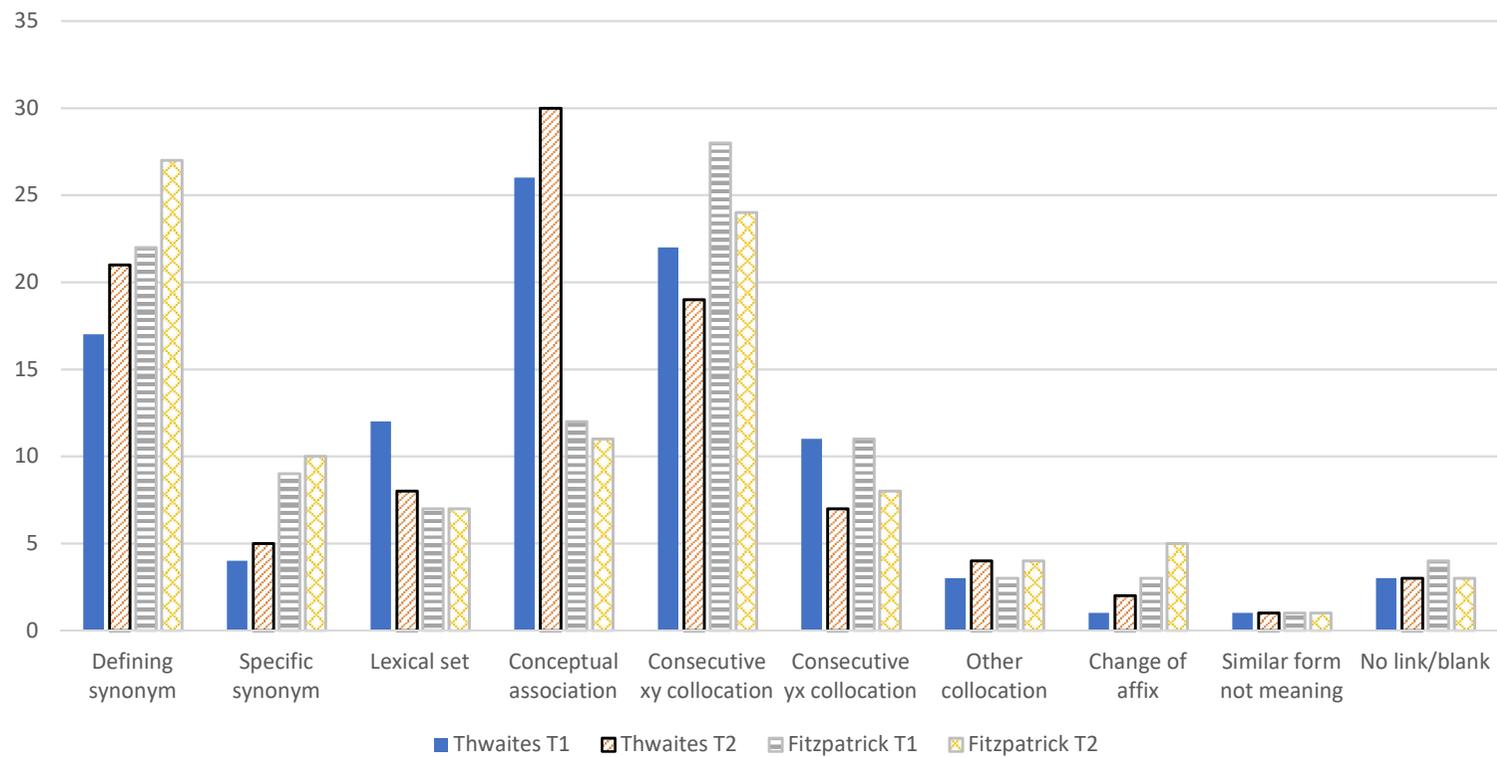


Table 2.5

Comparison of minimum, maximum, and mean number of responses given per participant in the current study and Fitzpatrick (2007). Each participant gave 100 responses.

Category	Current Study						Fitzpatrick (2007)					
	Min		Max		Mean (sd)		Min		Max		Mean (sd)	
	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
Defining synonym	0	2	46	57	17 (12)	21 (15)	2	1	55	57	22 (15)	27 (16)
Specific synonym	0	0	13	12	4 (4)	5 (3)	0	2	26	27	9 (6)	10 (6)
Lexical set	4	1	22	17	12 (5)	8 (4)	1	1	13	12	7 (3)	7 (3)
Conceptual association	12	12	48	42	26 (8)	30 (7)	4	3	27	29	12 (6)	11 (6)
Consecutive xy collocation	3	2	56	46	22 (14)	19 (13)	5	2	57	51	28 (15)	24 (14)
Consecutive yx collocation	2	1	25	20	11 (6)	7 (4)	1	2	29	30	11 (6)	8 (6)
Other collocation	0	1	9	10	3 (2)	4 (2)	0	1	8	12	3 (2)	4 (3)
Change of affix	0	0	4	6	1 (1)	2 (2)	0	0	13	30	3 (3)	5 (6)
Similar form not meaning	0	0	4	4	1 (1)	1 (1)	0	0	6	6	1 (2)	1 (1)
No link/blank	0	0	8	10	3 (2)	3 (2)	0	0	41	37	4 (9)	3 (8)

Table 2.6

Comparison of minimum, maximum, and mean number of responses given per cue in the current study. Each cue received a total of 32 responses.

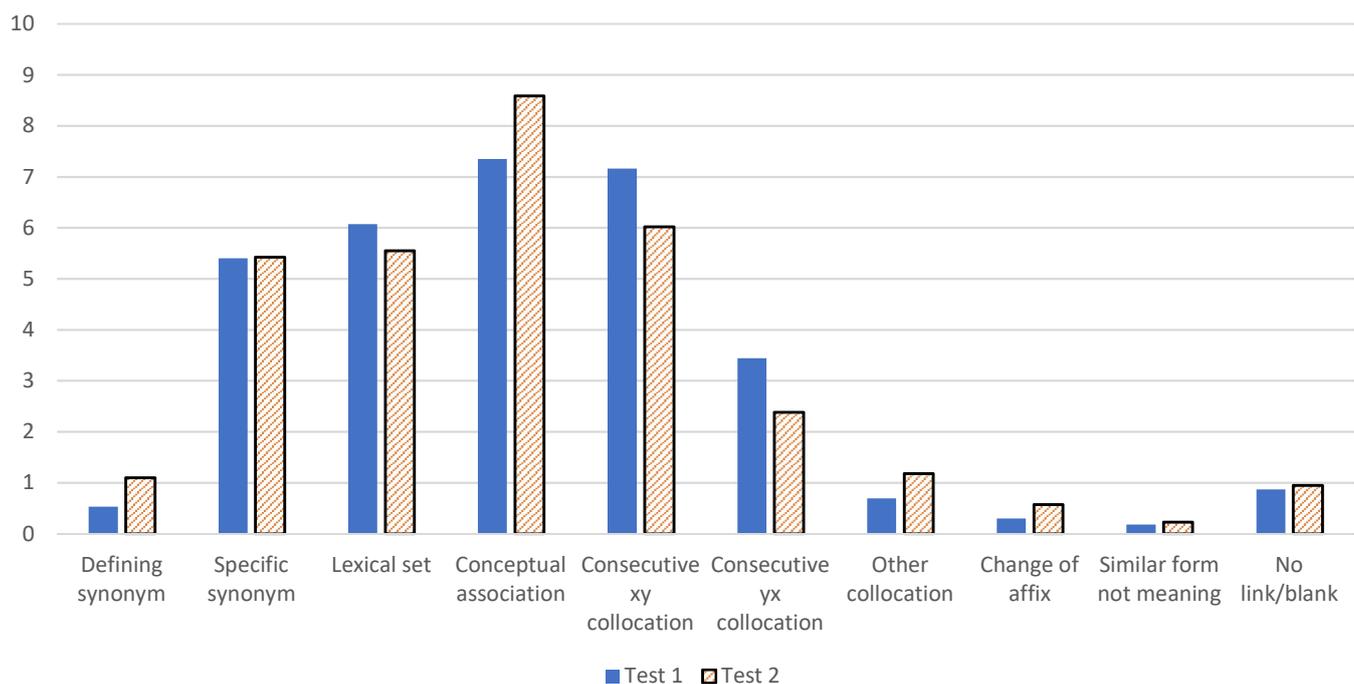
	Test 1			Test 2			Total		
	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)
Defining synonym	0	19	.53 (2.73)	0	18	1.1 (3.62)	0	19	.81 (3.21)
Specific synonym	0	20	5.4 (5.11)	0	24	5.43 (5.41)	0	24	5.42 (5.25)
Lexical set	0	27	6.07 (5.65)	0	22	5.55 (4.57)	0	27	5.81 (5.13)
Conceptual association	0	21	7.35 (5.16)	1	26	8.59 (5.64)	1	26	7.97 (5.43)
Consecutive xy collocation	0	25	7.16 (6.59)	0	26	6.02 (5.95)	0	26	6.59 (6.29)
Consecutive yx collocation	0	23	3.44 (4.79)	0	14	2.38 (3.28)	0	23	2.91 (4.12)
Other collocation	0	6	.69 (1.21)	0	14	1.18 (2.5)	0	14	.94 (1.97)
Change of affix	0	9	.3 (1.01)	0	8	.57 (1.38)	0	9	.44 (1.23)
Similar form not meaning	0	2	.18 (.41)	0	5	.23 (.68)	0	5	.20 (.56)
No link/blank	0	6	.87 (1.16)	0	7	.95 (1.35)	0	7	.91 (1.26)

in Figure 2.1, reveals the nature of some of the differences between the two groups. Similar numbers of lexical set and yx collocation-type responses, as well as form-based and erratic response types, were given. The most striking difference was in the number of conceptual associations, with participants in the replication study producing an average of 28 responses of this type, compared to 11.5 in the original study. This discrepancy appears to be made up largely by the greater production of synonyms, xy collocations, and affix manipulations by the Fitzpatrick group. The differences may be attributable to the characteristics of the two participant groups. This possibility is explored further in the discussion.

2.3.4 Cue level variation in responses.

During the coding of responses for the above analysis, it became apparent that each cue tended to receive responses of a certain type, much in the same way that each participant preferred responses of certain types over others. Some cues, for example, received numerous semantic responses, while others received more position-based ones. Examples of this variation are set out in Table 2.7. These examples are paired by grammatical class and frequency (as determined by the sublist of the Academic Word List from which they were drawn; see below for further discussion of this frequency measurement). As noted in Chapter 1, several researchers have suggested the need for a deeper understanding of how the properties of cues in WA studies influence responses (De Deyne & Storms, 2015; Fitzpatrick et al., 2015; Meara, 2009). As such, a brief exploration of this variation is presented below in order to assess the extent of the issue, and suggest some contributing factors.

Figure 2.2
Mean number of responses of each sub-type per cue in the current study



Some cue-level variation in response types is visible in Table 2.7. These cues are, however, only examples. A clearer picture emerges from Table 2.6, where the range in the number of responses of each sub-type to a single cue was wide. The lexical set category saw the widest range, with some cues (e.g. *vision, migration, significant, select*) receiving no responses of this type, while one (*currency*) received 27 (i.e. 85% of all responses to that cue; note that most cues with high scores in this category were nouns. A possible reason for this will be discussed in Chapter 4). A test of the interquartile range (IQR) between responses suggested that these high and low values were not statistical outliers: none of the response

Table 2.7
Response type distributions for a sample of cues, paired according to like frequency and grammatical class.

Cue	Test	AWL Sublist	GC	DS	SS	LS	C	XY	YX	OC	A	FO	N/B
Negative	2	3	adjective	0	6	12	4	10	0	0	0	0	0
Physical	1	3	adjective	0	1	4	4	23	0	0	0	0	0
Likewise	1	10	adverb	0	11	5	11	0	0	0	0	2	3
Nonetheless	2	10	adverb	0	7	12	7	0	2	0	0	0	4
Income	1	1	noun	0	0	16	6	5	5	0	0	0	0
Interpretation	2	1	noun	0	3	0	17	0	2	9	0	0	1
Subsidiary	1	6	noun	1	1	2	15	12	0	0	0	1	0
Transformation	2	6	noun	16	0	1	1	0	0	0	4	5	5
Preceding	1	6	verb	0	19	3	2	1	3	1	0	0	3
Underlying	2	6	verb	0	4	2	5	19	0	0	0	0	2

AWL Sublist = The frequency-based sublist from which the cue was selected; lower numbers are more frequent; GC = Grammatical Class; DS = Defining Synonym; SS = Specific Synonym; LS = Lexical Set; C = Conceptual Association; XY = Consecutive XY Collocation; YX = Consecutive YX Collocation; OC = Other Collocation; A = Change of Affix; FO = Form Only; N/B = No link/Blank.

Table 2.8
Comparison of minimum, maximum, and mean number of responses given per cue at the highest and lowest frequency bands. The maximum number of responses to each cue was 32.

	Frequency Band 1			Frequency Band 10		
	Min	Max	Mean (SD)	Min	Max	Mean (SD)
Defining synonym	0	17	1.86 (4.67)	0	13	1.05 (3.14)
Specific synonym	0	20	6.19 (5.85)	0	13	6.11 (4.52)
Lexical set	0	16	4.24 (4.34)	0	13	7.16 (3.7)
Conceptual association	1	18	8.29 (5.55)	1	20	8.21 (5.5)
Consecutive xy collocation	0	17	5.43 (4.95)	0	15	5.79 (4.54)
Consecutive yx collocation	0	9	2.71 (2.70)	0	11	1.26 (3.02)
Other collocation	0	9	1.29 (2.03)	0	12	.89 (2.73)
Change of affix	0	8	.81 (1.78)	0	3	.21 (.71)
Similar form not meaning	0	1	.14 (.36)	0	2	.11 (.46)
No link/blank	0	3	1 (1.18)	0	4	1.21 (1.36)

type analyses reached Hoaglin and Iglewicz's (1987) recommended IQR threshold of 2.2. Moreover, the high standard deviations in each sub-category, which generally approach and in some cases exceed 100% of the mean, further emphasise the heterogeneity in cue-level response patterns.

Table 2.9

Comparison of minimum, maximum, and mean number of responses given per cue according to grammatical class (adverbs are excluded because only six cues of this type were used in the study).

	Nouns			Verbs			Adjectives		
	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)
Defining synonym	0	18	.6 (2.86)	0	19	1.43 (4.18)	0	18	.73 (3.1)
Specific synonym	0	24	4.67 (5.18)	0	20	8.1 (5.86)	0	14	5.31 (4.59)
Lexical set	0	27	6.18 (5.75)	0	14	4.7 (3.61)	0	18	5.65 (4.56)
Conceptual association	0	26	9.01 (5.84)	1	16	6.63 (3.99)	0	21	6.39 (5.03)
Consecutive xy collocation	0	20	4.16 (4.98)	0	19	6.23 (5.02)	0	26	12.53 (5.87)
Consecutive yx collocation	0	23	4.8 (4.6)	0	4	.8 (1.13)	0	5	.2 (.83)
Other collocation	0	11	1.02 (1.78)	0	14	2.2 (3.35)	0	1	.12 (.33)
Change of affix	0	9	.5 (1.32)	0	4	.3 (.92)	0	8	.39 (1.23)
Similar form not meaning	0	5	.25 (.65)	0	2	.23 (.5)	0	1	.08 (.27)
(No link/blank	0	5	.81 (1)	0	7	.137 (1.59)	0	6	.61 (1.17)

Fitzpatrick (2006, 2007) controlled for two factors which she felt may influence responses. These were grammatical class and frequency. Given that several researchers have also reported an influence of these factors (e.g. Deese, 1966; Entwisle, 1966a, 1966b; Sökmen, 1993), their influence is worth assessing. Tables 2.8 and 2.9 isolate different categories of frequency and grammatical class respectively. In both sets of data, significant variation in response patterns within each sub-group of cues remains evident. The range in responses of each type remains wide, and standard deviations generally close to the mean, and occasionally far higher.

It would, of course, be desirable to conduct a more detailed analysis of this variation. However, the morphological makeup of the cues selected by Fitzpatrick makes such an investigation difficult. Fitzpatrick selected cues from the Academic Word List (AWL; Coxhead, 2000), which is built around word families. Each headword on the list subsumes between 1 and 20 derivations. For example, the headword *error* subsumes *erroneous*, *erroneously*, and *errors*. The cues in the present study were chosen from within these headword-based lists. Fitzpatrick's procedure was to take 10 words per test (i.e. twenty in total) from each of the 10 frequency bands of the AWL, matching these 10 words for grammatical class across the two tests. Because both headwords and derivations were eligible for selection, the resulting cue list contained numerous derived words.

There are two reasons why this complicates a deeper analysis of the influence of cue frequency and grammatical class. Firstly, the frequency of each cue was determined not its actual frequency, but by the frequency of its headword. For example, the frequency band of the word *subsidiary* was in fact determined by the frequency of *subsidy*. Thus, the actual frequency of the words in the study is not reliable. Secondly, derivational affixes can sometimes look like inflectional ones. Several of the cues in the study were ambiguous in this sense; it is not clear whether *devoted*, *compiled*, and *published*, for example, were intended as verb or adjective cues. An inferential analysis of the impact of grammatical class on response types would therefore be complicated by this uncertainty.

Finally, in view of the above discussion, it is important to note that Fitzpatrick's procedure for selecting cue words clearly resulted in two very similar tests, with highly correlated patterns of responses, in spite of the cue-by-cue variation in response types. This is demonstrated by the within-studies correlations in Table 2.4. It is also apparent in the similar distribution of response types per cue across the two tests, shown in Table 2.6 and Figure 2.2. This suggests that although there may have been variation in the way the cues in the two tests were responded to, they appear to have varied in roughly similar ways across the tests. As such, cue-level response variation does not seem to have affected the general pattern of results presented in Sections 2.3.1, 2.3.2, and 2.3.3, above.

In summary, the descriptive statistics given in Tables 2.6, 2.8, and 2.9 suggest that, firstly, cue-related variation in response types may be as large as the variation in responses attributable to participant-level differences; and secondly that this variation may not be reducible simply to cue frequency or grammatical class, because large variation exists even within these groups. However, it has not been possible to conduct an inferential analysis of this variation. This will be taken up in the section 2.4.3.

2.4. Discussion

In general, the results obtained in this replication support the findings of Fitzpatrick (2007). In particular:

- For every participant but one, their Test 1 pattern correlated strongly with the test 2 pattern
- There were also significant correlations between each participant's Test 1 and some other Test 2's in both tests, meaning that there is some degree of group-level homogeneity in responses. This homogeneity appears to have been slightly higher in the replication.
- A t-test comparing Euclidean distances within-subjects to between-subjects showed that the within-subjects proximities (i.e. the similarity between each participant's Tests 1 and 2) were significantly closer than between-subjects proximities, in both tests; the effect was very slightly weaker in this replication than in the original study).

The picture thus emerging here is of a given participant producing patterns of responses which are quite similar to some participants, but quite dissimilar to some other participants, within their group. At the same time, each participant is very likely to produce a similar response pattern each time they do the test, rather than conforming to any L1 response norm.

Beyond these similarities in the results from the two studies, several other issues and avenues for future research emerged in this replication. These will be explored below.

2.4.1 Differences between groups

While the group used in Fitzpatrick (2007) and the current group both showed individual patterns of response which were stronger than any group “norm”, they also produced group-level response patterns: correlations between each group’s test 1 and test 2 were stronger than those between the two groups. This suggests that while individual response patterns certainly exist, there is nevertheless some basis to the idea of group-level homogeneity in WA response patterns. Indeed, other researchers have commented on some potential sources of group level homogeneity. Fitzpatrick *et al.* (2015), for example, found that participant age influenced response homogeneity, with older participants producing more homogeneous sets of responses compared with those produced by younger participants (see also Hirsh & Tree, 2001), and also noted that “it is possible that other factors such as educational background or gender might also affect response norms” (Fitzpatrick *et al.*, 2015: 24).

In the case of the present study, it appears that the replication group responded in slightly more homogeneous ways than Fitzpatrick’s (2007) group: the latter yielded a lower *t*-value in the comparison of within- and between-subjects Euclidean distances. It is possible to speculate that the higher mean age of the participants in this replication, or their professional and educational background, may have been a factor here: the replication group did not contain students and therefore was likely to have had a higher mean age and greater professional homogeneity. However, given that neither the original study nor the replication collected detailed demographic data from its participants, it is not possible to test this hypothesis. A straightforward methodological recommendation for future studies, then, is to collect such data (see Fitzpatrick *et al.*, 2015, for suggestions as to specific types of information which might be worth collecting).

2.4.2 Subjectivity of coding

Perhaps the clearest difference between the two groups concerned the higher number of conceptual associations given by the replication group. While the simplest explanation of this finding is that the age-, profession-, or education-related differences between the groups influenced their response preferences, an alternative explanation might lie in the decisions made by the researchers during coding.

These decisions can be quite subjective. On occasions, differences in coding can be put down to differences in world knowledge, such as when responses derive from pop music (*chemical-brothers*; *chemical-romance*) or movies (*deep-impact*). On other occasions, a coder may fail to notice a particular sense or usage of a word. The cue-response pair *intervention-help*, for example, was initially coded as a conceptual link because *help* was taken to be a verb. It was only on a second run through the data that its potential role as a noun was noticed, and it was switched to the defining synonym category.

The discovery of individual response profiles in WA suggests a further reason to suspect that two different researchers, working alone, would make different decisions regarding the coding of some responses. Put simply, if individual WAT participants display individual response preferences, it may also be the case that coders have preferences in the way that they categorise responses. For example, this may result in them tending to code ambiguous responses in a certain way, or even to fail to notice a particular relationship between a cue and its given response.

On a practical level, one solution here is to employ a second coder. This would, at the very least, allow a greater awareness of the different possible relationships between cue-response pairs and reduce the chances that different possible codings would go unnoticed. This is the approach taken in the subsequent chapters of this thesis. A further solution, used by Fitzpatrick and Izura (2011) and Fitzpatrick *et al.* (2015), is to create response categories comprising more than one category (for example the “lexical set and collocation” categories used in Fitzpatrick *et al.*, 2015: 19). Such a scheme might help to reduce a coder’s bias when handling ambiguous cue-response pairs.

On a more theoretical level, the possibility that different coders code responses in different ways should be seen as one of the important research implications of the current study. As the early stages of this thesis have made clear, WA research up to the present has been troubled by inconsistent research findings. In addition to the numerous methodological issues discussed above, differences in coding preferences may be another source of this inconsistency. As such, it should be a priority for future research, and will be returned to in Chapter 7.

2.4.3 Cue-related response variation

Finally, while inferential analysis of variation in response patterns at the level of individual cues has not been possible in this study, the descriptive data does suggest that the distribution of response types given by participants is influenced by the cue itself; and that this variation is not reducible to the influence of cue frequency or grammatical class alone.

Research into the influence of cue-related factors on responses is still at an early stage, but numerous researchers have put forward variables which may play a part in determining response types. In addition

to frequency and grammatical class (Deese, 1962a; Entwisle, 1966a; Meara, 2009), these include concreteness and imageability (de Groot, 1989; van Hell & De Groot, 1998) and affective factors such as valence (the extent to which a concept is considered affectively positive or negative; Van Rensbergen, Storms and De Deyne, 2015; Van Rensbergen, Deyne and Storms, 2016). Chapter 3 of this thesis will explore this research in more detail.

This replication suggests that cue morphology may also have some influence over responses. In section 2.3.4, it was suggested that this factor complicated the analysis of the impact of cue frequency and grammatical class. However, it is also possible that morphology may itself be a useful candidate as an explanatory factor in word association. This is because the morphological characteristics of some cues appear to have influenced responses. While there were few responses of the affix manipulation type in either Fitzpatrick (2007) or this replication (on average around 2 per participant), in some cases, participants responded to derived words with an associate of the headword (*construction-build*; *contradiction-agree*). While it may simply be that participants in these situations are omitting affixes from their responses, other explanations remain possible. For example, it may be that participants are in some cases choosing to process only a word's stem, and not its affix (see Marslen-Wilson *et al.*, 1994, for a precedent for this type of processing).

This data presented in this replication therefore adds to calls for further research on the impact of cue-level variation on responses (De Deyne & Storms, 2015; Fitzpatrick, 2009; Meara, 2009).

2.5. Conclusions

This replication confirms the main findings of Fitzpatrick's (2007) original study. In particular, it supports the finding that individuals demonstrate preferences in the type of responses they produce in word association tests. While a certain amount of homogeneity does appear at the group level, these individual profiles appear to be stronger than the group "norm".

Several studies have now demonstrated consistent patterns of WA response for a given individual. Evidence as to the conditions under which these profiles emerge extends to the following circumstances:

- L1 participants responding to words taken from the AWL (Fitzpatrick, 2007, and this study)
- English L1/Welsh L2 bilinguals responding to words in both languages, taken/translated from the AWL (Fitzpatrick, 2009)
- Japanese L1/English L2 bilinguals responding to English words at two different levels of frequency (0-500; 500-1000 most common in English; Higginbotham, 2010)
- Japanese L1/English L2 bilinguals responding to English words from a single grammatical class (verbs, adjectives; Higginbotham, 2014)

Moreover, Higginbotham's (2014) investigations into the influence of cue-related features on these profiles led him to conclude that, for Japanese L1/English L2 participants, the frequency and grammatical class of the stimulus word had little effect on the consistency of the individual response profiles.

Two natural directions for future research are suggested by the findings presented above. The first is to look at WAT response profiles in more detail. This might involve exploring, for example, whether these profiles reflect differences in the organization of individual lexicons, or instead emerge from individuals' approach-to-task (Fitzpatrick, 2009). This might initially be tested by looking at whether responses of a participant's preferred type are produced more quickly than those of other types (Fitzpatrick & Izura, 2011). The second direction is to investigate the factors which co-vary with particular response patterns. Fitzpatrick (2007: 327-8; 2009) has suggested several factors which may predict individual response patterns. These include age, creativity, personality, intelligence, and emotional maturity. Another potentially fruitful distinction is cognitive style (Blajenkova & Kozhevnikov, 2009; Cools & Van den Broeck, 2007). Bergen (2012), for example, provides several examples of ways in which visual and verbal styles influence language processing, and suggests that these differences might extend to the way in which word associations are produced or recognised.

Neither of these routes will, however, be taken here. Instead, a closer analysis of lexical influences on word association, such as those of cue frequency and grammatical class will be undertaken. The chief reason for this choice is that the large variation in the of responses given to superficially similar cues (see Tables 2.7, 2.8, and 2.9, above) suggests that a principled examination of the differences between response patterns given by individual respondents is unlikely to yield straightforward interpretations without a more detailed understanding of the way in which cue properties influence responses.

There are other reasons for undertaking such research. Variation in cue characteristics may, for instance, be one explanation for the inconsistency of results in earlier research in word association. There has been a remarkably low level of consistency in cue selection for WA research. As noted in the previous chapter, several studies have made use of frequent, concrete, largely noun-based cues, such as those in the Kent-Rosanoff list (Kent & Rosanoff, 1910; Namei, 2004; Söderman, 1993). Others have attempted to balance factors such as frequency and grammatical class (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2015; Nissen & Henriksen, 2006; Wolter, 2001), but have varied in other ways – for example by selecting from different frequency levels or by using words of a specific register, such as the morphologically complex academic vocabulary selected by Fitzpatrick (2006, 2007, 2009). It is therefore not surprising that results have been inconsistent. A dedicated research agenda focused on cue-level variation might shed light on the extent to which cue selection protocols have contributed to

this inconsistency, as well as clarifying factors which need to be considered when selecting cues in future research.

Furthermore, the lack of understanding of cue-related factors in word association can be contrasted with the collective effort being made to map the influence of stimulus variation on other psycholinguistic tasks. Analysis of covariation between words and tests such as lexical decision and word naming is yielding new insights into lexical processing (e.g. Balota, Yap and Hutchison, 2007; Yap *et al.*, 2011; Balota *et al.*, 2012; Pexman, 2012; Taikh *et al.*, 2014; Yap, Lim and Pexman, 2015; Sidhu, Heard and Pexman, 2016). An illustrative example is Yap *et al.* (2011), which tested the influence of more than a dozen lexical and semantic variables, across more than 500 stimulus words, on response times and accuracy rates for three psycholinguistic tasks, in order to investigate “how the effects of different semantic dimensions are selectively and adaptively modulated by task-specific demands” (*ibid.*: p742).

Two assumptions underlie these studies. Firstly, it is assumed that differences in word characteristics equate to differences in processing. This means that, for example, concrete words will be processed differently (perhaps using visual or tactile processes) from abstract ones (which might involve episodic or emotional processes; Simmons *et al.*, 2008; Santos *et al.*, 2011; Wilson-Mendenhall *et al.*, 2011; Connell and Lynott, 2012). Secondly, these characteristics interact with psycholinguistic tasks in both task-general and task-specific ways. For example, semantic variables in general account for variation in response times on tasks such as lexical decision (LDT) and semantic classification (SCT). However, the strength of this influence varies according to both the specific features of each stimulus word, such as its concreteness or the number of senses in which it is used, and the demands of each task. For instance, Yap *et al.* (2012) found that the SCT is more strongly influenced by semantic variables than is lexical decision. On the other hand, LDT is more strongly influenced by distributional factors such as frequency and semantic neighborhood density. Comparisons of the influence of such factors across a wide range of tasks allow inferences to be drawn as to the types of knowledge which underlie lexical access – surface level (i.e. their phonological or orthographic realization), distributional (e.g. frequency, age of acquisition) and semantic (e.g. imageability, number of senses).

There are two ways in which the field of word association can gain from research of this type. Firstly, by adopting a method in which cue variables are systematically explored as independent variables in WATs, it will be possible to explore the extent to which cue-related variables encode differences in processing in the word association task. This might, in turn, lead to insights into the nature of the WA task itself. Secondly, studies such as those described above tend not to view word association as a psycholinguistic task in its own right, but rather treat it as an independent variable equating to semantic relatedness or diversity (Yap *et al.*, 2011, Hutchison, 2003; Hutchison *et al.*, 2008). As discussed in the previous chapter,

this is somewhat troubling because so little research has been dedicated to investigating the underlying nature of word association and the types of variables which influence it. For example, it remains to be seen whether the task is genuinely semantic in nature, as is frequently assumed (De Deyne & Storms, 2015). It also somewhat undervalues WA as a psycholinguistic task capable of contributing to our understanding of lexical processing in its own right.

As such, the remainder of this thesis will focus on the influence of individual word properties on responses. The next chapters will review previous research in this area. Subsequently, several empirical studies will be presented with the aim of providing new insights into the role of cue properties on the WA task.

Chapter 3: Cue word influence on word association behaviour.

3.1. Introduction

The previous chapter presented a replication of Fitzpatrick's 2007 study. The results supported Fitzpatrick's discovery that individual respondents generate consistent profiles of word association (WA) responses, and extended the scope of this finding to a population which had a more homogeneous professional and educational background than those previously studied.

The replication also raised the issue of the influence of cue selection on response categories. It was discovered that the cues used in the study were not responded to in homogeneous ways: there was considerable variation in response categories, exemplified by high standard deviations from the mean number of responses of each category to each cue, with accompanying wide ranges in the number of responses of each category (see Table 2.6, in the previous chapter). This marked variation remained even when the frequency and grammatical class of the cues were held constant.

The potential benefits of researching this area are compelling. They include the possibility that a deeper understanding of the influence of cue variation will allow a reanalysis of existing research, the production of guidelines for the selection of cues in future research, and the possibility that word association will be able to contribute more to, and be better understood by, future studies in the wider area of psycholinguistics, both as a dependent and an independent variable.

The aim of the current chapter is to review what has been established about cue-level variation to date. Because WA studies vary in the way in which they measure responses, the review will begin by summarizing the main WA measurements (i.e. dependent variables). It will then look at the impact of several aspects of cue-level variation on WA responses in existing research. A broad distinction will be drawn between distributional variables, such as frequency, contextual diversity, and age-of-acquisition; and semantic variables, which include concreteness, imageability, and emotionality. The chapter will close with a discussion of the aspects of cue-level variation best suited to further research².

² Numerous studies have looked (mostly indirectly) at the impact of lexical variables – particularly cue frequency – on second language (L2) word association (de Groot, 1989; Dronjic & Helms-Park, 2014; Fitzpatrick, 2009; Meara & Fitzpatrick, 2000; Nissen & Henriksen, 2006; Qian, 2002; Qian & Schedl, 2004; Read, 1993, 1995; Söderman, 1993; Sökmen, 1993; van Hell & De Groot, 1998; Wolter, 2001, 2002, Zareva, 2007, 2011, 2012; Zareva & Wolter, 2012). However, because the focus of the present thesis is L1 association, and because it must not be assumed that WA participants are influenced by lexical variables in the same way when responding in an L2 as in an L1 (Grosjean, 1989; Izura & Ellis, 2004; Klein, Mok, Chen, & Watkins, 2014) this review will focus on L1 studies. The

The first question to be addressed below is how word association has been measured. After that, attention will be turned to how these measurements interact with several aspects of cue-level variation.

3.2. How is word association measured?

In the study reported in Chapter 2, a single dependent variable, *response category*, was used to measure responses. This refers to the categorisation of cue-response pairs according to the linguistic relationship between them. However, numerous other WA measures have been employed. This variation in measurement type is fundamental to the flexibility of word association research because it allows researchers to address a wide range of questions. It is also, as will become clear below, critical to the investigation of cue-level variation, because some patterns of influence are largely invisible until they are revealed by the appropriate measure.

However, this variation also contributes to the sense of unfulfilled promise discussed in the Chapter 1, because the wide range of measurements used to date has made it very difficult to compare studies. One of the purposes of this review, then, is to offer a holistic picture of how cue-level variables interact with a range of dependent measures, and to suggest what these measures might mean in terms of lexical organisation.

The most widely-used methods of measuring WA data are described in Table 3.1. The first three columns of the table provide descriptive information. The fourth column of the table demonstrates that most measurements tend to be associated with specific research aims, such as the use of group norming procedures to investigate differences between populations, or the use of association strength to predict and explain variation in psycholinguistic tasks. Perhaps because of this goal-oriented approach to WA research, there currently exists only a small handful of studies which have simultaneously collected data of different types. This is a weakness in WA research because it means that overviews such as the one which follows are unavoidably fragmentary, in that they can only report information on all relevant variables from the point of view of the small number of dependent measures which have been used to investigate them. There are signs, however, that some researchers are beginning to address this problem (e.g. Playfoot & Izura, 2013).

only L2 studies to be discussed are those in which both L1 and L2 responses have been collected (e.g. de Groot, 1989; Nissen & Henriksen, 2006; Van Hell & De Groot, 1998; Wolter, 2001). In these cases, care will be taken to make clear which findings refer to first language participants, and which to second.

The final column of Table 3.1 gives examples of the influence of cue-level variation on these measurements, where such findings are available. These examples demonstrate that almost all WA measures offer some possibility for the investigation of cue-related factors. Only the use of group norming procedures, which are typically used to assess differences at the level of participants and groups, seem unsuited to cue-level analysis.

Several different ways of collecting WA data are implied in Table 3.1. The simplest and most common method is the discrete, single-response WA task, which involves the collection of a single response to each cue word. This task can be timed in order to collect response time data. This works either through the use of a voice-activated timer, which records voice onset time (see Simmons et al., 2008, for an example), or using a button-press timer, after which respondents are required to type responses into a computer (see Fitzpatrick & Izura, 2011). Discrete multiple-response tests are similar to single response tests, but require a specific number (often three) of responses per cue, and cannot be timed (although some studies do make the assumption that second and third responses are produced more slowly than first responses, and can thus be compared to items produced slowly in timed single-response WATs; see De Deyne & Storms, 2008). Lastly, in continuous WA tasks respondents are asked to produce responses continuously for a given period of time, or until they are able to give no more (see de Groot, 1989, for an example).

3.3 The lexicon as a network

Implicit in several of the measurement descriptions in Table 3.1 is a network model of the lexicon. Since the network paradigm has become dominant in most areas of psycholinguistics, including WA research, in recent years (De Deyne, Navarro, et al., 2012; A. W. Ellis & Lambon Ralph, 2000; Gruenenfelder et al., 2016; Miller & Fellbaum, 1991; Vitevitch, Chan, & Goldstein, 2014; Wilks, 2009), a brief discussion of the topic is required.

Network models are simulations of hypothesised neural structures. In the present case, this structure is the mental lexicon, which is the network in which word knowledge is stored. Such networks are based on the principle that *nodes*, which can represent either words (i.e. *localist* networks: De Deyne et al., 2013; De Deyne & Storms, 2008, 2015) or smaller lexical and semantic features such as phonemes or semantic features (*distributed* networks: Van Hell & de Groot, 1998), are connected via *links* (or *edges*). Networks can be either *directed*, meaning that they are sensitive to the direction, in-coming or out-going, of links between nodes; or *undirected*, meaning that they treat incoming and outgoing links as the same. Networks can also be *weighted*, meaning that the links between nodes are assigned a specific strength; or *unweighted*, meaning that all links are considered equally strong.

Several theoretical properties of the mental lexicon are implicated in these networks. The first is termed *connectivity*, which refers to the number of links, incoming and outgoing, possessed by either by a single node, or by the entire network. The second property, *connection strength*, refers to the weight of those connections, and is determined by the frequency with which two words are co-activated. In WA networks, this is estimated using association strength (note that only weighted networks model this property). A final network property, *centrality* (De Deyne, Navarro, et al., 2012; De Deyne & Storms, 2008), refers to the importance of a given word to the network. Centrality measures are derived from the two earlier properties: a word which is central to the lexicon will possess a large number of strong incoming and outgoing links to other words in the network (i.e. high connectivity and connection strength); certain measures of centrality will also reference the extent to which these words are themselves connected to still other words (De Deyne & Storms, 2008, 2015).

While the network metaphor has become dominant in lexical studies, a few words of caution are required. Most importantly, all artificial lexical networks, whether based on word association or another data source (Andrews, Vigliocco, & Vinson, 2009; Steyvers & Tenenbaum, 2005) are simplified simulations of the lexicon. As such, they can provide only a static approximation of what happens in the mind. Several examples of the differences between these models and human lexicons are worth noting. Firstly, research by D. L. Nelson et al. (2000) suggested that association strength, which is used in WA-derived networks to simulate connection strength in the lexicon, is merely “a manifestation of [connection] strength” (*Ibid*: 896). This manifestation is vulnerable to personal and contextual variation: “for a given individual, strength can vary from moment to moment, and for different participants, the same responses can be represented at different strengths depending on their experience” (*Ibid*.: see also K. Nelson, 1977). While this variation is to some extent captured in some WA models through the collection of multiple responses to each cue (De Deyne, Navarro, et al., 2012), these networks nevertheless remain static once data collection is completed.

Another problem with artificial lexical networks is that the data from which they are created, whether from word associations or from other sources, such as semantic relatedness judgments (Steyvers & Tenenbaum, 2005) or distributional data (Andrews et al., 2009), invariably excludes certain types of information which is likely to be available to humans. In the case of WA-derived networks, the context-free nature of the WA task means that such networks may fail to capture context-specific links in the lexicon. Barsalou (1982), for example, suggests that some connections between words may only be revealed in the presence of certain context-dependent information. This limitation may contribute to the sparse connectivity of WA-based networks (De Deyne, Navarro, et al., 2012).

Table 3.1
Methods for measuring word association responses.

Measure	How is it calculated?	Typical usages	Example of use to investigate cue-level variation
Distributional measures			
Group norm comparison	A control group responds to a set of WA cues. The most common responses produced by this group are calculated. Responses produced by another group are then compared to the control group to assess their level of similarity. Examples of control and experimental groups might include L1 (control) and L2 (experimental) respondents, or healthy (control) aphasic (experimental) participants.	Typically used to differentiate between populations with different characteristics. For example: <ol style="list-style-type: none"> 1. Psychological traits such as creativity and personality (Gough, 1976; Merten & Fischer, 1999) 2. Neurodegenerative conditions such as Alzheimer's disease (Eustache et al., 1990; Gewirth, Shindler, & Hier, 1984; Gollan, Salmon, & Paxton, 2006) 3. First and second language respondents (Kruse et al., 1987; Meara, 1978; Wolter, 2002). 	No studies to date have looked at interactions between lexical variables and this measurement.
Association strength	The number of response tokens given for a single response type. Generally presented as a ratio of total response tokens given to that cue. Can be further divided into forward (i.e. the tendency for A to elicit B; FAS) and backward (the tendency for B to elicit A) association strength (BAS).	Probably the most widely used measure of WA. It is used: <ol style="list-style-type: none"> a. As an explanatory variable in psycholinguistic studies (Hutchison, 2003; McKoon & Ratcliff, 1992, 1995; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Pecher, Zeelenberg, & Raaijmakers, 1998). b. In weighted network models, (De Deyne & Storms, 2008; Deese, 1966; Gruenenfelder, Recchia, Rubin, & Jones, 2016; Wilks & Meara, 2002), in which it manifests the strength of links between nodes (words) in the network. c. In the case of backward association strength, as an possible explanation for "associative" false memory in list learning studies (Asch & Ebenholtz, 1962; Brainerd, Yang, Reyna, Howe, & Mills, 2008; Epstein, Szymanski, & Daggett, 1977). d. In methodologically-focused investigations, as a way to investigate the reliability of different methods of WA data collection (D. L. Nelson et al., 2000). 	Brainerd et al. (2008) investigated covariation between BAS and a wide range of semantic variables. They found that many of these factors (e.g. meaningfulness, concreteness, valence, dominance, synonymy etc.) varied significantly as a function of BAS. In many cases, however, this variation did not appear to be systematic.

Measure	How is it calculated?	Typical usages	Example of use to investigate cue-level variation
Number of associates (NoA; Duñabeitia, Avilés, & Carreiras, 2008), AKA (Nelson's) Set Size (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004).	The number of response types given by more than one participant. It is calculated in discrete WA tasks, some of which allow respondents to give more than one response per cue (i.e. multiple response tasks). This corresponds to the network measure of out-degree. It is also referred to as <i>response heterogeneity</i> (the proportion of response tokens which are different types) and <i>response homogeneity</i> (the proportion of tokens which are the same type).	Used as an explanatory variable in regression studies investigating semantic processing. Numerous researchers refer to its semantic nature (Balota et al., 2004; Duñabeitia et al., 2008; Yap et al., 2012).	In a WA task which allowed up to three responses per cue, Zareva (2011) found that NoA increased along with cue frequency.
Response availability, AKA <i>m</i> (de Groot, 1989; Noble, 1952)	Similar to the above, except that it is calculated from continuous WA tasks – i.e. those in which participants are asked to give as many responses as possible in an allotted period of time.	Used to explore cue-level variation, as a measure of the connectivity of words within the lexicon.	De Groot (1989) recorded response availability in order to compare the influence of cue frequency and imageability.

Network measures; these operate on networks which use words as nodes and association strength as links between them.

Mutual association matrices	Such a matrix requires all responses to also be present as cues. The matrix depicts the frequency with which each word yields other words as responses, as well as how frequently it is given as a response to other cues.	Networks of this type are the basis of many studies. The networks aim to offer a simplified representation of the mental lexicon and its properties; methods for analysing these networks will be described below. Perhaps the most direct use of these networks is in Deese (1964, 1966) who used relatively small networks of around 300 words to identify mutually associated pairs and analyse their properties.	Deese (1964, 1966) investigated the impact of grammatical class on the structure of these associative links. He found, for example, that adjective pairs with strong mutual associations tended to be polar opposites such as <i>hot</i> and <i>cold</i> . Deese suggested that scalar relations apply to other adjectives from the same semantic set, to the effect that an adjective's meaning can be inferred from its mutual associations.
-----------------------------	--	--	--

Measure	How is it calculated?	Typical usages	Example of use to investigate cue-level variation
Network centrality measures: PageRank, in-degree, out-degree, clustering coefficient, and betweenness (De Deyne, Navarro, et al., 2012).	These measures estimate the extent to which a word is central to the lexicon by counting the number of associates a word generates as a cue (its out-degree; corresponding to NoA, above) and/or the number of times it is produced as a response (its in-degree). Some measures (PageRank, clustering coefficient) incorporate additional associative information (see De Deyne et al., 2013).	Network models aim to simulate the human mental lexicon. Centrality measures allow researchers to investigate the global properties of the network, as well as investigating words which are most important in the network. These measures can then be used to test, for example, theories of lexical attrition (Meara, 2004), or as explanatory variables in psycholinguistic tasks such as lexical decision and word naming (De Deyne, Navarro, et al., 2012).	De Deyne & Storms (2008) found that while distributional features such as frequency and age of acquisition have limited influence on the number of responses given to a cue (i.e. its out-degree), they correlate strongly with the number of incoming links with that word (i.e. its in-degree).
Cosine similarity (Landauer & Dumais, 1997)	The similarity of two words' associative neighbourhood is calculated, such that "two words that share no associations have a similarity of 0, while two words with the exact same associative responses have a similarity of 1" (Van Rensbergen et al., 2016).	By comparing the associative profiles of two words, the cosine similarity measure allows estimates of semantic similarity to be derived from WA-based networks.	Van Rensbergen et al.'s (2016) work found that extrapolations based on cosine similarity correlated strongly and significantly with human judgements of emotional properties (valence, dominance, and arousal; see below). In addition to suggesting an application for WA-based networks in estimating semantic variables, these results also suggest the importance of emotional properties in lexical structure.

Response category

Human coding	Human coders place all cue-response pairs into one or more pre-specified categories. The most common categorisation scheme is the syntagmatic (position-based; different grammatical class), paradigmatic (meaning-based; same grammatical class), clang (form-based/phonological) scheme (Saporta, 1955); many other schemes exist.	Different response categories are typically thought to represent different types of lexical knowledge or processing (Meara, 2009; K. Nelson, 1977). As such, researchers using a response categorisation method typically aim to investigate differences in word knowledge or processing across different populations. These include first and second language learners (Namei, 2004; Wolter, 2001) and different age groups (Entwisle, 1966b; Ervin, 1961).	Deese (1962) investigated the impact of grammatical class on the proportion of syntagmatic to paradigmatic responses. He found that nouns elicit syntagmatic responses in only around 21% of responses; verbs and adjectives in around 50% of responses, and adverbs around 73%.
--------------	--	--	--

Measure	How is it calculated?	Typical usages	Example of use to investigate cue-level variation
Corpus-derived categorisation	Rather than rely on human intuition regarding categorization, Zareva & Wolter (2012) attempted to determine whether word pairs could be categorised as collocations using a corpus-derived t-score threshold of 2.	Zareva and Wolter’s method was largely methodological, aiming to assess the “promise” of this corpus-derived measure. The method ultimately aimed to provide an estimate of the possible influence of co-occurrence-based lexical knowledge on the WA task.	No interactions of t-score with cue-level features were investigated in Zareva and Wolter’s study, and this line of research is yet to be adopted in other studies.
Other measures			
Response time	Measures the time taken for respondents to produce a WA response.	Several studies (de Groot, 1989; Fitzpatrick & Izura, 2011; Izura & Playfoot, 2012; Playfoot & Izura, 2013; Santos et al., 2011; van Hell & De Groot, 1998) have argued that response time may reveal differences in lexical organisation or processing. For example, Santos et. al. (2011) suggested that faster responses originate from a lexical system, while slower ones are processed by a deeper conceptual system.	In a study in which WA cues were all familiar acronyms (e.g. BBC, NATO etc.), Playfoot and Izura (2013) found that cue imageability and age of acquisition significantly speeded responses, while acronym frequency had no effect on response time.
Modelling WA using computerised data	Not strictly a measure of WA but rather a method of enquiry into its properties; some researchers (Bel Enguix, Rapp, & Zock, 2014b, 2014a; Mollin, 2009; Wettler, Rapp, & Sedlmeier, 2005) have attempted to compare WA norms lists to data or models derived from corpora.	Studies of this type aim to assess the potential influence of knowledge of lexical contiguity on WA responses. They seek to do this by constructing models or structured comparisons of WA using lists of collocations derived from corpora.	Bel Enguix et. al. (2014a, 2014b) suggest that models built from collocation data can produce WA responses with similar characteristics to human respondents. This suggests that a cue’s collocational profile should be considered a potential influence on WA responses.

Despite these issues, network models provide a useful and theoretically well-established framework upon which to base an exploration of cue-level influence on the word association task. Several researchers have suggested ways in which the properties of words influence their connectivity and network centrality. For example, De Groot (1989: 837) and De Deyne et al. (2013), in studies which will be described below, suggest that variables such as word frequency and age of acquisition shape network centrality, such that highly frequent or early acquired words are more central than less frequent or later acquired ones. This is because, according to Steyvers & Tenenbaum (2005) more frequent and early acquired words provide an anchor onto which later acquired words attach, thus increasing the number of links they possess. If true, this should be evident in measures related to connectivity, such as response availability, number of associates (NoA), and in- and out-degree. Semantic variables, on the other hand, have been hypothesised to influence connection strength. For example, Van Hell and De Groot (1998) suggested that connection strength is partially determined by the extent of semantic overlap between a word and its semantic neighbours, and that this overlap is influenced by variables such as concreteness and imageability (see Section 3.4.3, below).

3.3.1 Distributional and semantic access in word association

The above studies suggest that distributional features such as frequency, contextual diversity (Adelman, Brown, & Quesada, 2006), and age-of-acquisition influence network properties in different ways to semantic features such as concreteness and emotionality. A study which supports this hypothesis is that of Van Rensbergen et. al. (2015), who used a large WA-derived network made up of 4151 cue-response pair types, covering 665,461 cue-response tokens, to investigate network *assortativity* (i.e. the extent to which the features of cues predict those of their responses). It is important to note that the study was based on response tokens; this means that the number of times each word pair was entered into the analysis was determined by the number of times it was produced by participants. The results should therefore be taken to reference association strength. For each word in the network, measures of frequency, contextual diversity, age of acquisition (all distributional variables), valence, dominance, arousal, and concreteness (all semantic) were collected from existing databases. Van Rensbergen et. al. used a multiple regression model to explore co-variation in these variables. They found that distributional aspects of cues were poor predictors of the same distributional characteristics of the response. For example, cue frequency was able to explain only 1% of variance in response frequency. On the other hand, the semantic variables were each able to explain at least 15% of variation in the same property of their responses. The largest proportion of response variation explained was for cue valence (a subjective

measure of the emotional positivity or negativity of a concept), which explained 31% of variance in response valency. The authors interpreted this finding as suggesting the importance of emotional factors in determining the strength of connections in the lexicon (although other interpretations are possible: see Section 3.5.5, below).

There is some precedent for the differential effects of distributional and semantic factors in other areas of psycholinguistics. For example, Yap, Tan, Pexman, & Hargreaves (2011) compared the influence of lexical and distributional features (e.g. frequency, number of letters) with those of semantic ones (e.g. number of features, number of senses) on response times and error rates in three tasks – lexical decision, word pronunciation, and semantic classification. Using a hierarchical regression procedure, they found, firstly, that some variation in these measures for all three tasks was explained by cue frequency. The amount of variation explained was largest for lexical decision, followed by word pronunciation, and then semantic classification. The authors then added the semantic variables to the model. This time, the largest amount of variance explained was for the semantic classification task. Reflecting on results such as these, Taikh, Hargreaves, Yap, & Pexman (2015: 1515) suggest that “[word] naming and [semantic] classification tasks place different demands on lexical and semantic processes. Specifically, naming places greater emphasis on lexical variables such as word frequency, whereas classification places greater emphasis on semantic variables”.

As noted in Chapter 1, the psycholinguistic view of WA in recent years has been that it is largely a semantic task (Brainerd et al., 2008; De Deyne & Storms, 2015; Grossman & Eagle, 1970; Mollin, 2009; van Hell & De Groot, 1998), more similar in nature to semantic classification than to lexical decision or word naming. It is worth noting, however, that WA is considerably more complex than these tasks, in that it is likely to involve at least four processes – access to a word’s form, its semantics (though it is possible that this may not be accessed, particularly when form-based responses are given), response selection, and response articulation. Variables influencing each of these processes may also influence WA; alternatively, they may be lost in the longer time-course of associative response production. Whatever the reality, however, it seems that a principled approach to word association, which prioritises multiple measures of production, is particularly well-placed to investigate the nature of these effects, because of its ability to capture different aspects of cue-level influence via its numerous dependent measures.

3.4 Lexical variables

The above discussion has suggested that cue-level variables influence the organisation of the mental lexicon. The lexicon can be simulated using word association norms, and this simulated model can be used to explore cue-level influence. The influence of these variables on the network properties of connectivity, connection strength, and centrality are revealed through a range of dependent variables used to collect data from word association tasks. It has been suggested that a gross distinction between distributional and semantic variables may exist, and that these may influence the lexicon in distinct ways. The discussion will now turn to the lexical variables themselves, before returning to this interaction between variable, measurement, and lexicon in Section 3.5.

3.4.1 Grammatical class

Much of what is currently known about grammatical class influence on first language word association responses originates from two studies from the 1960's. These are by Deese (1962a, 1966), and Entwisle (1966a). Both studies used the response category measurement, and set out a relatively consistent pattern of findings for the lexical classes (nouns, verbs, and adjectives), based on relatively large-scale experiments involving undergraduate English L1 participants. In the Deese study, 100 participants provided responses to a total of 600 cues (253 nouns, 118 adjectives, 101 verbs, and 32 adverbs. For the Entwisle study, 200 adult participants responded to 24 nouns, 24 adjectives, 24 verbs, eight adverbs, eight pronouns, and eight miscellaneous words (prepositions, conjunctions; no details were provided on responses to these miscellaneous cues), for a total of 96 cues. Results from these studies are set out in Table 3.2.

The two studies agree to a considerable extent with regard to the proportion of syntagmatic to non-syntagmatic (i.e. paradigmatic and clang) responses given for noun cues. Both studies also showed consistently higher proportions of syntagmatic to non-syntagmatic responses for verb and adjective cues. The results for adverbs are, however, remarkably different. The data across the two studies were insufficient to compare the remaining classes, although Deese's (1966: 111-115) discussion suggests a generally paradigmatic trend for all function words.

The results of a meta-analysis by Cramer (1968) are in general agreement with the pattern of responding to the lexical classes. She found that nouns receive the highest proportion of paradigmatic responses, followed by pronouns, then adjectives, adverbs, and finally verbs. For syntagmatic responses, the order is adverbs, adjectives, verbs, and then nouns. This basic pattern, at least with regard to noun, verb, and

adjective cues, was also reported by De Deyne & Storms (2008), and for nouns and verbs by Bøyum (2016); although Van Rensbergen, Storms, & De Deyne (2015) reported a very high proportion (around 75%) of syntagmatic responses to verbs in their study.

Table 3.2
Percentage of syntagmatic responses for various grammatical classes in two large-scale studies.

	Nouns	Verbs	Adjectives	Adverbs	Pronouns	Prepositions
Deese, 1962a, 1966	21.4	48.1	49.9	72.8	N.D.	41.7
Entwisle (1966a)	22.9	40	34.2	21.1	22	N.D.
Nissen & Henriksen (L1 only)	43.5	58.8	59.7	N.D.	N.D.	N.D.

N.D. = No Data

Since the publication of these studies, few experiments have sought to probe grammatical class influence any more deeply. One exception is a bilingual study by Nissen and Henriksen (2006). The L1 (Danish) response patterns from this study are also reported in Table 3.2. Two aspects of this study’s findings stand out. First, the level of syntagmatic responding was, for all three lexical classes, markedly higher than that reported in earlier studies. Nissen and Henriksen suggest two cue-related explanations for this finding. First, they suggested that the high frequency of the cues (selected from the most frequent 2000-3000 words in English, corresponding to a frequency of 35-55 words per million: fpmw; Leech & Rayson, 2014) used in the test may have led to increased syntagmatic responding. They speculate that higher frequency words may have more general meanings than lower frequency words. These general words may more easily elicit specific, lower frequency paradigmatic (synonym or subordinate) responses than vice versa. Given this difficulty in identifying paradigmatic responses, the lower frequency cues used in Nissen and Henriksen’s study may have been more likely to elicit syntagmatic responses than those in studies with more frequent cues. However, as we will see in Section 3.3.2, the evidence on cue frequency influence on WA does not appear to support the conclusion that higher frequency will lead to increased syntagmatic responding. This is particularly true in the case of adjective cues, which generally yield more paradigmatic responses as frequency increases (Deese, 1962a, 1966; Entwisle, 1966a).

The second interpretation suggested by Nissen and Henriksen (2006:400-1) was that uncontrolled variables may have contributed to their findings on grammatical class influence. They suggested that factors such as frequency, concreteness, and imageability may have influenced their results (Nissen and Henriksen, 2006: 392). Chapter 4 will look at the potential impact of these factors in more detail. For the present, it is important to note that in drawing attention to uncontrolled variables in their own study,

Nissen and Henriksen revealed the same weakness in much existing WA research. Few word association studies have sought to limit the influence of potentially confounding variables. This is partly due to the continued use of pre-selected cue lists such as the Kent-Rosanoff list, which is largely made up of concrete, high frequency nouns (e.g. Namei, 2004; Söderman, 1993, Experiment 1; Sökmen, 1993; see also D. L. Nelson et al., 2000; Zareva, 2007), but studies using specially selected cues have also tended to control only for grammatical class and frequency (Deese, 1962a; Entwisle, 1966a; Ervin, 1961; Zareva & Wolter, 2012). Those studies which have controlled for additional factors have generally done so in order to investigate the influence of those variables specifically (i.e. they are independent, rather than control, variables; de Groot, 1989; Van Hell & De Groot, 1998; but see Fitzpatrick & Izura, 2011, for an exception). This creates challenges for the interpretation of findings such as those of Nissen and Henriksen because it makes the individual contribution of both controlled and uncontrolled variables is difficult to assess.

The second noteworthy aspect of Nissen and Henriksen's L1 findings is that, although syntagmatic responding is high overall, they follow the general trend for noun cues to receive considerably fewer syntagmatic responses than did either verb or adjective cues, as well as the trend for the number of syntagmatic responses to verbs and adjectives to be similar to one another. Nissen and Henriksen's discussion of this finding raises numerous rarely-discussed questions about the nature of word association. They suggest (cf. Entwisle 1966: 72, 120) that "words do not possess equal potential for yielding similar proportions of syntagmatic and paradigmatic responses in word association tests" (Nissen & Henriksen, 2006: 403). Several aspects of the nature of verbs and adjectives in the mind means that these cues are more likely to yield syntagmatic responses:

"(a) acquisition, (b) the manner and degree of integration into the word web, (c) the way they establish vertical and horizontal semantic relations to other words, (d) the degree of interaction with words on the sentence level, and (e) the degree of cognitive processing in a productive word association test" (*ibid.*).

Here, Nissen and Henriksen articulate a level of complexity to the WA task which goes beyond simple lexical vs. semantic distinctions. While semantic factors (i.e. point c) are assumed here to be influential, Nissen and Henriksen also assume an impact of several aspects of lexical storage (a, b), the cognitive demands of the WA task itself (e), and knowledge of a word's patterns of co-occurrence (d). In this final regard, they echo Deese (1962: 79-80):

Nouns are [...] largely independent of their verbal environments in association. They should yield associations which are mostly other nouns [i.e. paradigmatic]. Verbs and adverbs relate terms in discourse or modify relations. They should, therefore, be heavily dependent in associative meaning upon their environments. It follows that their associates should be largely from different form classes [i.e. syntagmatic]. [...] There is evidence to suggest that very common adjectives are paradigmatic in associative meaning; therefore, these should yield largely paradigmatic associates. Uncommon adjectives, however, depend upon the environments which they modify, and they should be mostly syntagmatic.

Here, both Deese and Nissen & Henriksen imply that the sentential roles played by different grammatical classes is a potential determinant of word association responses. This is an important point, not only because it helps to explain why cues of different grammatical classes elicit different proportions of syntagmatic and paradigmatic responses, but also because of the way it problematizes the apparently semantic basis of word association (Brainerd et al., 2008; De Deyne & Storms, 2015; Grossman & Eagle, 1970; Mollin, 2009; van Hell & De Groot, 1998) by suggesting that knowledge of co-occurrence and contiguity might be a significant source of associative knowledge. Indeed, research by Kang (2018) has suggested that the influence of knowledge of word contiguities on associative knowledge may vary as a function of grammatical class, since correlations between corpus and WA data differ across grammatical classes. This issue will be returned to in Chapter 7.

The above discussion has revealed a general pattern for response categories for each of the three lexical classes: nouns tend to yield fewer syntagmatic responses than both adjectives and verbs (see Section 3.4.2, below, for an amendment to this pattern for adjectives). Beyond these basic findings, however, the above review suggests that numerous problems remain:

- Proportions of syntagmatic responses to verb cues vary markedly between studies
- Little is known about the influence of grammatical class on WA measures such as response time or distribution
- Not enough has been done to remove the possibility of confounding effects of variables such as concreteness or imageability
- Little is known about response patterns to functional classes such as pronouns
- The psycholinguistic locus of grammatical class effects in word association remains largely unexplored.

3.4.2 Frequency

Word frequency refers to how often a word appears in a specific corpus of texts (Brysbaert & New, 2009; Kucera & Francis, 1967; Thorndike & Lorge, 1944). An influence of this variable on word association has long been assumed, for two main reasons. Firstly, frequency has a significant facilitative impact in several psycholinguistic tasks: more frequent cues are responded to more quickly. This is the case not only for predominantly lexical tasks such as lexical decision and word naming (Balota et al., 2007; Brysbaert & New, 2009; Keuleers et al., 2012), but also semantic ones such as semantic classification (Yap et al., 2011). One explanation of these effects is that more frequent (or earlier acquired: see Section 3.4.4) words have more connections to other words in the lexicon (de Groot, 1989; Steyvers & Tenenbaum, 2005). This suggests that WA response times should also be faster for more frequent cues, and that these cues will also show greater connectivity, as measured by response availability (the mean number of responses given to a cue by participants in continuous WATs) and number of associates (NoA; the same measurement, but taken from discrete tasks rather than continuous ones).

Secondly, with regard to response distribution, some researchers have noted that high frequency words can tend to elicit very predictable responses. Meara (1983), for instance, suggested high frequency adjectives (*black-white, soft-hard*) and gender-marked nouns (*king-queen, boy-girl*) as fitting this profile. Following this trend, Fitzpatrick (2007: 324) justified her selection of low frequency cues by claiming that "there is some indication that high-frequency words produce more predictable responses". The assumption here, then, is that frequency will influence association strength, to the effect that primary associations to some very frequent cues are so strong that they dominate the distribution of responses to that cue. Network models, however, do not necessarily support this prediction; as discussed in Section 3.3, the factors underlying connection strength are not well understood. This property is assumed to be influenced by personal and contextual factors (D. L. Nelson et al., 2000), and one such factor might be frequency, in that frequency provides an estimate of the number of lifetime exposures to a word. However, it is also influenced by semantic factors such as similarity (Mollin, 2009) and emotional ones such as valence (Van Rensbergen et al., 2015). Thus, Meara and Fitzpatrick's predictions hinges upon the assumption that frequency has a greater impact on association strength than do semantic and affective factors.

The available evidence on the influence of word frequency on WA in fact suggests a complex set of effects. Looking firstly at the pattern of results for response distribution, converging findings are reported in an

experimental study by Postman (1970) and a review of early research by Cramer (1968). In the former, 96 two-syllable nouns, representing four frequency bands, were given as cues to 1000 English L1 participants. Postman found that increasing word frequency resulted in a decrease in the number of response types (i.e. reduced response heterogeneity, as observed by Meara, 1983), and an increase in number of response tokens for the most common response type. The pattern revealed in Cramer's research was similar; with increasing frequency, response heterogeneity decreased, and the association strength of the most common response rose. In addition, Cramer reported that response availability increased along with cue frequency.

The pattern emerging from these early results was, however, called into question by a meta-analysis conducted by Brown (1971). This study assessed the impact of four variables (frequency, emotionality, pleasantness, and concreteness; see the corresponding sections below for findings regarding the last three variables) on 12 WA studies of varying sizes. From a total of 7 studies which measured response heterogeneity, only 3 revealed significant correlations with frequency; two of these were positive (the opposite of the pattern observed by Postman, 1970, and Cramer, 1968), and one negative. Brown thus suggested that the impact of frequency on response heterogeneity was negligible.

In an attempt to resolve this uneven pattern of results, De Groot (1989) conducted a series of Dutch language experiments in which she investigated the influence of cue frequency and imageability across both discrete (i.e. one response per cue) and continuous (i.e. as many responses as possible in a one-minute period) tasks. She made numerous measures from the discrete tasks, including association strengths, number of associates (NoA), and reaction time (RT). For the continuous tasks, the only dependent measure was the mean number of response types given to each cue (i.e. response availability; following Noble, 1952, De Groot calls this *m*). The results pertaining to RT will be described later.

In Experiment 1, 100 Dutch L1 participants were asked to provide a single response to each of 400 cues. Of these words, 240 were fillers³ (equally divided between nouns, verbs, and adjectives); the remaining 160 of the cues, all nouns, comprised the critical trials. These 160 nouns were classified as either high (i.e. rating > 3.5 on a 7-point scale) or low (< 3.5) imageability (see Section 3.3.3 for discussion of this variable), and either high (≥ 40 occurrences in a 620,000-word corpus) or low (≤ 20 occurrences) frequency.

³ These words were used in a separate WA norming study, and were not systematically varied according to frequency and imageability as the 160 critical cues were.

De Groot found that the only statistically reliable effect of cue frequency was on NoA: a larger number of responses types were given to high frequency (HF) cues (23.9 types per word) than to low frequency (LF) ones (20.7). In other words, as frequency increased, so did response heterogeneity. This pattern mirrors the positive correlations reported by Brown (1971), and is the opposite to that observed by Postman (1970) and Cramer (1968). There was no significant effect of cue frequency on any other distributional measure, including *m* scores.

De Groot (1989: 832) suggested that the absence of clear frequency effects could have been caused by the frequency range of the HF and LF sets being too low. This hypothesis was supported in her experiments 4, 5, and 6, in which lexical decision and word pronunciation tasks revealed only small differences in RTs between the HF and LF words (lexical decision: HF RT 42ms shorter than LF ($p < .001$); word pronunciation: HF RT 5ms shorter than LF, $p < .10$). De Groot hypothesised that such small differences in RTs would be lost in the much more time-consuming process of producing a WA response, and therefore created a new list of 100 noun cues in two frequency groups separated by larger differences in frequency. The HF cues had frequencies of between 70-250 occurrences in a corpus of 620,000 words (compared with ≥ 40 in the initial set); for the LF group this number was 0-9 occurrences (compared with ≥ 20). The new cues were matched pairwise on imageability to eliminate any confounding influence. This new list was again compared using a word pronunciation task (Experiment 7). A small but significant advantage for more frequent words was found (19ms; $p < .01$), suggesting that the new cues were more sensitive to frequency effects than the earlier ones.

Using this new set of cue words, De Groot then repeated the earlier discrete and continuous word association tests (experiments 7 and 8). This time, a frequency effect on response availability emerged: HF words now also yielded slightly larger *m* scores (0.5 more responses per cue) than LF words. However, the new cues revealed no effect of frequency on other aspects of response time or distribution. Thus, De Groot's findings did not strongly support any previously existing pattern of influence from frequency.

An interesting development of De Groot's findings is made in a study by De Deyne & Storms (2008). Their study made use of a directed lexical network in which all responses were also used as cues. The network was built from a multiple-response test (limited to three responses per cue) using 1424 Dutch words. This type of network allows for out-going and in-coming links to be viewed as distinct aspects of a cue's connectivity. The number of out-going links is determined by the number of response types which Cue A yields as responses. This is essentially the same measure as NoA, but is referred to in network models

using the term *out-degree*. For example, if Cue A yields 30 response types in a WAT, its out-degree would be 30. In-coming links are determined by counting the number of other cues which yield Cue A as a response. This is termed *in-degree*. This measure is rarely calculated in non-network investigations of WA. An example of the calculation of in-degree is Cue A being given as a response to 43 other cues (Cues X, Y, Z etc.). Thus, the in-degree of Cue A would be 43. This distinction gives a more nuanced view of the way in which a cue is integrated into the lexicon. As De Deyne & Storms (2008) showed, it also reveals different influences of variables to those already described. In the case of frequency, the authors found that a word's out-degree is only marginally influenced by its frequency ($r=-.14$). The negative correlation indicates that higher frequency cues yielded fewer response types in their study. This corresponds with the pattern observed by Cramer (1968) and Postman (1970), rather than those of De Groot (1989) and Zareva (2011); but the small effect size may resonate with the lack of agreement in these studies. On the other hand, there was a far stronger correlation ($r=.70$) between frequency and in-degree: frequent words were more often produced *as responses* than infrequent ones. De Deyne & Storms suggest that in-degree provides a better measure of network centrality than does out-degree because it references the frequency with which a given word is activated by other words. This conclusion is supported by the fact that in-degree explains more variance in lexical decision response times than other measures of centrality (De Deyne, Navarro, et al., 2012). De Deyne & Storms (2008) also noted significant positive correlations of frequency with other measures of centrality. Drawing on work by Steyvers & Tanenbaum (2005), they suggest that frequent words are encountered earlier and more often, which leads to other words building upon them as they enter the lexicon. This idea will be returned to in the discussion of age-of-acquisition effects, below.

In summary, the only consistent influence of cue frequency on WA response distribution appears to be the finding that response availability (i.e. m : the mean number of responses produced for a given cue in continuous tasks) increases along with frequency (Cramer, 1968; de Groot, 1989); this also appears to be the case on discrete multiple-response WA tasks (which requires a specific number of responses to each cue, greater than 1; Zareva, 2011). However, the large and significant correlation between frequency and in-degree revealed by De Deyne & Storms (2008), though not yet independently replicated, suggests that the strongest influence of frequency on connectivity is one which is invisible to studies looking at outgoing links from cues. Beyond these findings, there was, as discussed above, no clear support for Fitzpatrick (2007) and Meara's (1983) suggestion that cue frequency is a determinant of association strength and response homogeneity.

De Groot's (1989) study also looked at response time data. The study found no significant difference in RT to high (HF) and low frequency (LF) cues using the initial, narrowly distributed cue set. However, when new, more widely distributed cues were used, a small effect emerged: HF words (1,531ms) were responded to slightly more slowly than LF words (1,459ms). This effect was reliable only on a by-subjects analysis, however. The direction of the effect, with HF cues responded to more slowly, was the opposite of the effect observed in De Groot's word pronunciation task, suggesting that any initial processing advantage for these cues is cancelled out by whatever processes drive the later stages of WA response production. This finding is supported by Brysbaert, Wijnendaele, & De Deyne (2000), who also report longer response times for more frequent words in a spoken Dutch-language WAT involving 20 university student participants. However, this pattern does not agree with that reported by Cramer (1968), who suggested that response times are faster given higher frequency cues.

Brown's (1971) meta-analysis looked at eight studies involving RT measurements, and again reported an inconsistent pattern of influence from frequency. From four studies which reported significant correlations between these variables, three were negative (i.e. higher frequency correlated with faster responses – the pattern observed by Cramer) and one was positive (i.e. the same as De Groot, 1989, and Brysbaert, 2000). As such, no consistent pattern of influence from frequency on response times can be asserted. De Groot (1989) interprets this negligible influence of frequency as suggesting, firstly, that frequency does not impact on connection strength between words in the lexicon, and secondly that response times in word association are more strongly influenced by connection strength than by connectivity – thus the lack of influence of frequency on RT. The former part of this explanation is supported by the absence of any consistent pattern of frequency influence on association strength in the studies reviewed above, as well as the findings of Van Rensbergen et. al., who found no influence of frequency on assortativity (see Section 3.3.1).

Several studies have also looked at the impact of cue frequency on response category. Among the earliest of these was Deese (1962a, 1964), who investigated interactions between frequency and grammatical class on syntagmatic and paradigmatic responses, using a group of 100 male English-L1 undergraduate participants. Cues were selected from a range of frequency bands extending from 1-100 fpmw. These words were taken from the Thorndike and Lorge word list (Thorndike & Lorge, 1944).

Deese's results showed no effect of frequency on the proportion of syntagmatic responses for noun, verb, or adverb cues. Only in the case of adjective cues was a significant effect found: high frequency adjectives

were less likely to yield syntagmatic responses than low frequency ones. Looking more closely at the nature of these responses, Deese (1962: 82) found that the paradigmatic responses given to the high frequency adjectives were largely contrasting pairs (e.g. bad-good, big-little, fat-thin: note that some of the cues highlighted by Meara as having highly constrained response patterns also fit this category). Frequency was strongly and significantly correlated with the provision of this type of response ($r=.889$; $p<.01$). Deese suggested (1964) that this finding could be explained by reference to the associative structure of adjectives. More frequent adjectives form contrast-based relationships arranged around polar opposites (*hot-cold*), with other words with related meanings (*warm, freezing*) clustering around these basic distinctions. Responses to less frequent adjectives do not cluster in this way, either because they do not have polar opposites, or because their low frequency means that they are not encountered often enough to enable the formation of these contrast-based clusters in the lexicon. Instead, infrequent adjectives tend to become associated to words with which they co-occur.

Deese's general pattern, then, was for frequency to influence responses to adjectives, but not to cues of other grammatical classes. Three later studies found support for these findings, but called into question his interpretation of them. Entwisle (1966a), in a study looking at changes in response patterns as a function of participant age, also found that syntagmatic responding increased among adult participants as adjective frequency declined. However, the interaction between frequency and other grammatical classes "[did] not seem to [follow] any consistent pattern", and "the amount of variance accounted for [was] small compared to that accounted for by other variables [i.e. grammatical class and participant age]" (*Ibid.*; 564). Entwisle further suggested that the influence of frequency may have been confounded among her participants with cue familiarity; her younger participants were likely to have been unfamiliar with the less frequent cues, while older participants were much more likely to be familiar with them. Thus, the locus of any frequency effects within the responses of different groups may not have been the same.

The same influence of frequency on adjective cues was also uncovered by Söderman (1993, Experiment 2) in both L1 and L2 participants, who responded to 60 adjective cues. Söderman found that frequent words, defined as having a frequency above 50fpmw, received 62.7% (L1 respondents) and 52.6% (L2 respondents) paradigmatic responses, while infrequent (defined as lower than 10fpmw) cues received 44.3% and 30.3% paradigmatic responses respectively. Following Entwisle, Söderman suggested that familiarity with the cues might have been responsible for this effect.

Stolz & Tiffany (1972), looking solely at adjective cues of various frequencies, probed the impact of cue familiarity independently to that of frequency. They found that while frequency per se did not influence the extent of syntagmatic responding to adjective cues, participants' familiarity with the cues did: less familiar cues were more likely to receive "child-like" (i.e. syntactic⁴ and form-based) responses than were more familiar ones. The authors' interpretation contrasted with that of Deese: they suggested that frequency itself has no effect on WAT response categories, and that previously reported effects are better explained by familiarity. However, their manner of categorising responses collapses syntagmatic and form-based responses into a single group, making it difficult to compare this study with earlier work.

Wolter (2001) provided support for Stolz & Tiffany's interpretation using cues matched for grammatical class, frequency, and familiarity. He found a pattern of higher form-based responding to both low-frequency and low-familiarity cues, but no significant effect of frequency on the proportion of paradigmatic and syntagmatic responses. This is similar to the pattern reported by Stolz and Tiffany, but is based only on form-based responses, rather than the syntagmatic distinction reported by Deese. Wolter's findings also referred to verb and noun cues as well as adjectives. These distinctions suggest that the effects he reported may not be the same as those of Deese.

In a study into the influence of cue grammatical class and frequency on responses, Zareva (2011) found that while the combined effect of these variables did significantly affect the proportion of syntagmatic responses, neither independently influenced this measure, nor the proportion of paradigmatic responses. This may in part be due to a combination of the complexity of the study, which involved a 3 (high, mid, and low frequency) by 3 (noun, verb, and adjective) by 3 (L1, L2 Advanced, and L2 Intermediate) research design, combined with the small number of cues ($n=36$) used to represent the two cue-level variables.

Contrasting findings were reported by Guida & Lenci (2007), who found in a study of 312 Italian verbs that cue frequency was negatively correlated with noun responses (i.e. syntagmatic responding; $r=-.28$). This result is similar to that reported in a German language study by Schulte Im Walde & Melinger (Schulte im Walde & Melinger, 2005), which also made use of several hundred cues. Guida & Lenci's interpretation of these findings did not draw on the concept of cue familiarity. Instead, they suggested that lower frequency words are more lexically specific. This means that respondents associate them with more specific meanings and contexts. As a result, lower frequency verb cues are more likely to yield responses which refer to the specific entities involved in the event that the verb describes (*Ibid.*: p. 305). These entities,

⁴ Stolz and Tiffany appear to prefer this term to the more common "syntagmatic".

Guida & Lenci argue, will largely be nouns, as in the example of *cycle-bike*. Further research is needed, however, to demonstrate the relationship between lexical specificity and response GC, since Guida & Lenci's study did not investigate this directly, but rather used cue frequency as a proxy for lexical specificity.

In summary, two patterns of influence of cue frequency on response category have been observed; it remains unclear whether they are separate phenomena. Firstly, there is evidence that low frequency cues result in a higher proportion of form-based responses than higher frequency ones; this may be better understood as a word familiarity effect. Secondly, the proportion of paradigmatic and syntagmatic responses may vary as a function of cue frequency: less frequent cues receive more syntagmatic responses, and more frequent ones receive more paradigmatic responses. More research is needed to ascertain the locus of this effect: differing associative structures for infrequent words, word familiarity influence, and the impact of lexical specificity have all been put forward as explanations of this phenomenon. Finally, as with response distribution, it is notable that several researchers have commented on the greater influence of grammatical class than frequency on response categorization (Deese, 1962a; Entwisle, 1966a; Zareva, 2011).

The only consistent effects of cue frequency which emerge from the above discussion are:

- Frequency strongly correlates with a word's in-degree in WA-based networks. In other words, frequent words are more likely than less frequent ones to be given as associative responses in WATs (De Deyne & Storms, 2008).
- More frequent cues yield a higher number of response types in continuous or multiple-response WATs (Cramer, 1968; de Groot, 1989; Zareva, 2011).
- Less frequent cues tend to elicit a higher proportion of form-based responses; this might be better understood as a cue familiarity effect than a pure frequency effect (Stolz & Tiffany, 1972; Wolter, 2001).
- Frequent adjective cues are more likely to yield paradigmatic responses than infrequent adjective cues (Deese, 1962a; Entwisle, 1966a). This effect has been interpreted as stemming from the semantic structure of adjectives (Deese, 1964), and as a familiarity effect (Stolz & Tiffany, 1972).

Several weak or inconsistent patterns have been reported with regard to cue frequency:

- The impact of cue frequency on response times is inconsistent: some evidence points to slower responses for more frequent cues (Brysbaert et al., 2000; de Groot, 1989), but other studies show the opposite pattern (W. P. Brown, 1971; Cramer, 1968).
- The impact of frequency on response distribution is highly inconsistent.
- Network models suggest that the frequency of a cue is a poor determinant of the frequency of its responses (Van Rensbergen et al., 2015).

This pattern of results supports the view that frequency impacts upon the connectivity of words within the lexicon, but only when in-coming links to words are considered. There is a much less significant impact on connection strength. It also suggests that weaker familiarity with a cue increases the chances that its responses will be based upon its phonological or orthographic properties.

3.4.3 Age of acquisition

Many early accounts of age of acquisition (AoA) effects on tasks such as lexical decision suggested that the measure had no effect independent of word frequency. This led to questions about the its psycholinguistic validity, particularly given the subjective way in which AoA norms are collected, which involves asking adult respondents to estimate the age at which they first encountered a given word (Stadthagen-Gonzalez & Davis, 2006). However, several validation studies have supported these subjective measures by comparing them with more objective methods (Carroll & White, 1973; Gilhooly & Gilhooly, 1980; Jorm, 1991). Nevertheless, the range of tasks on which AoA explains RT and error rate variation independently of those of frequency are limited (Brysbaert & Ghyselinck, 2006). This has led some researchers to suggest that AoA effects, which saw early acquired words being responded to more quickly than late learned ones, were better described as cumulative frequency effects (Lewis, 1999b, 1999a; Lewis, Gerhand, & Ellis, 2001). This cumulative frequency hypothesis posits that AoA effects can be more parsimoniously interpreted as reflecting the number of encounters a person has with a word over the lifespan, rather than the result of the age of the first encounter itself.

Word association studies have played a role in several theories which have attempted to argue for the independence of AoA effects from those of cue frequency. Among the first such studies was Van Loon-Vervoorn (1989), who found that word association responses were produced more quickly for words rated as early acquired than later acquired ones; this effect was shown to be independent of cue frequency, and this independent from cumulative frequency. Van Loon-Vervoorn suggested that the effect occurred because early acquired words, by virtue of their early acquisition, attained a central place within the

lexicon; later acquired words are initially parasitic upon these central words. Early acquired words are therefore highly interconnected within the lexicon.

De Deyne & Storms (2008) found support for this notion in their study of factors contributing to network centrality. They found that while AoA did not affect the number of out-going links a word possessed (i.e. its NoA), there was a strong negative correlation between AoA and in-degree, such that early acquired words were more likely than late acquired words to be given as responses to other cues. Other researchers have commented on this centrality of early acquired words. Brysbaert & Ghyselinck (2006: 1001), for example, suggest that such words are frequently exemplars of common categories, such as banana, apple, and orange for fruit. Several other measures of centrality reported by the authors corresponded to this finding. For example, early acquired words were more frequently found to represent the fastest path between two other concepts than were late acquired words (i.e. this is ascertained using the *betweenness* measure of network centrality). However, in De Deyne and Storms' study, these effects were consistently similar to those for cue frequency (see above: the correlations for AoA were, in fact, slightly weaker than those for frequency); the study did not seek to establish the independence of these effects. As such, while De Deyne and Storms' study does suggest that early acquired words are more central to the lexicon than late acquired ones, it cannot be portrayed as suggesting that these effects are independent of frequency effects.

One study which has suggested this independence is that of Brysbaert et al. (2000), who suggest that there are two loci for AoA effects – one which is dependent on frequency, and another which is not. Their study used three WA cue lists. The first was matched for frequency and imageability, but differed in AoA; the second was matched for AoA and imageability but varied in frequency; and the last varied in imageability but was matched for frequency and AoA. Prior to the WA experiment, the authors demonstrated that their cue lists produced similar facilitative effects of AoA and frequency on word naming and lexical decision tasks. In the WA task, however, the effects of these two variables markedly diverged: early acquired words (matched for frequency) were responded to more quickly than late acquired ones, but more frequent words (matched for AoA) were responded to more slowly than less frequent ones (this pattern of results was briefly described in Section 3.4.2). These findings essentially repeat those of Van Loon-Vervoorn (1989). Brysbaert et. al. suggest that the results point to a frequency-independent AoA effect with its locus at the intersection of semantic and phonological processing. Later research has supported this multiple-loci account of AoA effects (Catling & Johnston, 2009).

For the purposes of the current review, the specific locus of AoA effects is less important than the consistent trend for AoA to facilitate WA response times, independently of the effects of frequency. This effect may be due to the increased centrality of early acquired words, compared to late acquired ones, which was demonstrated by De Deyne & Storms (2008). Unfortunately, no studies to date have further explored the impact of AoA on other measures of WA, such as response type.

3.4.4 Concreteness and imageability

Concreteness refers to a word's location on a scale which ranges from, at one extreme, physical referents which can be directly perceived by the senses (high concreteness; e.g. desk, peacock), to, at the other, abstract concepts which cannot be directly perceived (low concreteness; e.g. belief, absurdity; Brysbaert, Warriner, & Kuperman, 2014; Paivio, Yuille, & Madigan, 1968). Substantial evidence, including from brain imaging studies, investigations of patients with semantic dementia, and behavioural studies with healthy participants, points to concreteness as shaping semantic processing (Barber, Otten, Kousta, & Vigliocco, 2013; Bonner et al., 2009; Cousins, York, Bauer, & Grossman, 2016; Jessen et al., 2000). In addition, it is also notable that some evidence exists of a facilitative effect of concreteness on response times in tasks which are not primarily semantic, such as lexical access tasks (see Paivio, 1991, for a review).

Imageability is somewhat similar to concreteness in that it references the perceptual availability of a word or concept. However, where concrete concepts might be considered as such because of their capacity to be perceived by several of the human senses (e.g. they can variously be touched, seen, heard, smelt, or tasted), imageability refers exclusively to the visual modality. Instructions for participants in norm-generation studies typically ask learners to rate, on a 7-point scale, the ease with which a word calls to mind a visual image (Altarriba, Bauer, & Benvenuto, 1999; Bird, Franklin, & Howard, 2001; Stadthagen-Gonzalez & Davis, 2006). The psycholinguistic evidence points to a similar facilitative effect of imageability on both lexical and semantic tasks to that described for concreteness. A clear example of a lexical effect is provided by Balota et al. (2004), who addressed earlier concerns (Gernsbacher, 1984) that imageability may have been confounded with word familiarity. Balota et al. established that more imageable words are responded to more quickly and accurately than less imageable ones in both the lexical decision and word naming tasks (although the effects are attenuated in the latter; Yap & Balota, 2009), even when a wide range of other variables have been controlled. Furthermore, similarly speeded response times have also been found for more imageable words in semantic tasks such as semantic classification (Yap et al., 2012).

While concreteness and imageability are highly correlated (Schock, Cortese, Khanna, & Toppi, 2012), they are not synonymous. Examples of the differences are provided by Altarriba et. al (1999), who suggest that the two variables do not correlate in the case of emotion words such as *miserable* and *furious*, and Bird et al. (2001) cite *armadillo* as an example of a noun which is highly concrete but low in imageability – presumably because while participants know that an armadillo is an animal, few are able to visualise one. This suggests that concreteness ratings can be completed using encyclopaedic knowledge, whereas imageability may be more perceptually grounded. Furthermore, a psycholinguistic distinction between the two variables is suggested by Bird et. al. (2001), which show that imageability, but not concreteness, affects visual word recognition among normal and aphasic participants, and that imageability also explains a greater proportion of variation in word naming tasks among aphasics.

Recent research has suggested that a single variable, *maximum perceptual strength*, may underlie both concreteness and imageability effects (Connell & Lynott, 2012, 2014; M. M. Louwerse & Connell, 2011). This measure is calculated by first gathering ratings of the extent to which a word can be experienced by each of the five senses independently. This results in norms for the word's auditory, gustatory, olfactory, haptic, and visual strength (the latter measurement is strongly correlated with imageability: Connell & Lynott, 2012). The highest of these ratings is then considered as the word's maximum perceptual strength. For example, the concept *baby* might be experienced most strongly through the visual and auditory modalities; slightly less strongly through the haptic and olfactory channels, and barely at all in the gustatory modality. The maximum perceptual strength for *baby* is thus identical to the value for the visual modality, because that is the variable which returned the highest rating. Connell and Lynott (2012) found that maximum perceptual strength was able to explain greater variation than either concreteness or imageability on both lexical decision and word naming tasks. Unfortunately, no research to date has considered the impact of maximum perceptual strength on word association.

While the discussion below refers to concreteness and imageability as two distinct variables, they are grouped together. This is done largely for ease of comparison, but also because few word association experiments have explicitly investigated the latter, resulting in a lack of clear findings. Compounding this is the fact that, in two studies which provide much of what is currently known about the influence of these two variables in word association (de Groot, 1989; van Hell & de Groot, 1998), the authors essentially equate the two variables, using the terms interchangeably. This is done on the basis of a correlation of $r=.96$ between the imageability and the concreteness of the 160 noun cues used in the study. However, such high correlations between the two variables cannot be assumed for all studies. It should therefore

be kept in mind, during the discussion below, that concreteness effects in WA cannot be assumed to equally apply to imageability; more research disentangling the effects of the two variables, along with those of perceptual strength, is needed.

Regardless of the specific loci of concreteness and imageability, the significant effects of the two variables on tasks demanding both lexical and semantic access suggest that an influence on word association is likely. The few existing WA studies looking at these variables appear to confirm this, and findings pertaining to the two variables are generally similar. Perhaps the earliest study to probe concreteness effects in WA is the meta-analysis conducted by Brown (1971). As described in Section 3.4.2, this study assessed the influence of frequency, concreteness, pleasantness, and emotionality on response time and heterogeneity in numerous early WA experiments. The study found that concreteness had the largest overall impact across these variables. With regard to response heterogeneity, five of the seven studies included in the meta-analysis showed significant negative correlations (small-to-moderate in size: mean $r=-.31$ across both significant and non-significant findings) with concreteness: more concrete cues tended to elicit fewer response types. In addition, seven out of eight studies found significant negative correlations (generally moderate-to-high: mean $r=-.42$) between concreteness and response time: more concrete cues were responded to more quickly than less concrete ones.

De Groot (1989) found a similar, though more detailed, pattern of responses for imageability in her oral WA tasks, of which there were both discrete (single response) and continuous (as many responses as possible to each cue in a one-minute period) tasks. Contrasting high (H-I; >3.5 on a 7-point scale) and low (L-I; <3.5) imageability nouns (correlated at $r=.96$ with concreteness, as described above), she found the following significant differences:

1. Highly imageable cues were responded to more quickly (mean RT, across high and low frequency cues in experiments 1 and 2, = 1525.25ms) than less imageable ones (mean RT=1936ms; see Brysbaert et al., 2000, and Van Loon-Vervoorn, 1989, for similar findings)
2. H-I words also had lower omission scores (0.9% of responses blank) than L-I words (4.13%; see Bøyum, 2016, for a similar finding)
3. The most common response to each H-I cue was given more frequently (35.6% of the time) than for L-I cues (23.7%)
4. These primary responses were produced more quickly for H-I cues (1316.5ms) than primary responses to L-I words (1658.5ms)

5. H-I words had a lower heterogeneity (mean response types per cue = 28.1) than L-I words (40.65) in discrete WA, but
6. They resulted in higher *m* scores (i.e. the number of response types given) in continuous WA task (9.6 vs 7.3 mean responses per cue, given 60 seconds per cue).
7. H-I words yielded fewer “idiosyncratic” responses – i.e. response types produced by only one participant – than L-I words.

De Groot interpreted these findings as suggesting that imageability/concreteness influence both connection strength and connectivity. The former took the form of more imageable words becoming more strongly associated with their primary response. This is demonstrated by the greater homogeneity of responses to more imageable cues, the greater association strength of their primary responses, and their faster response times (de Groot, 1989: 837). The influence on connectivity was evidenced by the higher response availability scores. De Groot (Ibid.) suggested that “the concept nodes of concrete words contain more information than those of abstract words” (note here the equation of imageability with concreteness).

This latter interpretation is called into question by the findings of De Deyne & Storms (2008), who reported that out-degree (i.e. the number of response types given to a cue in a multiple-response WAT) did not correlate with imageability. They claimed that the only influence of this variable on connectivity was in the number of in-coming links to a given node (its in-degree), which correlated at $r=.30$ with imageability: more imageable words had a larger number of connections. There are several problems with this finding, however. Firstly, De Deyne & Storms did not investigate the independent effects of their variables, which included frequency and AoA. Given that concreteness correlates with both of these variables, and that frequency and AoA both produced stronger correlations than that of imageability, some confounding of these factors is likely here (although this would not impact the non-existent correlation between imageability and out-degree). Secondly, different data collection procedures were used in the two studies. While De Groot measured discrete and continuous responses, De Deyne & Storms collected three responses per cue. Thus, their measure is somewhere in between the two collected by De Groot, making it impossible to directly compare their findings with either discrete response homogeneity or continuous *m*. As such, the precise pattern of findings pertaining to the influence of imageability on connectivity is not known.

Van Hell & De Groot (1998) followed up on De Groot (1989) using a novel bilingual study. Eighty Dutch L1 participants took two word association tests at an interval of one month. Each participant took the tests in one of four language patterns – Test 1 Dutch-Test 2 Dutch; Dutch-English, English-English, or English-Dutch. The cues were selected in Dutch before being translated into English, and were varied orthogonally between cognates and non-cognates, concrete and abstract. There were 60 nouns and 30 verbs, along with 30 non-critical adjective cues. The authors used two dependent variables. First, they measured the frequency with which responses to these cues were a) the same word in Test 1 and Test 2 (for same-language tests), and b) translation equivalents of one another (for between-language tests). They found that concreteness was significantly associated with both measures: concrete words more frequently resulted in the same word being provided in both same-language tests, and in more translation equivalents between-languages, than did abstract words. The second dependent measure in the study was response time. The authors found that concrete words were responded to more quickly than abstract words, regardless of language, cognate status, and grammatical class. This finding supports the significant facilitative effect on response time of concreteness reported by Brown (1971), and imageability reported by De Groot (1989).

Van Hell & De Groot interpreted their results using a distributed network model, of which the critical aspect was that words share overlapping lexical (e.g. orthographic (time/tide) or phonological (hair/bear)) and semantic features. They suggested that word association responses are selected on the basis of this overlap, such that words with greater overlap will be selected more quickly and frequently. In the case of concrete words, this process will be comparatively fast and consistent because, the authors suggest, conceptual representations of concrete words (e.g. *apple*) share more conceptual features with similar words (e.g. *pear*) than do those of abstract words (e.g. *revenge-anger*). It can be speculated that same processes apply to De Groot's findings pertaining to imageability, except that in that case, at least some features shared by cue and response would be visual ones such as shape, size, and colour. Given that this explanation can account for a higher rate of response reproduction in delayed post-tests, it may also account for the greater response homogeneity observed for highly imageable cues. One problem for this hypothesis, however, is that it may struggle to deal with responses which are neither semantic nor form-based in nature.

Only one study to date has tested the impact of imageability on the categorical type of responses. Bøyum, in a Norwegian language study, found that the imageability of noun cues did not significantly influence the proportion of meaning-, position-, and form-based responses given to each cue. These results require

further investigation, however, since the study used a relatively low number of cues with a relatively uneven distribution in terms of imageability: only 10 high imageability noun cues were used in the study, compared with 31 of low imageability.

Finally, the study by Van Rensbergen et. al. (2015) described in Section 3.4 also points to an influence of concreteness on WA responses; in that study, the concreteness of cues emerged as a significant predictor of the concreteness of responses. This further suggests that this variable plays a role in determining the strength with which words are associated. As the authors noted, however, the study is essentially correlational, meaning that cue and response concreteness may simply co-vary, with no implication of causality (*Ibid.*: 6).

In summary, both concreteness and imageability appear to significantly influence word association responses. In the case of the former, research has uncovered the following effects:

- Concrete cues result in greater homogeneity of responses than abstract ones
- Concrete cues are responded to more quickly than abstract ones
- Responses to concrete words are more consistent (i.e. more likely to be produced again in a follow-up test, in the same or a different language) than abstract ones.
- The concreteness of a cue is correlated with the concreteness of its response: concrete cues yield similarly concrete responses.

This pattern of results is similar to that given above, pertaining to imageability. While the evidence for both variables is fragmentary and limited to some measures and not others, the consistency of the patterns demonstrates a strong and clear influence of both of these variables. Models provided by De Groot (1989) and Van Hell and De Groot (1998) suggest that both variables influence connectivity and connection strength. Moreover, the similarities between these patterns suggests a similar locus of effect: investigations into maximum perceptual strength (Connell & Lynott, 2012) may shed light on this.

3.4.5 Affective variables

Only a small handful of studies have investigated the impact of affective variables on WA. Those to have done so, however, have uncovered relatively consistent effects. Early studies in this fashion were compared in Brown's (1971) meta-analysis, which looked at the influence of two emotion-related variables: pleasantness, which references the extent to which words are considered agreeable; and emotionality, which refers to the extent to which a word evokes emotion. A key difference between the

two measures is that pleasantness is a bipolar measure, capturing both pleasant and unpleasant emotions, while emotionality refers only to the strength of emotion, rather than its polarity.

Brown's results suggested that the emotionality of a cue has a larger impact on WA measures than its positive or negative orientation. More emotional responses resulted in increased response heterogeneity, and were responded to more slowly than neutral words. Brown, however, notes that these findings, particularly those pertaining to response time, may have been confounded with those of concreteness, because more strongly emotional words also tend to be more abstract. In the case of pleasantness, more pleasant concepts were associated with reduced response homogeneity and faster response times. These findings were only significant, however, in three out of twelve of the studies analysed; the effect size in each of these significant studies was weak. Unfortunately, no dedicated attempts have been made to clarify the influence of emotionality and pleasantness in word association.

The impact of cue affective strength was taken up in two studies by Van Rensbergen et al. (2016, 2015). The first, which was initially described in Section 3.3.1, examined assortativity in a WA-derived network. The specific properties under investigation were different from those examined by Brown (1971). They were valence (the positivity of a word: an example of high valence is *Christmas*; low valence: *torture*); dominance (the sense of control a respondent feels over a concept high = more control: *project*; low = dominated by the concept: *earthquake*); and arousal, (the degree of activeness of a concept; high: *tornado*; low: *asleep*; all examples are from the English norms provided by Warriner, Kuperman, & Brysbaert, 2013; Van Rensbergen et. al. used Dutch cues and therefore collected norms from a range of other sources).

The study found that the valence, dominance, and arousal of cues explained unique variance in the same property of their responses (e.g. high cue dominance predicted high response dominance, but not response valence or arousal). The strongest effect was for cue valence, which explained 31% of variance in response valence in a regression model. The corresponding figure for dominance was 15%, and for arousal 17%. These figures, which are based on response tokens rather than response types and therefore reference association strength, suggest a role of emotional variables in the organisation of the lexicon.

Van Rensbergen et. al.'s follow-up study (2016) further suggests the importance of affective variables to lexical structure. The authors attempted to demonstrate that a large WA-derived network could be used to accurately estimate human ratings of valence, dominance, and arousal. This was done by calculating the overlap in response profiles of all words in the network. Two given cues were assumed to overlap to the extent that they yielded the same responses, with the same frequency, as one another. This overlap

was measured using cosine similarity, and was assumed to reflect semantic similarity. Estimates of a given word's emotional properties were then calculated with reference to the cosine similarity of other words for which human emotion ratings were already available. The study found that this WA-derived method correlated more strongly with human emotion judgements than similar methods based on overlapping corpus-based distributions.

While this method provides only indirect evidence for the centrality of emotional factors in lexical organisation, it extends Van Rensbergen et. al.'s earlier finding in that it suggests that overlap in associative response profiles (as opposed to the covariation between individual words demonstrated in the earlier study) is reflective of similarity in emotional properties. However, neither of these studies can be taken as evidence of causality in the relationship between emotional properties and response distribution. While it may be the case that e.g. high valence *causes* respondents to search for words with similar properties, it may equally be the case that such words become associated, and develop overlapping associative profiles, for reasons independent of their shared valence, such as their semantic similarity. As Van Rensbergen et. al. point out, however, a word's emotional properties can be viewed as one aspect of this semantic similarity.

The above summary suggests that while affective factors do influence word association responses, not enough research has demonstrated either the nature of this effect (e.g. on more common response measures such as association strength, response categories, and response times), nor its independence from other factors.

3.5 Discussion

The preceding section gave an overview of findings regarding the influence of cue-level variables on several measures of the word association task. Several important points can be made on the back of this analysis. Firstly, while the wide range of methods used to measure word association has frequently caused problems of comparability, it has become clear that this abundance of methods is fundamental to the value of WA research. This is because the use of multiple measurements reveals effects which would simply be invisible to a more limited methodology. The important influence of frequency on in-coming connectivity, for example, is an insight which would have been impossible using less sophisticated measurements than those designed to model network centrality.

Secondly, if we take the view that a perfectly clear picture of cue-level variation would require data from numerous different measures, including response categories, response distributions, response times, and

network centrality measurements, then it is clear that the viewpoint afforded by current research is incomplete. Such a complete picture of research findings is unavailable for any of the variables described above, save perhaps for frequency (although, in this case, critical findings such as those of De Deyne & Storms (2008) need to be tested for independence of effect). Numerous important findings are based on isolated studies or novel methodologies. As such, research needs to flesh out the basic findings presented above using a wide range of methodological approaches.

Third, it is notable that the same features which appear to influence in-degree (i.e. frequency and AoA) do not appear to strongly influence out-going links (De Deyne & Storms, 2008). This suggests that the processes involved in producing word association responses may not be the same ones which govern the structure of the lexicon. Frequency and AoA have been hypothesised to affect network centrality through a sort of early bird principle: those words which enter the lexicon earliest become anchors for later acquired words (De Deyne & Storms, 2008; Steyvers & Tenenbaum, 2005). This process, however, appears to have little effect on the heterogeneity of responses to these words (although early acquisition does appear to speed response times). This opens an avenue for new research into varying aspects of response production (semantic and distributional profiles, RT, response category, heterogeneity, response availability) for words which are highly central according to measures of in-degree. Such research might shed light on the relationship between WA response production and network structure.

Fourth, some initial support for the notion of distinct patterns of influence from distributional and semantic variables, suggested in Section 3.3.1, has been found, although the fragmentary nature of the picture presented above precludes the assertion of any strong pattern. Limited evidence suggests that distributional variables (in the above, frequency and age of acquisition) appear to influence network centrality: more frequent and/or earlier acquired words have more in-coming links to other words, and are thus more central to the lexicon. This is demonstrated by their tendency to be produced as WA responses. However, as noted above, the independence of these effects, both from each other and from other variables, needs to be established. There does not appear to be any consistent effect of distributional variables on connection strength. On the other hand, semantic variables (concreteness, imageability, and emotional variables) appear more likely to influence connection strength. Cues high in these values appear to have stronger primary associations, and these associations are produced more quickly than cues with lower values (though this pattern has not yet been shown to apply for emotional variables).

Fifth, there are several cases in which variables are potentially confounded with others. This problem has been noted above regarding the influence of frequency and AoA on network centrality; AoA is a potentially confounding influence on any study seeking to control for, or measure, frequency effects. Similar problems which emerged in various parts of Section 3.4 include possible confounds of imageability and/or emotional variables with concreteness, of frequency with familiarity, and of grammatical class with multiple variables. Future research needs to take care to avoid these pitfalls by designing cue lists which control for potentially confounding features.

Lastly, the analysis only rarely revealed patterns of interaction between variables. Perhaps the only consistent pattern was the interaction of grammatical class and frequency, which took two forms. Firstly, for all lexical classes, form-based responses rose as frequency declined (c.f. the possible confound with familiarity). Second, frequent adjectives tended to receive more paradigmatic responses than infrequent ones. That more interactions were not revealed, however, does not mean that they do not exist; it may simply be that further research into these variables is needed.

One of the main aims of this review was to identify the variable(s) most in need of further research. The above discussion suggests that all of the variables investigated so far, and others besides, are in need of additional study. One way to approach this research would be to emulate the large scale “megastudies” currently used in psycholinguistics to explore the influence of numerous variables across thousands of cues (e.g. Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Brysbaert et al., 2011; Keuleers, Lacey, Rastle, & Brysbaert, 2012). The main benefits of this approach are the great statistical power available from such large amounts of data, the eradication of the need to carefully select cues (as a consequence of the size and diversity of the cue set), and the comparability of the resulting data to studies in other areas of psycholinguistics, or involving other psycholinguistic tests.

However, there are significant difficulties pertaining to the use of these methods to research word association. The main challenge is statistical. Psycholinguistic megastudies depend upon hierarchical regression analysis to investigate response variation. This type of statistical analysis is appropriate for some WA measures, such as association strength, response homogeneity, and response time, which return interval data. This means that effective large-scale studies could be applied to WA using measures of these types.

However, perhaps the defining feature of WA studies is that they yield categorizable responses. As the above discussion reveals, these categorizations form the bedrock of much of what is known about WA

(e.g. Entwisle, 1966a; Entwisle, Forsyth, & Muuss, 1964; Ervin, 1961; Fitzpatrick, 2009; K. Nelson, 1977; Nissen & Henriksen, 2006). This variable is far less suitable for regression analysis than those listed in the previous paragraph. This is because, as a categorical variable, response categorizations need to be transformed from into interval data. This is theoretically possible (for example by counting how many responses of each type are given to each cue), but comes at the cost of the detailed response categories developed by Fitzpatrick (2006, 2007, 2009; Fitzpatrick, Playfoot, Wray, & Wright, 2015), because the use of so many categories would result in data which would be too sparse to be informative. Since these detailed categories have been demonstrated to reveal distinctions hidden by more basic categorization schemes (Chapter 2, and Fitzpatrick, 2007, 2009; Higginbotham, 2014), the curtailing of the efficacy of this measure implicit in large-scale studies makes the use of these studies to investigate WA undesirable. This is the case even when the significant practical challenges of conducting large-scale research with WA, which include the collection of large amounts of data from hundreds, or even thousands, of respondents, and the categorization of potentially tens of thousands of responses, are ignored.

Moreover, as described in Chapter 1, the view taken in this thesis is that the most effective WA research to date has stemmed from the use of careful, iterative research designs which systematically seek to test given hypotheses. Such an approach is best suited to the use of small-scale, single-variable factorial studies, the results of which can, if necessary, be replicated using large scale at a later date (Balota et al., 2012; Keuleers & Balota, 2015; Kuperman, 2015).

Which variable, then, offers the greatest potential for future research? The clearest answer is grammatical class. There are several reasons for this. Firstly, previous research has already uncovered a significant and consistent influence of grammatical class on response category: nouns elicit more paradigmatic responses than do either verbs or adjectives. This finding provides a base upon which to build new studies. Secondly, Fitzpatrick's (2007, 2009) innovations in the categorisation of WA responses (see also Higginbotham, 2014) mean that new subtleties in this pattern can now be explored. Third, it remains unclear how grammatical class influences other measures, such as response distribution. New studies on WA therefore need to have the objective of filling in these research gaps. Fourth, the potentially confounding effects of the variables discussed above are nowhere clearer than in the case of grammatical class. This is because concreteness, imageability, AoA, and possibly emotional variables, have all been put forward as potential explanations of grammatical class effects (Altarriba et al., 1999; Bird et al., 2001; Gentner, 2006; Nissen & Henriksen, 2006; Vigliocco, Vinson, Druks, Barber, & Cappa, 2011; Waxman et al., 2013). A natural aim of research on grammatical class, then, would be to establish the independence of its effects from potential

confounds. Fifth, the discussed offered by Nissen & Henriksen (2006) suggests several cognitive and distributional factors, rarely discussed in the WA literature, as possible explanations for grammatical class effects (see Section 3.4.1, above). Their discussion suggests that investigating this influence may lead to new insights into the WA task. Finally, there is a rich psycholinguistic literature on grammatical class, from which hypotheses for WA studies can be drawn, and to which word association data can be compared (Cordier, Croizet, & Rigalleau, 2013; Druks, 2002; Vigliocco et al., 2011). Thus, such studies can contribute to wider psycholinguistic research. For these reasons, the next experimental stage of this thesis will focus on elaborating the influence of grammatical class on WA responses.

Chapter 4: Do nouns and verbs elicit different types of word association response?

4.1 Introduction

Three main points regarding the status of word association research were established in the preceding chapters. Firstly, Chapter 1 established that results from existing WA studies have not provided a clear body of knowledge about the WA task. This was shown to be due to a combination of factors, including a lack of methodological consistency in existing studies and insufficient understanding or control of key variables. Chapter 2 suggested that the cues selected for WA studies are one such variable, because the category and distribution of responses to be influenced by the properties of individual cues. In Chapter 3, it was argued that several aspects of a cue's distribution and semantics contributed to this unpredictability of response patterns, and that analyses of these factors had the potential to contribute not only to the principled selection of cues in future WA studies, but also to discussions regarding the organisation of the mental lexicon.

The discussion in Chapter 3 concluded by suggesting that an examination of the influence of grammatical class on word association responses offers significant potential for the deeper understanding of cue-level influence on WA. The reasons for this suggestion were that previous research into the influence of grammatical class is limited by its failure to demonstrate the independence of that variable from others, such as concreteness and frequency; and secondly that several further avenues for exploration remain unexplored. These include the use of additional measures such as distributional information; a further level of granularity in response categorization through the use of a detailed scheme of responses, or through the investigation of under-researched classes such as adjectives, adverbs, and prepositions. All of these possibilities are in line with the suggestion, made in Chapter 1, that the field of word association would benefit from cumulative, iterative research approaches. In view of these arguments, the present chapter will explore the influence of the cue's grammatical class on WA responses.

4.1.1 Grammatical class selection

Chapter 3 suggested that the influence of all grammatical classes (i.e. not only nouns and verbs but also adverbs, pronouns etc.) on WA response was under-researched. However, it is difficult for a single study to address many classes simultaneously. This is because of practical limitations on the number of cues that can be presented to participants. Most WA studies limit themselves to no more than 100 cues, because exceeding this is assumed to be fatiguing for participants (Fitzpatrick, 2006, 2007; Namei, 2004;

Sökmen, 1993). For this reason, only two GCs will be studied in this chapter. This allows 50 cues of each class to be presented – a large enough sample to allow clear patterns to emerge.

In the interests of building upon existing research findings, this chapter will focus on noun and verb cues. These are the classes to have received most attention both in WA and other areas of psycholinguistics (e.g. Deese, 1962; Kemmerer, 2014; Kemmerer & Eggleston, 2010; Nissen & Henriksen, 2006; Vigliocco, Vinson, Druks, Barber, & Cappa, 2011). One feature of this research has been the use of the noun/verb distinction to test competing models of language processing. Vigliocco et. al. (2011), for example, compare the evidence for lexicalist, combinatorial, and emergentist models from noun/verb studies. Each of these models has distinct implications for the interpretation of WA research. This means that noun/verb WA studies offer an opportunity for exploration of general psycholinguistic models: something which is rarely done by WA researchers. While it is not the aim of the current chapter to investigate these models directly, their implications for WA will be returned to in Chapter 6.

Nissen and Henriksen's (2006) study has already hinted at the potential of noun/verb WA research to yield insights into the psycholinguistic nature of association responses. Their study, which compared responses to 15 noun, 15 verb, and 15 adjective cues drawn from the 2000- and 3000-word bands of Nation's Vocabulary Levels Test (Nation, 2001), supported earlier findings of a tendency for noun cues to receive more paradigmatic responses than did verb cues (Deese, 1962a; Entwisle, 1966a). Nissen and Henriksen (2006) then explored several potential explanations of these differing patterns. These factors, briefly discussed in Section 3.4.1, will now be examined more fully.

Nissen and Henriksen (2006) argued that several factors could contribute to differences in response category patterns to noun and verb cues:

- a) Differences in the age and nature of acquisition of items from the two classes;
- b) The syntactic properties of the classes. For example, they describe verbs as "relational" (2006: 403) in that their syntactic role is to relate other elements of sentences (e.g. nouns) to each other; this makes them more cognitively dependent upon other sentence constituents;
- c) The way in which they are integrated into the lexicon. For example, verbs may be stored together with the nouns to which they commonly refer;
- d) The differences in semantic relations between the classes, with nouns largely forming hierarchical relations of hypo- and hypernymy, while verbs tend not to do this;
- e) Differences in the extent of cognitive processing of the classes during WA response production.

The most straightforward of these factors is (d), which identifies differences in the structure of semantic relationships amongst nouns and verbs as a potential reason for the greater proportion of paradigmatic responses to nouns. Nissen and Henriksen (2006: 402) suggested that the extensive hierarchical relationships between nouns make these cues more likely to yield paradigmatic responses, because the availability of words at higher, equal, or lower levels of specificity is likely to be considerable for noun cues. In the case of verb cues, such words are not always available because of the less structured nature of verb semantics. Nissen and Henriksen (2006: 403) suggest, for example, that verbs have no “basic level” of meaning which allows respondents to give substantial information about a concept with minimal cognitive effort (Miller & Fellbaum, 1991).

This distinction may not apply equally to all nouns and all verbs, however. As Nissen and Henriksen (*ibid.*; 402) point out, hierarchical relations are typical for concrete nouns, but may apply less clearly to abstract ones (e.g. the relationship of *truth* to *justice* is not as clear as that of *table* to *chair*, which are related through the hypernym *furniture*). This suggests that a word’s concreteness may interact with the influence of grammatical class with regard to the cognitive availability of meaning-based WA responses. With this potential confound in mind, I conducted a test to investigate whether Nissen and Henriksen’s noun, verb, and adjective cues differed in concreteness. I gathered concreteness ratings for each of Nissen and Henriksen’s cues from a norms list provided by Brysbaert, Warriner, & Kuperman (2013), and calculated mean concreteness scores (rated from 1 to 5) for the (English)⁵ cues of each grammatical class. These ratings were then tested using a one-way ANOVA, with grammatical class as the factor. This analysis revealed significant differences in concreteness between the noun ($M = 3.74, SD = 1.04$), verb ($M = 2.90, SD = .87$), and adjective cues ($M = 2.70, SD = .75$), $F(2, 87) = 11.429, p < .001$. The size of the effect was large (eta squared = 0.21; Borenstein, 2009). A post hoc SNK test revealed that the effect was attributable to the significantly higher concreteness of the nouns ($p < .05$) compared to the verb and adjective cues. There was no significant difference in concreteness between the verbs and the adjectives. This analysis increases the possibility that earlier studies, including the Nissen and Henriksen study, may have partially confounded the effects of grammatical class with those of concreteness, and provides a rationale for the control of this variable in this chapter.

⁵ All cues in Nissen and Henriksen’s study were first selected in English, and then translated into the participants’ L1 (Danish). Both sets of words were used in the study. It is conceivable that the concreteness ratings would vary across languages.

Returning to Nissen and Henriksen's list of factors in noun/verb WA effects, point (a) implies a relationship between GC and the age at which words are acquired (age of acquisition, or AoA). Numerous researchers have found that nouns are, in general, acquired earlier and more easily than verbs (Gentner, 2006; Ma et al., 2009; McDonough et al., 2011; Waxman et al., 2013). The review in Chapter 3 found that, although the difficulty of norming AoA values leaves some room for doubt as to the construct's validity, there is nevertheless support for the notion that AoA influences the structure of the lexicon, suggesting that early acquired words may be more central to the lexicon than words with later AoA. This suggests a possible tendency for nouns to be more central to the lexicon than verbs. In terms of WA results, this should lead to faster RTs to noun than verb cues, as well as a higher in-degree (the number of times a word is given as a WA response to other cues) for nouns. The studies reviewed in Chapter 3 did not predict an influence of AoA on response category. Nevertheless, as in the case of concreteness, there are sufficient grounds here for also controlling AoA in this chapter.

Factors (b) and (c) differ somewhat from (a) and (d) in that they imply syntactic, rather than distributional or semantic, influences on WA responses. Factor (b) lays the groundwork for this influence by linking the syntactic roles played by cues in sentence processing (e.g. the role of verbs in relating syntactic components to each other) with their associative response patterns (e.g. the production of syntagmatic responses to verb cues). Factor (c) elaborates this connection by suggesting that verbs become associated, in the lexicon, with the subject and object nouns which they modify. This association is the result not merely of co-occurrence of two words, but of cognitive work undertaken in relating these words to each other during sentence construction. This implies a model of language processing which allows feedback between syntactic processing and lexical storage – an issue which will be returned to in Chapter 6.

Factors (a-d) all refer to the architecture of the mental lexicon. They are thus reflective of a common view of the WA task – that it provides, in the words of one researcher, “a window to the lexicon” (Sökmen, 1993; see also Boyum, 2016; De Deyne & Storms, 2015; Dubossarsky, De Deyne, & Hills, 2017; Fitzpatrick, 2007; Namei, 2004; Van Rensbergen, Storms, & De Deyne, 2015; Zareva, 2007, 2012). From this point of view, the production of WA responses constitutes little more than lexical access, activation, and selection. Wider processes, such as those involved in discourse participation, are not conceptualised as being part of WA.

Factor (e) refers to the possibility that additional cognitive processing, beyond lexical access, is a part of the WA response process, and that the extent of this processing can vary according to the features of the

cue. Specific cognitive processes are not detailed by Nissen and Henriksen, but might hypothetically include access to syntactic/integrative processing, conceptual knowledge, episodic memory, or still other processes, in addition to lexical access. The addition of these processes results in a much more dynamic conception of WA than is frequently put forward in research studies. However, considerable additional research is needed to confirm Nissen and Henriksen's hypothesis.

In summary, Nissen and Henriksen's discussion identifies several factors potentially contributing to the different patterns of response category to noun and verb cues. These points imply certain assumptions as to the nature of the lexicon: its organisation preserves, to at least some degree, the hierarchical relationships between words, although this may be modulated by concreteness and GC (factor d); it is influenced by the age at which words are acquired (a), and it appears to allow feedback between syntactic processing and lexical knowledge, in that words can become associated as a result of cognitive effort applied in the process of integrating concepts during sentence processing (b and c). Factor (e) additionally suggests that lexical access may not be the only process at work in the production of WA responses; unspecified cognitive processing may also be influential.

The study described in this chapter engages most directly with (a) and (d), though all five considerations will be revisited in the discussion. Factors (a) and (d) highlight the possible interaction of age of acquisition and concreteness with GC in earlier research studies, implying a need to control these factors in order to develop a clearer picture of the influence of GC on word association. The rationale for this work is that while the insights above offer a platform from which future work can move toward a psycholinguistic model of word association, such theoretical questions cannot be adequately addressed without a clear understanding of the influence and independence of cue-level variables.

4.2 The current study

The main aim of the current study is to establish whether the differences in responses to noun and verb cues reported by Deese (1962), Entwisle (1966a), and Nissen & Henriksen (2006) are independent of the influence of several potentially confounding factors, including concreteness and AoA. In addition, the study aims to expand earlier findings by testing for an influence of GC on distributional response patterns.

4.2.1 Participants

All participants in the study (N=71) were English L1 undergraduate students enrolled in an English language-related course at a university in the UK. All students were aged 18-21 years of age. None of the students were aware of the aims of the study, and all gave informed consent.

4.2.2 Cue selection

In order to establish the independence of noun and verb response differences, cues were selected in a manner which controlled for the effects of five key variables. These were, firstly, concreteness and age of acquisition, both of which were discussed above (concreteness ratings were taken from a large-scale norms database created by Brysbaert et al. (2013), and AoA ratings were taken from a database collected by Kuperman et al. (2012)). Secondly, word frequency, word length in syllables, and bigram frequency (“the average frequency of the letter pairs in the word”; Kuperman et. al., 2012) were also controlled. These variables were selected because of the strength of their influence on other psycholinguistic tasks (Brysbaert et al., 2011; Kuperman et al., 2012). Brysbaert et. al. (2011), for example, suggested that these variables together accounted for around 55% of variation in lexical decision response times. Bigram frequency in particular may be a possible influence on the number of form-based responses.

The selection of a source for frequency values was not straightforward. Word association studies to date have drawn variously on corpora (e.g. the British National Corpus (BNC) used in Higginbotham, 2014) and pedagogical word lists or vocabulary tests (which are derived from corpora but group words into frequency bands rather than providing detailed frequency information; e.g. the Academic Word List (AWL; Coxhead, 2000, used in Fitzpatrick, 2007), and the Vocabulary Levels Test (VLT; Nation, 2001; used in Nissen & Henriksen, 2006)).

In the present study, frequency values were taken from the SUBTLEX_UK corpus (van Heuven, Mandera, Keuleers, & Brysbaert, 2014), which is composed from subtitles from a range of BBC television channels broadcast between January 2010 and December 2012. This corpus was selected for several reasons. Firstly, it provides more detailed frequency counts than pedagogical vocabulary lists. Secondly, at more than 200 million words, it comfortably exceeds the 16-17 million words recommended by both Brysbaert & New (2009) and Balota, Cortese, Sergent-Marshall, Spieler, & Yap, (2004) as a minimum corpus size necessary for providing accurate frequency estimates.

Finally, as Brysbaert and New (2009) point out, perhaps the most important concern regarding corpus selection is representativeness. Even very large corpora are not necessarily representative of a given

participant's language exposure because they can be collected from sources unfamiliar to the participant(s) in question. As such, of an ideal source of frequency counts will also take into account demographic aspects of participant group. The SUBTLEX_UK corpus is well matched with the participants used in the current study because it is composed of subtitles from a wide range of BBC television programmes, including TV aimed at children and young adults; drama, documentary, and political programming. The participants in this study are likely to be familiar with at least some of these programmes, which were broadcast while they were in their teens. This gives it an advantage over corpora such as the British National Corpus, which is constructed from texts published before these participants were born. Support for the selection of this corpus is provided by Van Heuven, Mandera, Keuleers and Brysbaert (2014), who compared frequency counts from a number of corpora with response time data obtained from a large-scale lexical decision norming study using participants demographically similar to those in the current chapter⁶ (Keuleers, Lacey, Rastle, & Brysbaert, 2012). Van Heuven et. al found that the SUBTLEX_UK corpus provided the best fit, explaining almost 3% more of the variation in response times than frequency values drawn from the BNC.

The SUBTLEX_UK corpus uses a novel logarithmic measure, the Zipf Scale (van Heuven et al., 2014: 1178), to convey word frequency information. The scale is essentially a simplification of the more standard frequency per million words (fpmw) measure. It uses a seven-point spectrum, further delineated by two decimal places, where a value of 1 represents words with a frequency per million of 0.01, and 7 denotes an fpmw of 10,000 (see Table 4.1, below). The authors created this measure in order to address a common misunderstanding in the interpretation of the fpmw measure; namely that the difference between 0.1fpmw and 1fpmw appears to be a relatively small difference when in fact it accounts for the same amount of variation in response times as the difference between 1 and 10fpmw. Note that the Zipf measure is a *scale* – it does not categorize the frequency of responses. Van Heuven et. al. report values to the fifteenth decimal place. The frequency values reported in this study will be presented as Zipf values.

Measures for length in syllables and bigram frequency were obtained using the N-Watch program (Davis, 2005). As well as collecting information on a range of variables from numerous norming studies (e.g. Bird, Franklin, & Howard, 2001; Coltheart, 1981), this software calculates neighbourhood size values, such as bigram frequency, by referring to a default vocabulary of more than 30,000 words.

⁶ However, it should be noted that data were collected not only from undergraduates, as in the present study, but also from postgraduates and academic staff.

One hundred cue words (50 nouns and 50 verbs) were selected for the study. Each word was chosen for its proximity to a baseline value chosen for each of five control variables. These were:

- Concreteness = 3, on a scale of 1 (very abstract) to 5 (very concrete)
- Age-of-acquisition = 10 years
- Frequency = 3, on a scale of 1-7 (corresponds to a FPMW of 1; see discussion above)
- Word length in syllables = 2.5 (items were two, three, or four syllables in length)
- Bigram frequency = 45

Table 4.1
Comparison and examples of Zipf values (from Van Heuven, Mandera, Keuleers, and Brysbaert (2014).

Zipf value	fpmw	Examples
1	0.01	antifungal, bioengineering, farsighted, harelip, proofread
2	0.1	airstream, doorkeeper, neckwear, outsized, sunshade
3	1	beanstalk, cornerstone, dumpling, insatiable, perpetrator
4	10	dirt, fantasy, muffin, offensive, transition, widespread
5	100	basically, bedroom, drive, issues, period, spot, worse
6	1000	day, great, other, should, something, work, years and,
7	10,000	and, for, have, I, on, the, this, that, you

Table 4.2
Comparison of variable values for noun and verb cues

	Nouns				Verbs				<i>p</i>
	Mean	SD	Min	Max	Mean	SD	Min	Max	
Frequency	3.09	.22	2.66	3.39	2.99	.21	2.68	3.38	.101
Concreteness	2.97	.20	2.59	3.39	2.99	.18	2.59	3.41	.506
AoA	9.98	.60	8.67	10.80	9.81	.57	8.58	10.85	.300
Length (syllables)	2.5	.60	2	4	2.28	.52	2	4	.347
Bigram frequency	41.73	30.11	9.86	151.14	48.95	17.89	20.67	88.50	.396

Each of these baseline values was selected to reduce the likelihood of interactions with the independent variable (grammatical class) in unpredictable ways. An example of this would be if words with a concreteness of 4.5 or above (on a scale of 1-5) were selected. Because relatively few verbs reach this level, it is possible that responses to such verbs in word association tests might be somewhat atypical. As such, median-level values were generally selected for all variables.

In order to confirm that the noun and verb cues did not differ significantly with regard to these variables, t-tests were performed to compare mean values for each. The results (see Table 4.2) show that the differences between the noun and verb groups did not approach significance for any of the variables. This is supported by the similar means and low standard deviations shown in Table 4.2.

4.2.2.1 Additional cue selection criteria

Two additional criteria were applied to the cue word selection. Firstly, words which could potentially occupy more than one grammatical class (e.g. *scuttle*, *mimic*) were not selected. These words could potentially have influenced results because participants may not have responded to the intended grammatical class. Secondly, words which could form a different grammatical class through the deletion of a single morpheme (e.g. *raider*, *enforcer*) were not permitted. This decision was made in view of the possibility that derived words may be processed in a compositional manner, with the stem being processed first and the meaning contributed by the affix only attached afterwards (see e.g. Marslen-Wilson, Tyler, Waksler, & Older, 1994).

4.2.3 Materials

The cue words were presented on four sheets of double-spaced A4 paper. The cue list is presented in Appendix 2, in the same format in which it was given to participants, except that the items marked with an asterisk were not so marked in the participant version. These items will be discussed below. Respondents were asked to “Please write the first word that comes to mind for each of the following words”. They were reminded that there were no correct answers, and were also asked not to change their answers once they had been written. No time limit was given for providing answers. All participants completed the list within 17 minutes.

4.2.4 Dependent measures

Two dependent measures, response category and response distribution, were collected. A modified version of the categorization scheme used in Chapter 2 was used to categorise responses. This new version was taken from Fitzpatrick, Playfoot, Wray, & Wright (2015). The differences between the two schemes all pertain to the detailed level of categorisation. Firstly, the “defining” and “specific” synonym categories from the earlier scheme were collapsed into a single “synonym” category. Secondly, the stipulation that collocations which were not likely to be consecutive words had to be placed into the “other collocation” category was removed; the new scheme allows collocations to be cue-response, response-cue, or both. Thirdly, a two-step category was added to capture responses connected via an intermediary (e.g. *wallow*

– *tree*, connected via *willow*). Finally, responses were allowed to be categorised as “dual-category” in the event that both meaning-based and position-based interpretations were permissible. The full scheme is produced in Table 4.3.

It should be noted that while the basic level categories closely resemble the paradigmatic/syntagmatic/clang system, and will be compared as such, they are not entirely synonymous. The largest difference concerns the role of grammatical class in response coding. Firstly, while in most cases responses categorised as “meaning-based” were from the same grammatical class as the cue, this was not a criterion for entry into this class, as it is in both Deese (1962⁷) and Nissen and Henriksen (2006). Similarly, the position-based category did not consider membership of *different* grammatical classes to be a necessary criterion for selection. For example, the response *ice* to the cue *cream* would be coded as position-based due to its frequent occurrence in the lexical item *ice cream*. This is despite that fact that both words are nouns. Coding schemes determined by grammatical class membership may occasionally therefore code responses differently. As discussed in Chapters 1 and 2, the main rationale for the adoption of this scheme was that its detailed level allows for previously invisible distinctions to emerge without sacrificing comparability with existing research.

Response distribution data was collected using the scheme developed by de Groot (1989), also described in Chapter 3. This involves the collection of four different response counts:

1. The “associative frequency of the primary response” (AFPRIM) – i.e. the number of response tokens for the most common response type given to each cue,
2. The number of response types with exactly one response token, to each cue ($N=1$)
3. The number of response types with more than one token, to each cue ($N>1$)
4. The total number of response types to each cue (i.e. the sum of values 2 and 3, above; $N>1 + N=1$)

As Chapter 3 points out, little is known about the influence of grammatical class on response distribution; the rationale for the collection of this data is therefore simply to establish research findings in this area.

4.2.4.1 Coding procedure

All responses to the 100 cue words ($N=7100$) were first placed into one of the four basic-level categories. They were subsequently placed into a detailed group, which may include a dual-category coding.

⁷ Deese (1962) does not use the terms “paradigmatic” or “syntagmatic” in his study. Nevertheless, his system of categorizing responses into either same or different form class categories is essentially the basis of the syntagmatic-paradigmatic distinction.

All responses were categorised by two coders, who worked independently. The same two coders were used in both this study and the following one. The first coder was the author. The second was an English Language teacher which around 10 years of classroom experience. He possessed an MA in Applied Linguistics, and also had a knowledge of Latin and Greek which offered a useful insight into possible etymological interpretations of responses. In both this and the following chapter, the second coder was naïve to the aims of the study. The percentage agreement between the two codings was then calculated to be, for the basic level, 91.1%; and for the detailed level, 81.26%. All cue-response pairs which had not been agreed upon during this independent coding process were then resolved through discussion between the two coders. In cases of disagreement, the opinion of the (naïve) second coder was preferred in order to reduce any potential bias on the part of the first coder. As in Fitzpatrick et. al. (2015), the most common reason for disagreements was the failure of one or other of the coders to notice a particular sense of a cue word's meaning. An example of this is the cue *compose*. A coder focused on the sense of *create a musical composition* might fail to notice a response, such as *compose-yourself*, in which *compose* is responded to in the sense of *regain calmness*.

4.2.4.2 Excluded dependent measures

One of the key assertions in this thesis so far is that a fully-realised model of WA will require the collection of diverse WA measurements, including response time and network-based measures. The absence of these measures in this chapter therefore requires some explanation. Looking first at response time measures, one of the goals of the chapter is to compare existing findings with data which is similar in as many ways as possible, excepting the control of potentially confounding variables. With this in mind, it was felt that the complexity of setting up either voice- or keyboard-activated RT triggers could influence comparability with earlier studies, which were all conducting using pencil-and-paper methods. Regarding network-based data, the present study is simply incompatible with such an approach, which requires the collection of responses to a far greater number of cues than are to be presented here. The addition of network measures, then, will have to wait until existing databases become publicly available.

4.2.5 Research questions

Three questions were being addressed in this study:

1. Do noun and verb cues differ in terms of the stereotypy of the responses which they elicit, when key variables are controlled?
2. Do noun and verb cues differ in terms of the category of the responses that they elicit, when the same variables are controlled?

3. What can be learned from the analysis of responses to nouns and verbs using detailed levels of analysis?

4.2.6 Removal of unsuitable cues

It emerged during the coding process that a large number of cues failed to meet the initial criteria set out for cue selection. In particular, a number of words which appeared on the list turned out to have additional senses which had been missed during initial cue selection. Where these senses changed the grammatical class of the word, it was necessary to remove the word from the analysis. An example is the word *scuttle*, which had been selected as a verb but occasionally yielded the response *coal*, which pointed to its noun meaning. Another problem was the selection of derived words which could change grammatical class through the removal of an affix. An example which illustrates the problem is *recharge*. This word was selected as a verb, and also on the basis that its stem, *charge*, is also a verb. The noun sense of *charge* was not noticed until during the coding process. These and similar examples were removed from the analysis, resulting in the deletion of a total of 26 cues. These words are marked with an asterisk in Appendix 2.

Analysis of the responses to these cue words appears to bear out the reasons for their removal. *Recapture*, for instance, elicited a number of responses which appeared either to be prompted by the stem (*kidnap*, *claim*, *caption*), or by the affix (*repeat*, *return*). This might suggest an influence of derivational morphology on the processing of word associations (cf. Bozic, Tyler, Su, Wingfield, & Marslen-Wilson, 2013; Diependael, Grainger, & Sandra, 2012; Marslen-Wilson et al., 1994). These examples also suggest the difficulty of noticing all possible grammatical and semantic properties of words during categorization. As such, future studies need to take precautions to ensure that all cue words meet all of the criteria set out for their selection.

4.3 Results

4.3.1 Response stereotypy

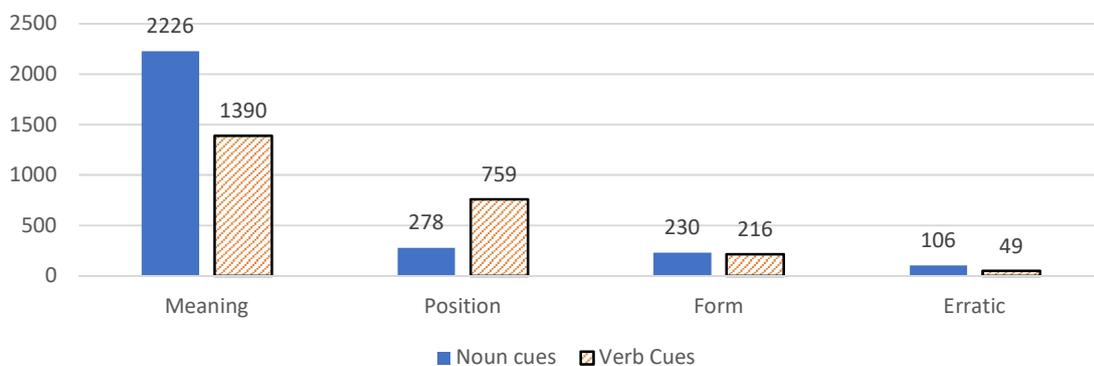
In order to determine whether the noun and verb cues resulted in significantly different patterns of stereotypy, independent samples t-tests were conducted for each of the four stereotypy variables (AFPRIM, signifying the frequency of the primary response; $N=1$, signifying the number of response types given by exactly one respondent; $N>1$, signifying the number of response types given by more than one respondent; and $N=1 + N>1$, signifying the total number of response types) with grammatical class as the independent variable. Data in all four groups was approximately normally distributed. These tests

revealed no significant differences between the noun and verb cues for any of the four stereotype measures (AFPRIM $t(72) = -.081, p=.935$; $N=1 t(72) = -.493, p=.623$; $N>1 t(72) = -.724, p=.471$; $N=1 + N>1 t(72) = -.319, p=.751$). These results suggest that the grammatical class of cues does not systematically influence the strength of association between cue and response, or any other aspect of response distribution.

4.3.2 Basic-level response type

The second research question asked whether participants produce different categories of responses to noun and verb cues. In order to investigate this, a chi-square test of independence was conducted between grammatical class and basic-level response codings. In all of the tests which follow, expected cell frequencies were higher than five. The test revealed a significant association between grammatical class and response type ($\chi^2(3) = 405.91, p < .001$). The association was moderately strong (Cramer's $V = .278$; Cohen, 1988). Figure 4.1 shows the response type counts in more detail. A glance at this data suggests that it is the differences in meaning- and position-based responses to noun and verb cues that explain this significant association, rather than differences in form and erratic responses. In order to test this, a post-

Figure 4.1
Total response counts of each basic type for noun and verb cues



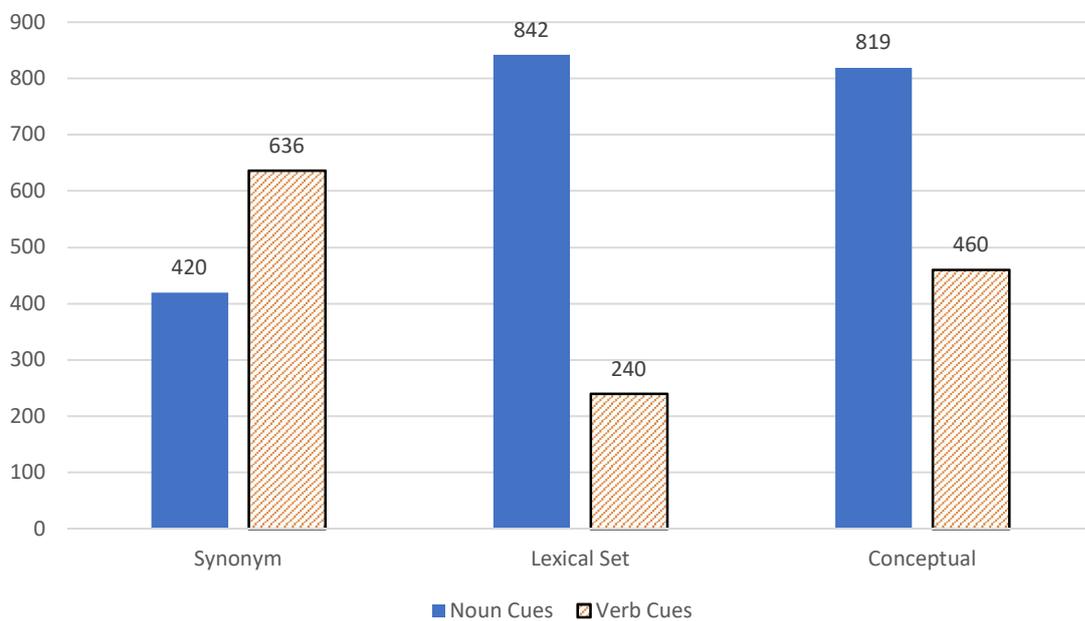
hoc chi-square test was conducted comparing responses from only the meaning and position-based categories. This test again revealed a significant association ($\chi^2(1) = 391.58, p < .001$). The association strength was similar to that in the previous test (Cramer's $V = .29$). These results suggest a clear distinction in the way noun and verb cues are responded to, in line with previous findings of differences in proportions of paradigmatic/syntagmatic responses (e.g. Deese, 1962; Nissen & Henriksen, 2006).

4.3.3 Detailed response types

The third question was a more general one, asking what a more detailed coding of responses (Fitzpatrick, 2007) could reveal about the nature of the differences in response types to noun and verb cues. In order to investigate these differences, chi-square tests of independence were conducted to test for grammatical class effects on the detailed subcategories of the meaning- and position-based categories. For the meaning-based category, a chi-square test compared response counts between grammatical class and synonym, lexical set, and conceptual responses. The test revealed a significant association ($\chi^2(2) = 333.30, p < .001$). The strength of the association was moderate (Cramer's $V = .312$). As Figure 4.2 illustrates, this significant finding was based on higher synonym responses to verbs, along with higher lexical set and conceptual responses to nouns. These results are addressed in more detail in the discussion. A further chi-square test was conducted to analyse differences in response patterns to noun and verb cues within the position-based category. This test again revealed a significant association between the grammatical class of the cue and the response type ($\chi^2(2) = 170.93, p < .001$). The association strength

Figure 4.3

Total response counts of each detailed meaning-based category to noun and verb cues.



was moderate (Cramer's $V = .418$). The most striking aspect of the distribution of these findings (see Figure 4.3) was the very large number of cue-response associations to verbs. While the noun responses did not

display any marked directionality, responses to verb cues were very heavily biased toward cue-response associations.

One explanation of this finding is that there may have been effects related to the transitivity of the verb cues. Transitive verbs might be expected to yield more cue-response associations, since they require an object noun. Intransitive verbs, on the other hand, might be expected to result in the opposite pattern. This hypothesis is supported by the case of *backfire*, which was the only intransitive verb in the dataset. This cue yielded the highest number of response-cue associations in the entire study (7 response-cue types, totalling 20 response tokens – a quarter of all response-cue associations to verb cues). This suggests that a more structured analysis of the effects of transitivity might be warranted.

Figure 4.5
Total response counts of each detailed position-based category to noun and verb cues.

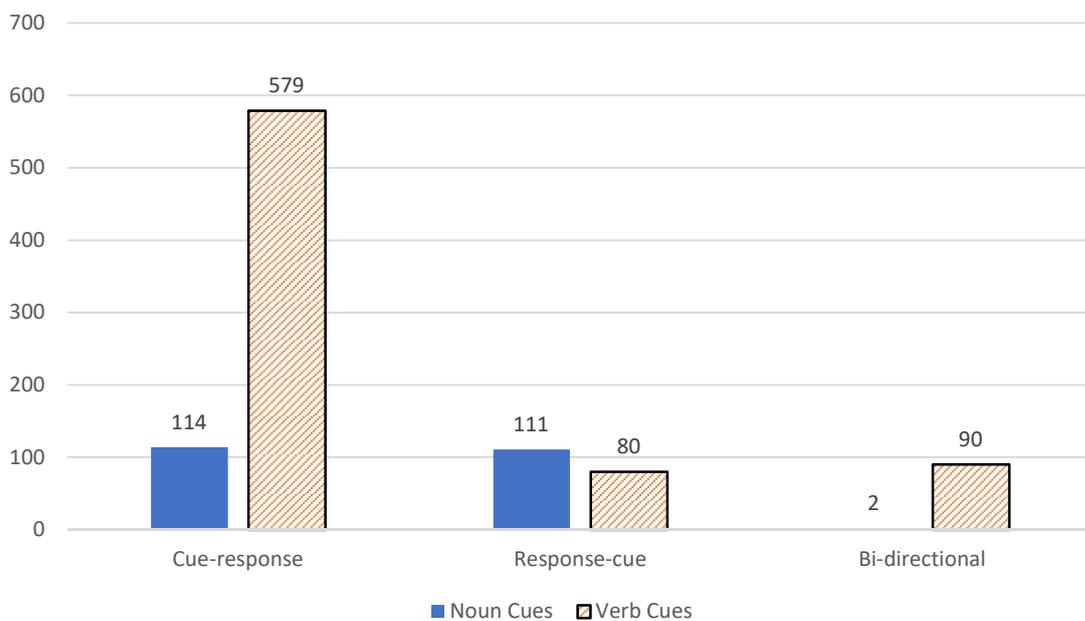


Table 4.3

Scheme for categorization of responses (adapted from Fitzpatrick, Playfoot, Wray, and Wright, 2015).

Basic category	Detailed category	Definition	Examples
Meaning-based	Synonym	Cue and response are synonymous in some situations	<i>detach-separate, wallow-bathe, excursion-trip</i>
	Lexical Set	Cue and response share a hyponym, or one word in the pair is an example of the other; includes antonyms	<i>contaminate-rot, autopilot-driver, crockery-fork</i>
	Other conceptual	Cue and response are related in meaning, but are not synonyms or in the same lexical set	<i>infancy-innocence, ancestry-age, exterminate-daleks</i>
Position-based	Cue-response collocation	Cue is followed by the response in common usage; includes compound nouns	<i>rookie-cop, migrate-overseas, voodoo-priest</i>
	Response-cue collocation	Cue is preceded by the response in common usage; includes compound nouns	<i>medley-pop, culprit-caught, knighthood-bestow</i>
	Bi-directional	Cue could precede or follow the response in a common phrase(s)	<i>circulate-paper, detonate-bombs, nightlife-good</i>
Form-based	Affix manipulation	Cue is the response with the addition, deletion or changing of an affix	<i>propel-propeller, nutrition-malnourish, migrate-emigrate</i>
	Similar in form only	Cue and response are similar in orthography and/or phonology but do not share meaning	<i>entity-entry, buffoon-baboon, wallow-willow</i>
	Two-step	Cue and response appear linked only through another word	<i>layman-bed (via lay), hover-vacuum (via hoover), wallow-tree (via willow)</i>
Erratic	Erratic	The link between cue and response seems illogical. Includes repetition of the cue	<i>stifle-quite, facet-button, culprit-silkworm, medley-say</i>
Dual-category	Lexical set <i>and</i> cue-response collocation		<i>autopilot-mode, symptom-disease, repel-attract</i>
	Lexical set <i>and</i> response-cue collocation		<i>crockery-china, commotion-noise, moron-stupid</i>
	Synonym <i>and</i> cue-response collocation		<i>voodoo-magic</i>
	Synonym <i>and</i> response-cue collocation		<i>ancestry-family,</i>

4.4 Discussion

4.4.1 Do noun and verb cues differ in terms of the stereotypy of the responses which they elicit, when key variables are controlled?

The heterogeneity of responses suggested that the effects of grammatical class on response stereotypy were minimal. No significant differences between nouns and verbs were found for any of the four stereotypy variables in isolation (AFPRIM, $N+1$, $N=1$, $N+1 + N=1$). As Chapter 3 discussed, response distribution measures provide an indication of the strength of association between a cue and its primary response, as well as the number of words to which a given cue is associated (although continuous WA methods tend to provide more exhaustive measures of this variable). The results above suggest that grammatical class is not a strong determinant of either the associative strength of the most common response type to each cue, nor of associative heterogeneity, at least in the case of noun and verb cues. This finding is in line with previous findings (de Groot, 1989; van Hell & De Groot, 1998) which suggest that imageability/concreteness is a stronger determinant of response distributions than grammatical class. Van Hell and de Groot (1998), for instance, found that in an L1/L2 comparison study using three variables – concreteness, cognate status, and grammatical class – the latter was the weakest predictor of response homogeneity.

It is important to note that the above results do not provide a measure of the influence of grammatical class on network centrality, because this measure needs to be extracted from large-scale network studies which are currently not publicly available. While previous research suggests that age of acquisition may be a key factor in centrality (De Deyne, Navarro, et al., 2012; Steyvers & Tenenbaum, 2005), it remains possible that some influence of GC will also be evident here. Further research is needed to uncover this.

4.4.2 Do noun and verb cues differ in terms of the category of the responses that they elicit, when the same variables are controlled?

Previous research has found that nouns tend to yield more paradigmatic responses than verbs, while verbs elicit more syntagmatic ones (Bøyum, 2016; Deese, 1962a; Entwisle, 1966a; Nissen & Henriksen, 2006). The results above extend these findings by demonstrating that this effect still emerges after five linguistic variables - concreteness, age-of-acquisition, frequency, word length, or bigram frequency - are controlled.

Returning to the earlier discussion of Nissen and Henriksen's (2006) factors in noun-verb differences in WA (see below), the above finding has implications for factors (a) and (d).

- a) Differences in the age and nature of acquisition of the two classes;
- b) The syntactic properties of the classes. For example, they describe verbs as “relational” (2006: 403) in that their syntactic role is to relate other elements of sentences (e.g. nouns) to each other; this makes them more cognitively dependent upon other sentence constituents;
- c) The way in which they are integrated into the lexicon. For example, the differing syntactic roles of the classes implies that, for example, verbs may be stored together with the nouns to which they commonly refer;
- d) The differences in semantic relations between the classes, with nouns largely forming hierarchical relations of hypo- and hypernymy which are not shared by verbs;
- e) Differences in the extent of cognitive processing of the classes during WA response production.

Firstly, the present chapter has established that noun and verb influence on patterns of WA response categories cannot be reduced to differences in age of acquisition. This finding has important implications for interpretations of WA that suggest that differences in response categories are indicative of different degrees of word knowledge. Specifically, research in the 1960’s showed a shift from syntagmatic to paradigmatic responding through childhood (Entwisle, 1966a, 1966b; Ervin, 1961; K. Nelson, 1977). Several researchers have interpreted this finding as evidence that paradigmatic responses reflect a higher degree of knowledge, or of integration into the lexicon (Namei, 2004; Nissen & Henriksen, 2006; Söderman, 1993; Wolter, 2001, 2002). The evidence in this chapter challenges this viewpoint: both syntagmatic (position-based) and paradigmatic (meaning-based) responses occur when cues are uniform in their age of acquisition and frequency. This suggests that familiarity or extent of experience with these words is unlikely to explain the category of the responses they received. Instead, the proportion of responses of different categories received by each cue is strongly influenced by the cue’s grammatical class.

Factor (d) refers to the semantic structure of nouns and verbs (e.g. their organisation into hyponyms, hypernyms, or troponyms). The results of this chapter suggest that this remains a viable interpretation of the noun-verb response category differences. In Section 4.1.1, a potential confound with concreteness was discussed with regard to this factor. Specifically, it was suggested that the opaque hierarchical relations between abstract nouns might make these words less likely to receive meaning-based responses. However, in order to test the above hypothesis, it would be necessary to systematically vary the concreteness of the cues used in a study. The cues selected in this chapter had

a uniform concreteness of between 2.97 and 3.41 on a scale from 1 to 5, meaning that they were neither highly concrete nor highly abstract (examples of cues include *smuggle*, *prescribe*, *vigil*, and *loophole*). As such, the present chapter is not able to test the influence of concreteness on the availability of meaning-based responses. On the other hand, the results above do clearly show that, as in the case of AoA, response category continues to vary when concreteness is held constant. This demonstrates that noun and verb response patterns cannot be reduced to differences in concreteness.

4.4.3 What can be learned from the analysis of responses to nouns and verbs using detailed levels of analysis?

This chapter provides a clear justification for the use of Fitzpatrick's scheme for categorizing WA responses based on detailed semantic and position-based distinctions (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick et al., 2015): the use of this scheme revealed noun/verb differences with regard to directionality of position-based responses to verbs, as well as a tendency for verbs to yield synonym responses. These effects would have remained hidden had a less detailed scheme, such as the syntagmatic/paradigmatic distinction, been used.

4.4.3.1 Direction of position-based responses

Section 4.4.3 describes a strong tendency for the verb cues used in this study to yield cue-response, rather than response-cue, associations. This trend was entirely absent from the noun data, where the number of position-based responses in each direction was almost identical. One interpretation of this finding is that there is an influence of verb transitivity on responses. This hypothesis is in line with factors (b) and (c) of Nissen and Henriksen's interpretation of noun/verb differences, which suggest that the syntactic biases of cue words influences their storage in the lexicon. Cue transitivity is also suggested by Bøyum (2016) as an explanation for her finding that, in a Norwegian L1 study, noun cues yielded significantly more response-cue associations than did verbs. Bøyum suggested that one reason for the lack of response-cue associations to the verbs in her study may have been their transitivity (although, as in this chapter, Bøyum's data did not sufficiently distinguish transitive and intransitive cues to allow this hypothesis to be tested).

Other interpretations of the effect are possible, however. Firstly, Clark (1970) looks at word associations from a generative-transformational perspective, positing that the lexicon and the grammatical system are discrete modules in the brain. As such, syntactic knowledge does not influence the structure of the lexicon – a view which is at odds with Nissen and Henriksen’s interpretation. Instead, Clark argues that when producing syntagmatic associations, respondents apply a left-to-right rule which is derived from general language production processes. Clark’s theory would therefore explain the response-cue associations to *backfire* as a failure to identify an adequate

Table 4.4
All meaning-based primary responses to verbs, categorized by semantic relationship

Synonym		Troponym		Contrast		Different part of speech	
Cue	Response	Cue	Response	Cue	Response	Cue	Response
Depart	Leave	Electrify	Shock	Detach	Attach	Cater	Food
Exclude	Leave out	Confiscate	Take			Descend	Down
Intrude	Interrupt	Devour	Eat			Infect	Disease
Reside	Live	Dissect	Cut				
Conceal	Hide	Migrate	Move				
Shudder	Shiver	Notify	Tell				
Exterminate	Kill						
Dissolve	Disappear						

Table 4.5
All meaning-based primary responses to verbs, in their original categorizations

Synonym		Lexical set		Conceptual		Two-step	
Cue	Response	Cue	Response	Cue	Response	Cue	Response
Depart	Leave	Migrate	Move	Cater	Food	Infect	Disease
Exclude	Leave out	Detach	Attach	Descend	Down		
Intrude	Interrupt						
Reside	Live						
Conceal	Hide						
Shudder	Shiver						
Exterminate	Kill						
Dissolve	Disappear						
Notify	Tell						
Electrify	Shock						
Confiscate	Take						
Devour	Eat						
Dissect	Cut						

left-to-right association, and subsequent recourse to another transformational rule leading to the eventual response-cue result.

A second alternative interpretation is that the strength of collocation between a word and its response determines its likelihood of selection. In the case of transitive verbs, this would imply that verbs collocate more strongly with their objects than with their subjects. Precedents for this interpretation exist in the work of several researchers (e.g. Bel Enguix, Rapp, & Zock, 2014; Wettler, Rapp, & Sedlmeier, 2005) who have suggested that word associations depend critically upon collocational knowledge. However, attempts to correlate word association responses with corpus-derived collocation data have yielded mixed results (Mollin, 2009).

Serious comparisons of these two interpretations of transitivity-related response variation are impossible, however, without a far larger sample of data on transitivity in WA. As such, discussion of these competing interpretations of the response patterns detailed in this chapter will be delayed until Chapter 6.

4.4.3.2 Synonym responses to verbs

The second finding uncovered in the detailed analysis, namely the prevalence of synonym responses to verbs, went against the general pattern for nouns to yield more semantic associations. This result is, however, the same as that reported by Bøyum, who found significantly more synonym responses to verb than noun cues (2016), and supports the conclusions of Nissen and Henriksen (2006: 403), who suggested that the organisation of verbs is dominated by synonym, troponym, or contrast relationship. Table 4.4, above, suggests that all of these types of relationship were represented in the current data.

This data also reveals complications regarding the coding of verb responses. Table 4.5 demonstrates that the original coding of these words using Fitzpatrick's (2015) scheme resulted in a very different set of results to what would have emerged from the verb-specific categories suggested by Nissen and Henriksen⁸. This suggests that the granularity offered by Fitzpatrick's scheme may be curtailed in the case of verb-verb response pairs, because of the specific relationships between these words: it may be that the detailed categories set out in Table 4.3 are better suited to the coding of nouns than verbs

⁸ It should be noted here that these issues of categorization did not affect the main finding of the study – namely the basic category finding that nouns yield more meaning-based responses than verbs. This is because all of the examples given above were first placed in the basic category “meaning-based” group, before being further categorized into the detailed category groups which included synonyms and items from the same lexical set. As such, the foregoing discussion affects only the detailed category findings.

when responses are words of the same GC. Fitzpatrick's descriptions of these schemes generally do not make explicit mention of verb-specific relationships such as troponymy. For example, the description for Fitzpatrick's "Lexical Set/Context related" category (Fitzpatrick, 2007) contains the following text: "x and y same lexical set/related coordinates/meronyms/ superordinates/provide context". The choice of terms here ("coordinates", "meronyms", "superordinates") suggests noun-specific relationships. This leaves some ambiguity as to the categorization of verb-verb responses. Should troponyms such as *devour-eat* be considered synonyms, as they largely were in the present study, or do they more reasonably fit the description of members of a lexical set, in the manner of *dachshund-dog*?

Similar problems exist with the coding of adjective responses. Aitchison (2012: 124), for example, suggests that adjective relations include "ascriptive adjectives" (words which give a quality, such as *clever*, to an entity) and "pertainyms" (words which denote a "pertaining to" relationship, in the way that *physical* in *physical education* pertains to the type of education in question). The adaptation of such categories to general schemes presents significant challenges. On the other hand, a major drawback of switching to GC-specific categorizations would be to render difficult any attempt to compare response patterns across classes. Future research could therefore look into alternative manners of coding WA responses which minimizes these cross-class complexities (c.f. Guida & Lenci, 2007)

The implication of this discussion for the present study, however, is that the significant difference between nouns and verbs in the "synonym" detailed category cannot be made too strongly. This is because if the words in the "synonym/troponym" section of Table 4 had been categorized as lexical set items, rather than synonyms, the larger number of responses to verb cues in the "synonym" sub-category would not have reached significance.

4.5 Conclusion

The main aim of this chapter was to establish the independence of a previously reported grammatical class effect on word association response categories. The results above demonstrate that this effect is independent of the frequency, concreteness, age of acquisition, bigram frequency, and length in syllables of the cue. It has also expanded these findings by, firstly, finding differences in response directionality and synonym responding as a function of GC, and, secondly, finding that GC does not influence distributional aspects of responses, such as their strength of association.

The study has additionally established a precedent for WA work on the influence of cue-level variables in WA. In doing so, it offers a route for future research to emulate work on other psycholinguistic tasks such as lexical decision and word naming (e.g. Brysbaert et al., 2011; Keuleers et al., 2012; Yap, Tan, Pexman, & Hargreaves, 2011). This is important work: Chapter 3 suggested that such variables influence the structure of the lexicon itself, while findings such as those presented by Van Hell and de Groot (1998, de Groot 1989) appear to imply that robust effects from other psycholinguistic tasks, such as the frequency effect in word naming and lexical decision, cannot be assumed to also influence word association.

It is therefore important that this work now moves forward to address further influences on WA response patterns. Perhaps the most pertinent question to emerge from this study pertains to verb transitivity. As discussed in Section 4.5.3.1, this issue offers a route into discussion of several important but under-elaborated models of associative responses. Further research on verbs also permits the development of coding schemes better suited to the relationships between verbs. Chapter 5 will therefore pursue this research direction in more detail.

Chapter 5: Does verb transitivity influence word association responses?

5.1 Introduction

In Chapter 4, an original experiment explored the influence of grammatical class (GC; specifically, nouns and verbs) on WA responses. The study supported the findings of earlier work (Deese, 1962a; Entwisle, 1966a; Nissen & Henriksen, 2006) by suggesting an influence of grammatical class on response categories: nouns elicited more meaning-based responses than verbs, and verbs more position-based responses than nouns. Further to earlier studies, the experiment also established the independence of these effects from variables including concreteness, age of acquisition, and frequency. The study found no significant effect of GC on response distribution – a measure which had not previously been examined with reference to GC.

Through the use of the detailed response categories devised by Fitzpatrick (2007), the experiment also revealed a potential influence of transitivity on response patterns. The locus of the effect was the directionality of position-based responses: while verb cues in general yielded significantly more cue-response than response-cue associations, the sole intransitive cue in the dataset demonstrated the opposite pattern – it elicited seven response-cue types, compared with just one coded as cue-response. These results suggest that verb transitivity may influence the directionality of position-based responses (see section 4.4.3, above). This finding, though based on a single cue, warrants further research into influence of transitivity in word association tests.

An additional benefit of an investigation into the influence of transitivity on WA responses is that it may offer an opportunity to test different psycholinguistic models of language processing. This is because differences in response patterns as a function of transitivity would have implications for how we understand specific syntax to influence lexical storage and processing during WA response production. Psycholinguistic models would differ in how they interpreted such patterns, and some models would be unable to explain them at all. This research would therefore offer an opportunity to determine which psycholinguistic models offer the best fit for existing WA data. Before such models can usefully be explored, however, the influence of transitivity on WA needs to be established.

5.2 Defining transitivity

In a seminal article on the nature of transitivity, Hopper and Thompson (1980) advance the position that transitivity is a property not of individual verbs, but of entire clauses. They outline several features of transitive clauses, only one of which is the presence of two participants (i.e. both a grammatical subject and object). Other features include the telicity of a clause (i.e. whether the action described is

seen as complete or incomplete), its volitionality (i.e. whether the agent in the clause is acting purposefully), and the degree of affectedness of the grammatical object of the clause (in more transitive sentences, the object is strongly affected by the action of the agent). For example, Hopper & Thompson (1980: 253) suggest that the sentence *Jerry knocked Sam down* is highly transitive not only because it possesses both a grammatical subject and an object, or because the verb *knock* is “transitive”, but because the clause possesses the qualities of an action which is transferred from one participant to the other, it is telic (i.e. viewed as complete), punctual (in that no delay occurs between inception and completion), totally affects the object, and features a highly individuated object (in that the object is a referential, animate, proper noun). Hopper & Thompson contrast this example with another clause, *Susan left*, which has only one participant but, by virtue of possessing the characteristics of an action, telicity, punctuality, and volitionality, is more transitive than a sentence such as *Jerry likes beer*, the only transitive aspect of which is its two participants (*Ibid.* p254).

There are several implications of this position. Firstly, it implies that transitivity is a gradable phenomenon, not an absolute one. Clauses should therefore be seen as varying along a scale of transitivity according to how many transitivity features they display. As the authors point out, this means that “just as a clause may have an overt second participant, and still be aligned with the intransitive clause, so also it may lack a second participant, and yet have transitive features” (1980; 266). In other words, the absence of a grammatical object in a clause is not the sole determinant of the transitivity of that clause. The second implication of Hopper and Thompson’s work stems from this continuous view of transitivity: if entire clauses, rather than individual verbs, are the locus of transitivity, then it is not accurate to speak of “transitive” or “intransitive” verbs.

This creates a problem for the current study, which aims to assess the impact of the transitivity of individual verbs on WA responses. There is, however, a precedent for the negotiation of this problem in the existing psycholinguistic literature, which is to refer to the transitivity bias of individual words. This approach acknowledges the clause-level nature of transitivity, but refers to the predisposition of some verbs to appear more often in highly transitive sentences than others (e.g. DeDe, 2010, 2014; MacDonald, 1994; Osterhout, Holcomb, & Swinney, 1994; Theakston, Lieven, Pine, & Rowland, 2001).

This approach is not perfect. In particular, verb transitivity bias continues to be defined as the tendency for a given verb to take a grammatical object. This is problematic because it ignores the possibility that the biases related to other features of transitivity, such as those described above, might influence processing in psycholinguistic tasks. To return to the example of *Jerry knocked Sam down*, it is possible, for instance, that the verb *knock* tends not only to use a grammatical object, but

also to be used in telic or punctual sentences, or for its object to be highly affected or individuated. Thus, the transitivity bias of a verb, properly defined, would be far broader than the definition generally used in psycholinguistic studies. This situation persists because no systematic analysis has been made of the covariation between individual verbs and the transitivity features of their clausal contexts, making a more principled selection of verbs difficult.

Nevertheless, the selection of verbs based on the presence of a grammatical object is appropriate as a starting point for word association studies, because it is, as will be shown below, specifically the presence or absence of an object which is assumed to be an influence on WA responses. Therefore, while it is acknowledged that a future study based upon a broader definition of transitivity might itself yield valuable insights into the influence of clause-level factors on WA, the present study (as was the case of those reviewed below) will define cue transitivity as a strong bias towards the appearance or non-appearance of a grammatical object in the company of the verb in question.

5.3 Previous research on transitivity and lexical processing

Recent psycholinguistic research has uncovered effects of verb transitivity bias, particularly in the field of sentence parsing. Debate in this area has centred around whether or not a verb's transitivity bias affects the parser's attempts to attach noun phrases to the verbs which precede them. For instance, Mitchell (1987) suggested that the parser ignores the transitivity bias of a verb when creating interpretations of sentences. He suggests that, in sentences such as "after the child had sneezed the doctor prescribed a course of injections", the parser attempts to attach the noun phrase "the doctor" to "sneezed" as its direct object, in spite of the fact that "sneeze" is an intransitively biased verb. Staub (2007), however, challenged these findings and, in a series of experiments, concluded that transitivity bias *does* influence the parsing of sentences such as these: the parser does not attempt to attach noun phrases to verbs found mostly in intransitive constructions, unless ambiguity in the rest of the sentence necessitates a reanalysis. Mitchell's earlier results to the contrary were explained as a methodological confound. Staub interpreted these results as suggesting that the parser makes use of all available information while generating interpretations of sentences, including information regarding a verb's transitivity bias.

The Lexical Bias Hypothesis (Gahl, 2002) offers a theoretical interpretation of these results. The hypothesis posits that statistical regularities in a language, such as the likelihood of a given verb appearing in intransitive sentences, are an influence on language storage and processing. Thus, probabilistic knowledge of a verb's transitivity is likely to be one source of information available during sentence processing. In support of this hypothesis, DeDe (2014), for example, asked groups of aphasic

and healthy control participants⁹ to read sentences in which the transitivity bias of verbs was systematically varied to either match (e.g. “the agent called the writer” – a transitively biased verb in a transitive sentence) or not match (e.g. “the agent called from overseas” – transitive verb, intransitive sentence) the sentence’s transitivity. DeDe found that both groups of participants took longer to read the mismatched sentences than the matching ones, even though all sentences were equally grammatically acceptable. This suggests that participants’ expectations about a verb’s argument structure influence the way they parse sentences containing that verb.

While these studies suggest an influence of a verb’s transitivity bias on sentence processing, it does not necessarily follow that methods such as word association elicitation, which require participants to respond to single words rather than whole sentences, will also show such effects. In fact, the influence of transitivity has attracted only brief attention in previous word association research, with only a handful of studies explicitly addressing this aspect of language. None of the earliest of these studies sought to examine the influence of transitivity on WA responses directly, but instead were attempts to explain the widely-observed shift from syntagmatic to paradigmatic responses in children’s WA behaviour (Woodrow & Lowell, 1916). One such study was Brown and Berko (1960). These authors hypothesised that the syntagmatic-paradigmatic shift stemmed from the emergence of syntactic classes in the linguistic system of young children. Young children depend, for their associations, on contiguity of words in continuous speech. For example, children associate *dark* and *night* because of the frequency of co-occurrence between the two words. However, as they mature, they are increasingly aware that words such as *light* and *bright* can all fill sentence frames such as “The sky is _____”. This awareness leads to the emergence of grammatical class distinctions in the minds of children, and these distinctions then become the foundation for word association.

In order to test this theory, Brown and Berko asked four groups of 20 participants (1st, 2nd, and 3rd grade primary school students, and adults) to provide word association responses to cues from a range of grammatical classes, including both transitively and intransitively biased verbs. The authors found that both age and grammatical class were significant determinants of participants’ tendency to provide “homogeneous” responses (i.e. those of the same grammatical class), with older children and

⁹ The reason for the use of both participant groups is that DeDe wished to test the hypothesis that sentence processing in aphasics is inhibited by the same factors as in healthy participants – namely difficulty accessing syntactic information for words which appear in contexts which do not match their transitivity bias. DeDe thus argued that the difference between the two groups is one of extent, not of qualitative difference. The study was interpreted as supporting this hypothesis, since the nature of the effects found in the study was the same for both groups.

adults producing more homogeneous responses than did younger ones. They interpreted their results as supporting their earlier hypothesis that the syntagmatic-paradigmatic shift could be explained in terms of the “gradual organization of [a child’s] vocabulary into the syntactic classes called parts of speech” (Brown & Berko, 1960: 13). They further suggested that grammatical class effects could be interpreted in terms of the stage of acquisition of each individual class. In other words, those classes which had been most thoroughly acquired would yield a higher proportion of homogeneous responses, while a low proportion of homogeneous responses to a given class could be taken as evidence of the incomplete acquisition of that class. Differences in proportions of homogeneous responses given by adults to different grammatical classes were interpreted, in line with this theory, as being the result of the greater frequency of words from these grammatical classes than those from others, such as intransitive verbs or adverbs. For Brown and Berko, then, it is not simply the frequency of a given word which influences the category of response given, but the frequency of occurrence of its grammatical class as a whole, since it is this which determines the speed with which a concept of each grammatical class emerges.

With regard to the influence of transitivity, Brown and Berko found that participants from all age groups gave slightly more homogeneous responses to intransitive verbs than to transitive ones. This finding is difficult to interpret, however, since the authors did not attempt statistical comparisons of these categories. In addition, the cue list included only six verbs of each sub-type – several of which (e.g. *to stand*, *to walk*) are not strongly intransitively biased, because although their primary usage is likely to be in intransitive sentences, transitive usages (or independent senses) are also possible (e.g. *I can’t stand John*; *You’re walking a fine line*). For these reasons, although Brown and Berko’s study hints at increasing verb responses to intransitive cues, it cannot be taken to provide evidence for an effect of transitivity on word association.

Another paper, by Ervin (1961), offered a similar theory of the syntagmatic-paradigmatic shift, but one with different implications for the understanding of transitivity effects. Like Brown and Berko, Ervin suggested that a maturing child’s exposure to an increasingly wide variety of words and sentence frames influenced the structure of the lexicon. However, Ervin elaborated this theory more fully than Brown and Berko. She suggested that in the early stages of language development, associative relations between words are formed through perception of forward contiguity in text. This means that if A frequently precedes B in text, then A would likely elicit B as a WA response. Importantly, Ervin asserted that backward association is dispreferred; therefore, B should only rarely elicit A in a test of association. In addition, Ervin (1961: 368) suggested that “functional words rarely occur in isolation

and will, therefore, be improbable as associative responses". This can be interpreted to mean that respondents prefer not to produce these words as responses because they sound or appear unfamiliar to respondents when produced in isolation; Ervin does not mean to suggest that words occurring in isolation (such as *yes*) are particularly likely to appear as WA responses.

The process guiding the emergence of paradigmatic responses in Ervin's theory is not the gradual emergence of grammatical class, as in Brown and Berko (1960), but is rather the result of a widening perception of linguistic choices. Children become increasingly aware that a sentence such as *Would you like a cup of...* can be completed not only by *milk*, as a young child might assume, but also by *water*, *tea*, and *coffee*. These words therefore become associated through a principle which Ervin describes as "contiguity during competition" (Ervin, 1961: 362), "on the basis of occurrence in the same preceding verbal contexts" (Ervin, 1961: 372). That is, as children encounter and produce sentences (e.g. *Would you like a cup of...*), alternatives for selection arise (*milk*, *water*, *tea*, *coffee*) and become associated as a result of their competition for selection. Thus, Ervin saw both syntagmatic and paradigmatic responses as deriving from contiguity; but the contiguity in question was textual in the case of syntagmatic responses, and cognitive in the case of paradigmatic ones.

In Ervin's theory, intransitive verbs are predicted to elicit a greater number of paradigmatic responses than do transitive verbs. This prediction rests on two assumptions. Firstly, since forward association is a key determinant of WA responses, transitive verbs should become associated with their grammatical object. However, since intransitives do not have an object, and also because of their high frequency of sentence-final occurrence, they are unlikely to develop forward associations, and thus less likely to yield syntagmatic associations, than are transitive cues. Secondly, given that backward association is not considered to be an important determinant of WA responses in Ervin's theory, it should not be the case that intransitives become associated with their grammatical subject. Therefore, they will become dominantly associated with their rivals for selection during sentence processing, and thus yield paradigmatic responses.

Unfortunately, while Ervin interpreted her study as supporting this hypothesis, the experiment suffered from similar problems to that of Brown and Berko: it used only six verbs, and just one of these (*come*) was classified, in the study, as unambiguously intransitive. As such, no inferential investigation of the influence of transitivity was possible. The descriptive statistics presented by Ervin in fact showed similar levels of syntagmatic and paradigmatic responses to *come* as were found for the study's transitive cues.

Brown and Berko (1960) and Ervin's (1961) theories are somewhat similar to modern usage-based views of language development (Behrens, 2009; Bybee, 2010; Bybee & Beckner, 2010; N. C. Ellis, Romer, & O'Donnell, 2016; Gerola, 2016; Lieven & Tomasello, 2008; Tomasello, 2003) in that they stress the influence of language exposure and usage on lexical storage and organisation. These ideas will be returned to in the following chapter. An alternative position is taken in a later study by Polzella and Rohrman (1970), who derived a theory of transitivity based on generative grammar. Chomsky (1965) hypothesised that a verb's argument structure is stored as a part of its entry in the mental lexicon. For example, the entry for a transitive verb would specify a subject preceding the verb, and an object following it. Lexical entries for intransitive verbs, on the other hand, would specify a subject but no object. Polzella and Rohrman suggested that this would lead to a greater number of noun responses to transitive verbs since, when presented in isolation (i.e. without their direct objects), transitive verbs are "incomplete cognitive objects" (Polzella & Rohrman, 1970; 538). Respondents would therefore attempt to "complete" the transitive verbs by producing their grammatical object. Transitives were, as such, hypothesised to have a closer relationship to noun phrases than intransitives have. Intransitives, not having this close relationship with nouns, would yield verb responses.

The authors tested this theory by asking 34 adult participants to produce spoken responses to 28 verb cues (equally divided between transitive and intransitive). The dependent variables in the study were the grammatical class of responses, and their response time latencies. The results supported Polzella and Rohrman's theory; transitive verbs yielded significantly more noun responses than intransitives, while intransitives received significantly more verb responses (note that this finding also supports Brown and Berko's finding of more homogeneous responses to intransitives than transitives, above). In addition, the noun responses to transitive cues were produced more quickly than noun responses to intransitives. These findings support the view that the different argument structures fundamental to transitive and intransitive verbs are represented at the lexical level, and therefore influence lexical processing. It is worth noting, however, that Polzella and Rohrman's results can also be explained by models which propose that knowledge of a verb's syntactic biases, rather than generative rules, account for transitivity effects in psycholinguistic tasks (DeDe, 2014; Gahl, 2002; Staub, 2007). One such model is the Lexical Bias Hypothesis (Gahl, 2002), which proposes that statistical regularities are noticed and made use of during language production.

One caveat regarding Polzella and Rohrman's study is that the cues used in the study were not provided; the only description given was that they were controlled for frequency and were equally meaningful, as tested through a continuous WA task. As such, it is not possible to assess the extent of

their transitivity biases. As with the other studies reported above, then, their results must be interpreted with caution.

In summary, then, there appears to be a tendency in WA for transitive verbs to elicit a greater number of noun (i.e. syntagmatic, heterogeneous) responses than do intransitives. This basic pattern was found by all of the above studies. However, each study offered a different interpretation. Brown and Berko (1960) suggested that the gradual, unequal emergence of grammatical categories in the lexicon explain all GC differences in WA; Ervin's (1961) theory suggests that learning forward associations through textual contiguity (and the absence of such contiguities for intransitive verbs) accounts for the results; and Polzella and Rohrman (1970) propose that differences in the lexical entry for transitive and intransitive verbs are the reason for their different response patterns. Specifically, the entry for transitive verbs specifies a noun phrase in the object position (which the entry for intransitives does not); this leads to respondents producing nouns in response to transitive verbs. Further evidence for this view comes from the finding that noun responses are faster for transitive cues than for intransitive ones, while verb responses are faster for intransitives than for transitives (Polzella & Rohrman, 1970). Finally, it was noted that all of the studies above suffered from technical problems such as insufficient cues and a lack of detailed information about cue selection.

5.4 The current study

5.4.1 Predictions

The above studies offer a set of predictions for the investigation to be reported in this chapter:

Prediction 1: Transitive verbs will yield more noun responses than intransitive verbs.

Prediction 2: Intransitive verbs will yield more verb responses than transitive verbs.

These predictions are based on the findings of all three WA studies discussed above. As already noted, two of these studies (Ervin, 1961; Polzella & Rohrman, 1970) suggested (though for varying reasons) that the provision of noun responses to transitives was related to syntax: transitive verbs require a grammatical object, while intransitives do not. However, an alternative hypothesis might state that intransitive verbs, in the absence of a need for a direct object, will receive responses intended as *subject* completions. In this case, they will also receive noun responses. This possibility is evidenced by the fact that, in all of the studies cited above, noun responses still occurred to intransitive cues, even though in smaller numbers than to transitives.

Fitzpatrick's detailed coding scheme (2006, 2007), also used in the previous chapter, offers a way to investigate this. Assuming that transitively biased verbs are likely to elicit a direct object, they should

be more likely than intransitively biased verbs to elicit responses coded as cue-response, as this category would capture verb-object relationships. Intransitively biased verbs, on the other hand, may yield responses coded in the opposite direction (i.e. response-cue) signifying (perhaps amongst other things) subject-verb relationships. As such:

Prediction 3: Transitive verbs will yield more cue-response associations than intransitive verbs.

Prediction 4: Intransitive verbs will yield more response-cue associations than transitive verbs.

5.4.2 Method

5.4.2.1 Participants

The participants in this study were 53 adults, all of whom were students at a UK university. All identified English as their first language.

5.4.2.2 Cue word selection

The cue words were 50 transitively biased and 50 intransitively biased verbs. The intransitive cues were selected from an online list of verbs¹⁰. Data on the concreteness (Brysbaert et al., 2014), frequency, and contextual diversity (CD; van Heuven et al., 2014), and age of acquisition (AOA; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) of these verbs was collected, and several steps were taken to ensure that each verb was exclusively intransitive (see below). Transitive cues were then selected through a process of pair-matching verbs with very similar values on the above variables. Because of the difficulty of identifying sufficient intransitive verbs, no other variables were controlled. It should be noted that this approach differs from that taken in the previous chapter, where all cues were taken from a predetermined range. In the present study, the pair-matching control method means that wide ranges of frequency, CD, AoA and concreteness were used.

Additional steps were undertaken in order to address two potential problems related to grammatical class membership. First, many English verbs can also function as nouns (or other grammatical classes). In order to reduce the possibility of influence from other classes, all cues were checked against the Longman Dictionary of Contemporary English (Longman, 2014) to ensure that the primary usage of each cue was as a verb, rather than another grammatical class. Secondly, few English verbs occur exclusively in transitive or intransitive phrases. Frequent verbs, in particular, can appear in both transitive and intransitive sentences, often with slightly different meanings. Thus, the transitivity bias of each cue was further checked (using the same dictionary) to make sure that all of its listed meanings

¹⁰ Available at https://en.wiktionary.org/wiki/Category:English_intransitive_verbs.

corresponded to the appropriate verb type (i.e. cues selected as transitive had only transitive meanings, and intransitive cues only intransitive meanings). This process resulted in the elimination of many potential cues. In particular, the final cue list contains few cues from lower frequency levels.

The second level of this preparatory analysis revealed that numerous intransitive cues were listed as “transitive with prepositions” (Longman, 2014). This was problematic for the present study because the tendency for a verb to appear along with a preposition in a phrasal construction could interfere with the influence of transitivity itself. If, as suggested above, intransitive verbs are less likely to yield cue-response associations because they do not require a direct object, then frequent occurrence in transitive phrasal constructions could interfere with this effect, either as a result of the production of potential direct objects of a phrasal construction, or through the production of adverbs or prepositions used in the phrasal expression. In order to control for this possibility, mutual information scores were calculated between each cue and any preposition with which it commonly occurred. Any potential cues which shared an MI score of 4 or above with any preposition were excluded from the cue list. This eliminated cues such as *apologize* (*for*: MI = 6.48), *depend* (*on/upon*: MI > 15), and *belong* (*to*: MI > 20). A stricter MI criterion would have been desirable here, but would have resulted in the cue list dropping below 100 items. Of those items which remained, *wince* (*at*: MI = 3.84) and *inquire* (*into*: MI = 3.80) were the only items which approached the selected threshold. All MI scores were taken from the TenTen English corpus hosted by Sketch Engine (Jakubiček, Kilgariff, & Kovář, 2013; Kilgariff et al., 2014). A summary of cue characteristics is provided in Table 5.1. As the table shows, mean, minimum, and maximum values are very similar for all variables, and standard deviations are similarly low. This suggests that the two groups did not markedly differ. T-tests also revealed no significant differences between the transitive and intransitive verbs.

Table 5.1
t-tests and Min/Max statistics for transitive and intransitive cues.

	Transitive				Intransitive				t-test
	Mean	SD	Min	Max	Mean	SD	Min	Max	<i>p</i>
Frequency (Zipf)	3.29	.78	1.73	4.94	3.15	.81	1.59	5.32	.910
Contextual diversity	.03	.06	.01	.265	0.3	.08	.01	.15	.681
Concreteness	3.12	.74	1.83	4.52	3.15	.73	1.78	4.52	.939
Age of acquisition	8.88	2.01	4.52	14.13	8.49	2.30	4.32	14.13	.405

5.4.2.3 Materials

The cue words are presented in Appendix 3. The order of the cues was randomized, with transitive and intransitive cues mixed to limit the possibility of practice effects. Each cue was assigned a number

between 1 and 100. A random number generator was then used to select the order of the cues¹¹. These were given to participants on 2 double-sided pieces of A4 paper. Respondents were asked to “Please write the first word that comes to mind for each of the following words”. They were informed that there were no correct answers to the test, and were also asked not to change their answers once they had been written. No time limit was given for providing answers. All participants completed the list within 15 minutes.

5.4.3 Categorization of data

62 responses (from a possible total of 5300) were left blank, leaving a total of 5238. Responses were analysed in three ways. Firstly, in order to investigate predictions 1 and 2, regarding the possible influence of transitivity on the grammatical class of responses, the primary grammatical class of each response was collected from the SUBTLEX_UK list (van Heuven et al., 2014).

Secondly, stereotypy measures were calculated for each cue. While numerous researchers have used measures of stereotypy in their studies (e.g. Playfoot & Izura, 2013), the specific measures used in the present study, as in the previous chapter, were the same as those used by de Groot (1989):

1. The “associative frequency of the primary response” (AFPRIM) – i.e. how many individual participants produced the most common response to each cue.
2. The number of unique responses produced by exactly one person, to each cue ($N=1$)
3. The number of unique responses given by more than one person, to each cue ($N>1$)
4. The total number of responses to each cue (i.e. the sum of values 2 and 3, above; $N>1 + N=1$).

Finally, each cue-response pair was categorized according to an adapted version of the Fitzpatrick (Fitzpatrick, 2007, 2009; Fitzpatrick et al., 2015) scheme used in the previous chapter. In view of the discussion in Section 4.5.3.2, which revealed difficulties in the categorization of verb-verb associations, it was deemed necessary to change some categories in order to have them better reflect the types of relationships which exist between verbs. As noted in that Chapter, Fitzpatrick’s existing scheme may have an inadvertent bias towards noun-noun responses (e.g. categories such as “Lexical Set” and descriptors such as “hyponym” and “meronym”, which refer less comfortably to verbs than nouns; see also Aitchison, 2012: 124).

In order to address this problem, three changes were made. Firstly, the “Lexical Set” category was split into “Troponym”, representing verb-verb associations which denote different ways of performing the same basic actions (e.g. *stroll* is a way of *walking*), and “Contrast”, which aimed to capture oppositional relationships (*sit* and *stand*). Secondly, the “Conceptual” category was split into “Other

¹¹ Available at <https://www.random.org/lists/>

Conceptual – same part of speech” and “Other Conceptual – different part of speech”. The “different part of speech” conceptual category allowed for the coding of non-verb responses which were not thought likely to be position- or form-based. Finally, a “Meaning- and Form-Based” dual-link category was added to address the absence of this category in previous coding schemes. The full coding scheme is reproduced in Appendix 4.

As in the previous chapter, each verb was first categorized according to the four basic level categories, and then into detailed groups. All responses were coded separately by two coders, as described in Section 4.2.4.1. Likewise, in cases of disagreement the second coder’s view was preferred, in order to reduce any potential coding bias on the part of the first coder (i.e. the author). Initial agreement was 77.2% overall. This is lower than the level of agreement achieved in the previous study (basic level 84.1 vs 91.1, detailed level 70.4 vs 81.3), but is nevertheless similar to that reported in Fitzpatrick et al. (2015: 76.9%).

An analysis of the responses which provoked the greatest disagreement revealed that those categories which involved combining two existing categories into a single one (i.e. categories 8 and 13-19, in Appendix 4) received very low levels of inter-rater agreement. Of 53 responses placed by both coders firstly into the position-based basic category, and subsequently by at least one coder into the “Bi-directional” detailed category, only 8 (15%) were placed in this latter category by *both* coders. In the case of categories 13-19, 52 total responses were placed into one of these detailed categories by at least one coder (irrespective of the basic category), but the coders initially agreed on only 10 of them. Taken together, pairs placed by at least one coder into categories 8 or 13-19 accounted for less than 6% of all cue-response pairs, but more than 20% of all coding disagreements.

Dual-link responses were first used in Fitzpatrick & Izura (2011) in order to address concerns over cues which could conceivably be categorized as one of two different response types. Fitzpatrick and Izura (*ibid.*; 377) gave the example of *black-white* as a cue-response pair which could equally well be understood as either position- or meaning-based. Another example was *pen-pencil*, which share both form-, position-, and meaning-based relationships. Fitzpatrick and Izura went on to demonstrate that responses categorized as dual-link were produced faster than those categorized as either equivalent or non-equivalent meaning. This may suggest a psycholinguistically real, potentially cumulative, effect of more than one type of relationship between cue and response. However, given the uncertainty which these categories appear to have introduced into coding, it may be helpful to establish clearer guidelines on their use than in the coding scheme developed for this study. This issue will be returned to in Chapter 7.

Following initial coding, the two coders reached agreement on all remaining cue-response pairs through discussion. In order to counteract the possibility of bias in these decisions, the second coder was not made aware of the study’s research questions. To further reduce bias, where initial agreement between coders was difficult to reach, the second coder’s decision was chosen.

5.4.3.1 Removal of cues

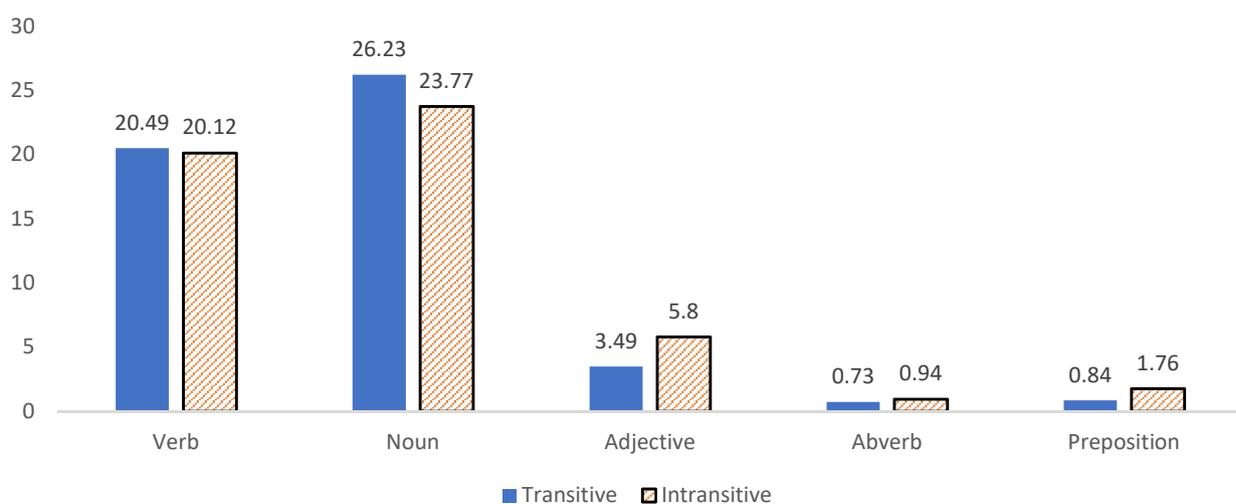
During the processing of responses, it emerged that one transitive cue, *afford*, had been duplicated in the cue list. The second iteration of *afford* was therefore removed from analysis along with its intransitive pair, *remain*. This left a dataset of 98 cues.

5.5 Results

5.5.1 Grammatical class of responses

Predictions 1 and 2 suggested that transitivity status would influence the grammatical class of responses, with transitives expected to yield a greater number of noun responses than verbs. To provide an initial test for this hypothesis, a chi-square test of independence was used to test whether the grammatical class of responses was independent from the transitivity of cues. All expected cell frequencies were greater than five. The test showed an association between transitivity and grammatical class of responses ($\chi^2(11) = 69.09, p < .001$). The size of the association was small (Cramer’s $V = .115$).

Figure 5.1
Mean responses of four grammatical classes to transitive and intransitive cues. Note that all proper nouns were included in the “noun” column.



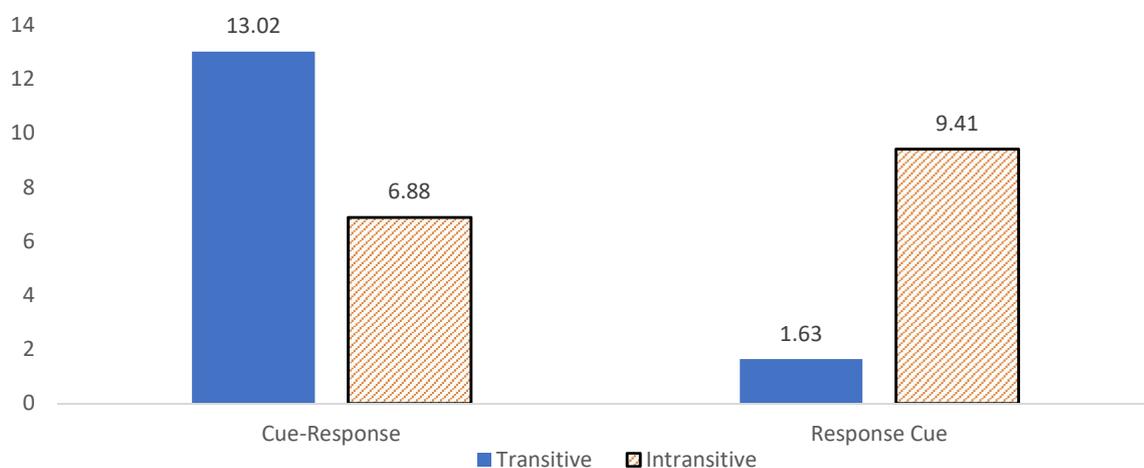
However, while the above test demonstrates that the transitivity of cues significantly influenced the grammatical class of responses, it does not test the specific hypothesis that transitive cues would receive more noun responses, and intransitives more verb ones. In order to do this, a planned comparison was conducted using only noun and verb responses. This test revealed no significant association between cue transitivity and the grammatical class of responses ($\chi^2(1) = 1.474, p = .225$), although the general effect was in the direction predicted. One problem with the above analysis, however, is that the categorization of grammatical class (i.e. that of van Heuven et al., 2014) includes a separate categories for proper nouns (examples from the present data included *Buddha*, *Rugrats*, and *Taylor*). In order to test for the effect of this, these words were re-categorised as nouns, and a further chi-square test was conducted. This test again revealed no significant effects of transitivity on the grammatical class of responses ($\chi^2(1) = 1.753, p = .186$). Thus, while the direction of the effect of transitivity on noun responses was in the predicted direction, it was not found to be significant. Prediction 2, which suggested that intransitive verbs would receive more verb responses than did transitives, was not supported. The initial significant difference in the grammatical class of responses was likely caused by the accumulation of differences across the noun, adjective, and preposition classes, as illustrated by Figure 5.1.

5.5.2 Response type

Predictions 3 and 4 state that transitive cues will receive more cue-response associations, while intransitives will receive more response-cue ones. These hypotheses were tested using a chi-square test of independence between position-based associations (cue-response and response cue) and transitivity. Expected cell frequencies were all above five. The test revealed a significant association

Figure 5.2

Mean responses to transitive and intransitive cues categorised as cue-response and response-cue.



between cue transitivity and directionality of position-based coding ($\chi^2(1) = 358.02, p < .001$). The strength of the association was large (Cramer's $V = .486$; Cohen, 1988). Transitive cues yielded more responses categorised as cue-response, while intransitives elicited more response-cue associations. Figure 5.2 illustrates this.

These results offer initial support for Predictions 3 and 4. However, two potentially confounding issues which emerged during cue selection require further investigation. The first of these, discussed in Section 5.4.2.2, is the potential for some cues to fulfil both noun and verb roles, or to form phrasal verb constructions with an adverb or preposition. In order to investigate the influence of this factor, during response coding the two coders highlighted:

- a. Any response which could not be explained as a response to the cue in its verb form, but which could be explained as a response to the cue in another grammatical class. Examples of this included examples where the cue had been taken as a name (*chuck-Taylor/Bass/Berry/Norris; stitch-Lilo; waddle-Chris*), as an adjective (*limp-floppy/hair/biscuit/wrist*), as a noun homograph (*bark-tree/wood*), or as a compound noun (*glow-stick; snooze-button*).
- b. Any response which formed a phrasal verb with the cue (*rise-up/again/above; inquire-about*).

Final agreement from the coders resulted in the removal of 200 tokens of 26 association pair types. The above chi-square test of independence was then re-run. The significant interaction between transitivity and response directionality remained ($\chi^2(1) = 434.831, p < .001$); the effect size was slightly larger than in the initial test (Cramer's $V = .568$). Thus the significant interaction of transitivity and directionality of position-based responses cannot be explained by the tendency of a small number of cues to be interpreted as nouns, or to yield phrasal verb completions.

The second potentially confounding variable was the identification of a number of cues which were thought likely to have limited response profiles due to their narrow semantic range. Post-hoc analysis of all cues identified five cues (four intransitive – *bark, purr, roar, and erupt*, and one transitive – *extinguish*) for which the associative frequency of the primary response (AFPRIM: i.e. the number of times the most common response was produced) was at least two standard deviations above the mean. These cues, along with their matched pairs (*crush, underline, bury, furnish, and flinch*, respectively) were removed from the data, and a new chi-square test of independence was run. This test again revealed the significant association between transitivity and directionality of position-based responses ($\chi^2(1) = 140.86, p < .001$). There was a somewhat diminished effect size here (Cramer's V

= .345), suggesting that some of the statistical power of the earlier tests originated from cues which were strongly associated with a single position-based response. However, the general effect remained.

Finally, when all potentially confounding factors (responses to noun, phrasal verb responses, and semantically narrow cues) were removed at the same time, the significant effect of transitivity remained ($\chi^2(1) = 195.28, p < .001$; Cramer's $V = .438$).

The above results offer counterpoint to those offered by Polzella and Rohrman (1970), who found that transitives elicit more noun responses than intransitives. The present results suggest that while the two verb types are equally likely to yield noun responses, there may be qualitative differences in the specific nature of those responses. A further chi-square test, identical to the above except that only noun responses were analysed, bears out this conclusion. The test revealed a significant association between transitivity and directionality of position-based associations ($\chi^2(1) = 504.81, p < .001$; Cramer's $V = .666$). Note that the effect size here is larger than the above analyses, suggesting that the general trend for transitives to yield cue-response associations and intransitives to yield response-cue ones was stronger among noun than non-noun position-based responses.

Taken together, the above analyses suggest that, contrary to the predictions of existing models of word associations to verbs (H. H. Clark, 1970; Ervin, 1961; Polzella & Rohrman, 1970), the nature of the difference in responses to transitive and intransitive cues is not one of grammatical class, but rather of the directionality of responses. This pattern was most marked when responses were nouns. This finding is taken up in more detail in the discussion.

Table 5.2
Response stereotypy means and t-tests. Bonferroni corrected $\alpha = .0125$

	Transitive		Intransitive		<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD		
N>1+N=1	21.10	7.355	18.41	7.07	1.849	.068
N>1	7.00	2.11	6.04	2.18	2.213	.029
N=1	14.10	6.35	12.37	5.49	1.447	.151
AFPRIM	16.49	9.505	20.73	10.88	-2.057	.042

5.5.3 Response distribution

The final stage of analysis was to investigate the stereotypy of responses. As an initial Kolmogorov-Smirnov test revealed that this data was normally distributed, t-tests were used to probe significant differences. As Table 5.2 shows, several aspects of these response profiles showed differences which approached significance. However, none remained significant at the Bonferroni correction α of .0125. In addition, because the intransitive verbs appeared to yield slightly more homogeneous response profiles, a further analysis was conducted after the removal of the five semantically narrow cues (four of which were intransitive), and their pairs, described above. The results of this analysis are presented in Table 5.3. This analysis demonstrates that the two verb types did not significantly differ in response stereotypy after the removal of these cues.

Table 5.3

Response stereotypy means and t-tests after removal of semantically narrow cues. Bonferroni corrected α = .0125

	Transitive		Intransitive		<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD		
N>1+N=1	21.32	7.332	19.41	6.29	1.311	.193
N>1	6.98	2.02	6.32	1.88	1.587	.116
N=1	14.34	6.47	13.09	5.08	1.008	.316
AFPRIM	16.32	8.71	18.82	8.25	-1.382	.170

5.6 Discussion and conclusion

The above results demonstrate a clear effect of transitivity on the directionality of position-based responses; transitive verbs were more likely to yield post-modifying responses (i.e. cue-response; the same direction as a verb-object relationship), while intransitives were more likely to elicit responses in the response-cue direction (i.e. the same direction as a subject-verb relationship).

These results offer new insights into existing theories of the influence of verb transitivity on word association. Firstly, Brown and Berko (1960) suggested that paradigmatic responses will dominate once a grammatical class has been fully acquired by respondents. They explained differences in WA responses as a function of GC in terms of the differing speed of acquisition of each GC – something that was itself hypothesised to be determined by the frequency of exemplars of that GC. This hypothesis is difficult to disprove in the absence of detailed information about the frequency of items in each grammatical class. However, the current results, as well as those of chapters 3 and 4, suggest

that GC influence runs deeper than Brown and Berko claim. In particular, Brown and Berko's hypothesis cannot explain the directional influence of transitivity found in the current study, because they suggested that the only influence of GC on WA responses would be in the proportion of syntagmatic to paradigmatic responses. Their theory offers no mechanism for the explanation of differences in directionality of responses.

Ervin's (1961) theory is only compatible with the current findings with modification. Ervin suggested that transitive verbs should become associated with words which commonly follow them in discourse, through the simple process of forward association. This would lead to syntagmatic (cue-response) associations. In the case of intransitives, since fewer possibilities for forward association would exist due to the absence of a grammatical object and the higher chance of being encountered in a sentence-final position, intransitives should instead become associated with words which share their semantic contexts and compete for selection during sentence production. This would lead to paradigmatic responses. This is not, however, the pattern found in the present study. What appears to be missing, in Ervin's theory, is the possibility that, in the absence of possibilities for forward association, intransitive verbs become associated with their grammatical subjects. In other words, the present results suggest that backward association may be a more important determinant of WA responses than Ervin claimed.

Polzella and Rohrman (1970), like the other studies discussed above, found that noun responses were more common for transitive cues than to intransitives, and were also produced more quickly to transitive cues. They explained this by suggesting that transitive verbs have a closer relationship with nouns than intransitives do, because they require a grammatical object in order to render them complete "cognitive units" (*Ibid.*: 538). While the present results do not directly support this hypothesis because there was no significant difference in noun responses as a function of transitivity, they are compatible with the view that a word's syntactic structure is stored along with the word's lexical entry (Chomsky, 1965; DeDe, 2014; Gahl, 2002; Polzella & Rohrman, 1970; Staub, 2007). Thus in order to accommodate the current findings within Polzella and Rohrman's theory, it would be necessary to suggest that a verb's syntactic argument structure (as in generative grammar: Chomsky, 1965) or syntactic biases (as in usage-based theories such as the Lexical Bias Hypothesis: Gahl, 2002) could lead a verb to become associated not only with its grammatical object, as Polzella and Rohrman suggested, but also with its subject. The different predictions made by generative and usage-based models of language processing will be discussed further in the next chapter.

Another issue which requires further investigation is the influence of two lurking variables which had a minor impact on the above results. The first of these can be loosely termed “semantic range”. There is currently no available variable which captures precisely this feature of language; perhaps the closest attempt is the concept of Semantic Diversity (Hoffman, Lambon Ralph, & Rogers, 2013; Hoffman & Woollams, 2015; Jones, Johns, & Recchia, 2012). While the five cues identified as having a particularly strong primary association strength (i.e. AFPRIM) made only minor differences to the findings of this study, it remains an open question as to the extent to which semantic narrowness is a feature of (in)transitivity itself. Intransitive verbs tend to cluster around related concepts (Hunston & Francis, 2000). Examples of this include involuntary noises (*cough, sneeze, hiccup, burp*), animal noises (*bark, roar, purr, hoot, howl*), words denoting conflict (*argue, bicker, tussle*), manners of movement (*dawdle, trundle, crawl*) and words related to procrastination or inactivity (*dally, dither, fidget*). Some of these words denote very specific actions, often with very narrow semantic ranges and quite specific actors. It may be, then, that semantic narrowness is a feature of the semantics of intransitivity itself, rather than something which stands independent of transitivity. Since numerous words from these and other semantic categories were included in this study, with some crossover of meaning, this issue remains a potentially confounding factor in the present study, and is a matter for future research.

The further investigation of this variable is challenging, however, for the reason that word association itself has often been the source of data on the semantic range of words. For example, Polzella and Rohrman (1970) used continuous WA measures to determine the “meaningfulness” of a word, while Chapter 3 discusses numerous further uses of both discrete and continuous WA to determine semantic range. However, the use of WA as a measurement for this variable makes it impossible to use as a variable in WA studies – a low associative set size is assumed here to be an *effect* of narrow semantic range, rather than its definition. This issue echoes wider problems in the WA literature, regarding the use of WA data as both an independent and a dependent variable (De Deyne & Storms, 2015; Hutchison, 2003). The issue of semantic range will be returned to in Chapter 8.

Finally, in Section 5.2, above, it was noted that transitivity is more complex than merely referring to the presence or absence of a grammatical object. Sentences vary in their transitivity not only through their argument structure, but across features such as their punctuality, telicity, and the degree to which the grammatical object is individuated or affected (Hopper & Thompson, 1980). Given that the current results have established an influence of transitivity on WA responses, further study is needed to establish whether these effects are related purely to the presence or absence of a grammatical object (as transitivity was defined in this study), or whether these additional aspects of transitivity

might also play a role in determining responses. For example, it is possible that verbs which denote a high degree of affectedness of the object (e.g. *destroy*) might yield that object as a WA response more often than is the case for verbs which tend to have a smaller effect on their object (e.g. *see*).

The following chapter, however, will return to the core findings of the experiment reported above, and will look at theoretical frameworks capable of explaining the effects of transitivity presented here.

Chapter 6: Theoretical explanations for transitivity effects

6.1 Introduction

In the previous chapter, different patterns of WA responses to transitive and intransitive cues were uncovered. These patterns pertained neither to syntagmatic/paradigmatic divergence in responses, as predicted by Brown & Berko (1960) and Ervin (1961), nor to grammatical class, as predicted by Polzella & Rohrman (1970). Instead differences were found in the directionality of position-based responses: transitive cues were more likely to yield cue-response associations (corresponding to the provision of a grammatical object to the verb cue), while intransitives were more likely to elicit response-cue associations (corresponding to a grammatical subject).

Chapter 5 closed by concluding that no current theory of word association predicts these findings. The purpose of the present chapter is therefore to review wider literature in search of a broader framework within which to accommodate them. The discussion will begin with a description of a generative-transformational model of word association, which asserts that response generation involves the unconscious application of a hierarchical set of rules. After pointing out some weaknesses of this view, the discussion will turn to usage-based (UB) perspectives. A contrast will be drawn between contiguity-based UB models, which focus on the role of lexical co-occurrence in the formation and generation of word associations; and semantic models, which highlight the cognitive work involved in structuring associative knowledge and propose that the majority of WA responses are derived from this structure, rather than from participants' knowledge of lexical co-occurrence.

6.2 Clark's Generative-Transformational WA model

Before discussing specifically semantic or contiguity-based accounts of word association, however, Clark's (1970) generative-transformational explanation of word association will be discussed. Clark's model is important to the current debate for three reasons. Firstly, it provided the background to Polzella and Rohrman's (1970) work, which was found, in the previous chapter, to incorrectly identify the nature of transitivity effects in WA. An understanding of the wider background from which Polzella and Rohrman generated their assumptions may cast light on why their predictions were not borne out in Chapter 5. Secondly, Clark's explicitly generative background provides a useful reference for the usage-based theories which will follow. Finally, Clark's work remains, almost 50 years after its publication, perhaps the most detailed attempt to provide a model of WA response generation – even

if, as we will see, the highly specific predictions made by the model only explain WA data at a relatively shallow level.

Clark's theory was intended to explain WA in terms of the same generative-transformational processes which guide language comprehension and production; i.e. it is governed by a finite set of rules which determine syntactic structure, and a store of lexical items which are slotted into this structure according to their own syntactic and semantic constraints. Clark argued that although WA cues are presented in isolation (unlike in contextual discourse), the generative rule-based system was nevertheless activated during response generation.

Clark's WA model asserts that three stages are involved in associative response production:

1. initial understanding of the cue
2. a set of operations upon it
3. the production of a response (H. H. Clark, 1970, p. 273)

Clark argued that the second of these stages involves the sequential application of a set of hierarchical transformational rules in order to generate a response, with the process terminating at the highest level at which a permissible response has been identified. He posited that the specific rule(s) to be applied depends on the features of each cue word. Cue features include syntactic constraints such as grammatical class (GC), countability, and transitivity; semantic features such as animacy and gender; and selectional features which determine the type of things which can permissibly co-occur, or which feature together in idioms.

The full list of rules suggested by Clark is:

1. The minimal contrast rule: the most detailed feature of the cue is changed in order to produce a response. The example of *man* is discussed by Clark. This word has the features [+Noun, +Det., +Count, +Animate, +Human, +Adult, +Male]. Since the most specific of these is [+Male], the minimal contrast rule would involve changing this feature to [+Female], and thus returning the response *woman*.
2. The marking rule: given a "marked" word, the rule selects an "unmarked" equivalent (e.g. "long" yields "short"); but the rule stops us from operating in the opposite direction (i.e. "short" does not elicit "long").
3. The feature deletion and addition rule: features (e.g. [+Male]) can be added or deleted in order to arrive at a new (often taxonomic) word. For example, adding [+volition] to "die" gives

the response "kill". Clark suggest that the rule prefers deletion, where possible, to addition of features (1970, p. 279).

4. The category preservation rule, which is essentially a sub-rule stating that lower level (i.e. more detailed) features of words will always be changed before higher level ones. These purportedly higher-level features reveal the syntactic underpinnings of the theory – the highest-level features are, according to Clark, grammatical ones such as GC and countability; key semantic features such as animacy are placed lower down the list.
5. The selectional feature realization rule: the lexical representations of some words (but, notably, not nouns) include selectional features which determine the type of thing that the word can modify, and thus co-occur with. The rule states that we can select from within these options. Clark (1970, p. 282) provides evidence from preposition stimuli, suggesting that where possible, we prefer the minimal contrast rule (i.e. for prepositions which have an antonym); but in cases where the minimal contrast yields nothing, we use selectional features.
6. The idiom-completion rule: if the word forms part of an idiom, we may complete the idiom in our response. This works only from left to right: "cottage" may yield "cheese", but "cheese" does not yield "cottage".

Importantly, Clark states that the first four of these rules would result in paradigmatic responses, while the last two would yield syntagmatic ones – a detail which explains the predominance of paradigmatic responses in adult WA.

An important aspect of Clark's model is its assertion that responding to WA cues is a serial, rule-governed process in which successive transformations of the cue are attempted in order to locate a suitable response. This is a very different conception of WA from the process of activation, search, and retrieval implied in network-based models of WA (cf. Collins & Loftus, 1975; De Deyne & Storms, 2015). Implicit in the network metaphor is a spreading activation mechanism, in which associated nodes (words) are activated (or suppressed) with a strength equal to the weight of the connection between them, and within which central words will be most likely to be selected as responses. Clark's model does not deny the possibility of the existence of such a network, but suggests that it is the rule-based operations upon the cue which determines response selection, not the network-based search-and-retrieval process.

The remarkable specificity of Clark's model means that it yields exceedingly clear predictions regarding common WA responses. In many cases, these predictions appear to be well supported by WA norms data. The two most detailed features of *burglar*, for example, are [+steals] and [+steals from houses].

Any change to the latter (or its deletion) would yield *thief*, except if the change yielded [+steals from banks], in which case the result would be *robber*. As such, the model predicts that *thief* and *robber* should be the most common responses to *burglar*, and indeed they are¹². Other examples, however, raise important questions about the knowledge which respondents would need to possess in order to generate appropriate responses. *Crocodile*, for example, yields as its primary associate *alligator*. According to the minimal contrast rule, this association would result from respondents changing the highly specific features of snout shape and/or fresh/saltwater habitat which distinguish crocodiles from alligators. Yet it is unlikely that a majority of respondents come to WA tests equipped with this knowledge¹³. While this problem can, to an extent, be countered by suggesting that the minimal feature used to create the association need not be explicitly understood by participants (or even factually correct), it is nevertheless probably true that, for many people, the only thing that distinguishes these two animals is that one is named “crocodile” and the other “alligator”. In the absence of more detailed knowledge, it seems probable that many respondents would perform minimal contrast transformations on known features, such as transforming [+big], yielding *lizard*, *gecko*, or *salamander*; or [+amphibious], resulting in *Komodo dragon*. Yet these responses are uncommon or absent. In such cases, the contrasting network interpretation that the association between *crocodile* and *alligator* arises through conceptual similarity and lexical co-occurrence seems a likelier interpretation than minimal contrast. A similar argument can be applied to abstract nouns: it is not clear what feature transformations separate *justice* from its primary associate *law*, or *creation* from *God*.

Another significant weakness of Clark’s model is its prediction that noun cues should *only* receive syntagmatic responses where that response completes an idiom. Several previous studies make this appear unlikely: De Deyne and Storms (2008) report that, in a databases of responses to around 1400 cues, at least 24% of responses to noun cues were verbs or adjectives. Examples such as *burglar-rob/steal*, *chorus-sing*, *sandpaper-rough*, and *knowledge-smart* argue against the idiom-completion interpretation. Some of these associations could be explained if Clark’s model were modified to allow nouns cues to specify selectional features, but doing so would violate Chomsky’s assertion that selectional features are not stored in noun representations (Chomsky, 1965). Furthermore, it seems

¹² All examples in this chapter are, unless otherwise stated, taken from the University of South Florida norms list (D. L. Nelson et al., 2004).

¹³ Note that, according to the norms gathered by Kiss, Armstrong, Milroy, & Piper (1973), *alligator* remains the primary associate to *crocodile* for respondents in Edinburgh, where people have somewhat less exposure to either animal than in Florida.

that in many cases a simpler explanation of some responses can be achieved without recourse to Clark's rules. Deese (1966, p. 109), for example, suggests that associations such as *grass-green* and *sky-blue* are simply stored as properties of the cue, and do not reflect rules of the type suggested by Clark's model.

As discussed earlier in this section, Polzella and Rohrman's model of transitivity effects in WA are derived from Clark's model. The authors do not explicitly state which of these rules their modification concerns, but Rule 6 appears most relevant. Their hypothesis, previously discussed in Chapter 5, is that transitive cues strongly bias the selection of a grammatical object because doing so results in a cognitively complete unit of meaning (implicitly, an idiom). Intransitives would not be biased toward noun responses in this way, partly because they are viewed as being cognitively complete, and partly also because the left-to-right nature of the idiom completion rule should preclude the selection of a grammatical subject. As Chapter 5 made clear, this prediction was not supported by the research in this thesis: transitives and intransitives were equally likely to yield both noun responses and position-based responses, and intransitives *did* yield apparent right-to-left associations, in contrast to the prediction made by Clark and Polzella & Rohrmann.

One final point concerns Clark's efforts to connect WA responses with contextual language processing. While this goal has clearly yielded a theoretically interesting model of WA, it is also very clear that WATs do not share many of the features of language comprehension or production. For example, WATs generally require only single-word responses, which furthermore do not need to convey any communicative information; there is no pragmatic intent on the part of either the test designer or the respondent. Responses do not need to be oriented to the needs of a given situation or interlocutor using such methods as the application of grammatical tense, aspect, or number; there is no need for pronoun or preposition usage for purposes of orientation (Levelt, 1989). As such, there is no *a priori* reason why it would be necessary assume that the same processes which guide language production would also guide word association.

Before leaving generative models of WA, it should be pointed out that Clark's model is based on a relatively early incarnation of generative language processing, inspired by the early work of Noam Chomsky (1957, 1965). Later generative models (e.g. Pustejovsky, 1991) would, if addressed specifically to WA, offer different sets of predictions which may resolve some of the above issues. An exploration of these models will not, however, be attempted here.

6.3 Usage-based theories

Generative explanations of linguistic phenomena have come under attack from researchers working with usage-based (UB) theories of language. UB theories posit that an individual's linguistic system is a product of their lifetime's experience of comprehending, processing, and producing language. The UB position is extremely broad, taking in viewpoints from cognitive psychology, cognitive linguistics, psycholinguistics, natural language processing, corpus linguistics, complex systems theory, and numerous other related academic disciplines (N. C. Ellis et al., 2016). This reflects the fact that, as Behrens (2009) states, UB theories are non-reductionist – they assert that language development can be influenced by every level of cognitive activity and linguistic analysis. As a consequence, the networks of knowledge which result, in UB theory, from linguistic experience include connections between phonological, morphological, syntactic, semantic, and pragmatic features of input. For example, Bøyum (2016, p. 22) suggests that the words *bed* and *leg* share phonological (through the shared vowel /e/), morphological (through the shared plural marker -s and its phonological realisation /z/), and both semantic and collocational (through the compound term *bed legs*) links in the mind. However, the words' representations would differ in many other ways, including other aspects of their phonology, semantic associations, collocations, contexts of usage, and pragmatic functions. Each of these elements is derived from a combination of linguistic experience and cognitive processing. It should be noted, too, that these networks do not exist only at the word level; they are instead based on the concept of *constructions* (Croft, 2001; N. C. Ellis et al., 2016; Evans, 2006a; Fillmore, 1988; Goldberg, 1995, 2006). A construction is a form-function pairing, of any size, which is encountered repetitiously in linguistic input. Examples of constructions include morphological and derivational affixes such as plural -s and adjectival suffix -able; words, collocational pairings, n-grams such as *it would be good if*, and syntactic structures such as the English ditransitive (Goldberg, 2006).

The cognitive processes which identify and operate upon these constructions are viewed as being domain general, and as operating on two levels. Firstly, processes operate on linguistic input by recognising, distinguishing, counting, and storing information from the environment. Taylor (2012) suggests that these processes result in the creation of a “mental corpus” which stores vast quantities of data, analogous to contemporary electronic corpora. The mental corpus is in fact far more detailed than computer corpora because it stores not only memories of linguistic experience itself, but additional details such as the place, time, and interlocutors involved in that experience. This vast cognitive structure provides the raw material for a second level of processing, including categorisation, abstraction, and generalisation processes. Through these processes, the material of the mental corpus

is sorted and made efficient, leading to the emergence of complex phenomena including grammar, which is viewed in usage-based models merely as “the cognitive organisation of one’s experience with language” (Bybee, 2006, p. 711). According to this view, the complex networks, semantic and otherwise, which structure linguistic knowledge (De Deyne & Storms, 2015; Deese, 1966) are also products of domain-general language processing.

One of the key implications of usage-based theories of language is that, unlike in generative linguistics, no innate, language specific processes are required to explain language development processes (Abbot-Smith & Tomasello, 2006; Bybee, 2010; Gentner, 2003; Langacker, 1987, 2000; Lieven & Tomasello, 2008; Norvig, 2012; Tomasello, 2000, 2003). The most compelling evidence for this viewpoint has come from research into child language development. Here, the generative position holds that since children are born with an innate knowledge of grammar, their acquisition of syntax should simply be a case of acquiring those linguistic parameters appropriate to their own language. This should, according to Abbott-Smith and Tomasello (2006), result in abrupt shifts in the grammatical accuracy of young children’s language, as settings are learned and automatically applied to all relevant lexical items. The usage-based position, on the other hand, argues that because language development is explicitly dependent on linguistic experience, children’s acquisition of syntax should be gradual, item-specific, and frequency-sensitive (Behrens, 2009). Numerous studies have now provided evidence for this latter cluster of features in language development (Akhtar, 1999; N. C. Ellis, 2002; N. C. Ellis et al., 2016; Lieven & Tomasello, 2008; Matthews, Lieven, Theakston, & Tomasello, 2005; Theakston et al., 2001; Tomasello, Akhtar, Dodson, & Rekau, 1997).

The discussion will now turn to how usage-based models can (and have) been viewed as underpinning both contiguity-based and semantic explanations of word association.

6.3.1 Contiguity-based WA models

Contiguity between words has long been assumed to be a determinant of association. Warren (Warren, 1916), for instance, traces the concept all the way back to Plato, who identified contiguity and similarity as the two foundations of associative thought. Slightly more recently, the theory was neatly summed up by William James: “Objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also, in the same order of sequence or coexistence as before” (James, 1890).

In the light of the above discussion, this classical conception of associative learning through contiguity can be viewed as a form of statistical, usage-based learning. Its cognitive basis is in the perception and

storage of repetitious sequences in experienced language. That such processes are implicated in UB theories is evidenced by numerous statistical accounts of language development, such as Hoey's *lexical priming hypothesis* (2005; see also Taylor, 2012, described above). In Hoey's model (as in Taylor's), an individual's sum linguistic experience is retained in a "mental concordance". From this structure, transitional probabilities between linguistic units (e.g. between a given word and another word, or a word and a grammatical construction) can be calculated in order to generate dynamic, probabilistic estimates of the likelihood of co-occurrence between (sets of) words or syntactic structures in given contexts (the lexical bias hypothesis, discussed in Chapter 5, makes similar assumptions DeDe, 2014).

Applied to word association data, these models make three key assumptions. The first is very general: that the human brain is capable of noticing and storing huge volumes of data, and of analysing it in such a way as to allow transitional probabilities to be derived. With more specific reference to word association, a second assumption is that a method of storage and retrieval dependent on a mental concordance-like structure provides a more parsimonious explanation of WA responses than other forms of cognitive architecture; and thirdly, that this mental concordance is actually referred to during WA production. Evidence in support of these models will be reviewed below.

Prior to the development of large text corpora, the presence of paradigmatic responses in WA was taken as evidence against contiguity-based account interpretations of WA, since it was assumed that these word pairs would rarely appear contiguously. One potential solution to this problem was offered by Ervin (1961; 1963), who suggested that paradigmatic pairs become associated through competition for activation during sentence processing. For example, words such as *walk*, *amble*, and *stroll* sometimes compete for selection. Ervin suggests that this will lead to the words becoming associated, through a mechanism which she terms "contiguity during competition" (1961, p. 362). This results in paradigmatic responses.

One weakness of Ervin's theory, which has subsequently become known as the substitutability hypothesis, is that it only accounts for paradigmatic responses which are directly substitutable – i.e. synonyms or taxonomically related words. Other types of paradigmatic responses, such as those characterised by Fitzpatrick as "conceptual" (*sweater-autumn*; *doctor-needle*), cannot be explained this way. A further criticism is that such potential uncertainties may in fact be too rare to account for many paradigmatic responses because of the way that contextual factors constrain word selection in fluent speech (cf. Griffin & Bock, 1998; Jurafsky, Bell, Gregory, & Raymond, 2000; Pace-Sigge, 2018).

However, the analysis of computerised corpora has revealed that the difficulty of accounting for paradigmatic responses within a contiguity-based model may not be as large as first thought. For example, numerous noun cues receive compound-completion responses such as *site-construction*¹⁴ and *hat-straw*, while other cues form frequent collocational pairings with their primary associate, as in *bat-ball* and *ever-never*; all of these would be considered paradigmatic in spite of their frequent textual co-occurrence. Indeed, looking through lists of word association norms, it is difficult to find examples of paradigmatic associations which would categorically *not* co-occur. Supporting this, two studies (Charles & Miller, 1989; Justeson & Katz, 1991) have both suggested that pure contiguity gives a better account of associations between antonymous adjectives than does Ervin's substitutability hypothesis (although both authors caution that their results may not explain paradigmatic responses in other grammatical classes).

A first step toward a more principled comparison between WA and text corpora was conducted by Spence & Owens (1990), who found that the associative strength between words correlated with their corpus co-occurrence. Several researchers have developed these findings. One such study is Bel Enguix, Rapp, and Zock (2014a). The study involved creating a network from co-occurring nouns, verbs, and adjectives in the British National Corpus, such that the link between two words was strengthened each time the two words were found in direct contiguity with each other. The resulting weighted network was then compared with responses to 5910 WA cues from the Edinburgh Associative Thesaurus (EAT; Kiss et al., 1973), with the weight of each link between words used as a predictor of the strength of association between them. While the level of correspondence between these two data sources was somewhat unclear in Bel Enguix et. al.'s description, they drew two conclusions. Firstly, they found that this model was able to predict both syntagmatic and paradigmatic responses. Secondly, they claimed that their model selected common associative responses with a similar frequency to a human respondent: an average of 6.2 respondents, across each of the 5910 cues, selected the same primary response as that selected by the model, compared with an average of only 5.8 respondents selecting the same response as that of a randomly selected respondent in the EAT. Bel Enguix et. al. claim that this demonstrates that "the system's answers could hardly be distinguished from human answers" (2014a, p. 3029).

While Bel Enguix et. al.'s findings are thought-provoking, they obscure some important differences between WA and corpus-derived data. In particular, they do not quantify (or qualitatively explore) the

¹⁴ While associative strengths between the pairs mentioned in this section vary, all represent the most common response to their cue, in the South Florida norms.

number of occasions on which their model predicted responses which were never produced by human respondents (cf. Kang, 2018; Wettler et al., 2005).

The magnitude of this issue was revealed in another comparison of WA and corpus data conducted by Mollin (2009). She compared responses to 30 cues randomly selected from the EAT (Kiss et al., 1973) with the collocational profiles of those words, as collected using a span of +/-4 words from the British National Corpus (BNC). Mollin found that the correspondence between these data sources was low. This is in part because far more words collocate than are ever produced in WATs: of 20,003 bigrams identified in her BNC analysis, only 626 also appeared as WA response pairs (compared with 577 pairs which occurred in the EAT but did not co-occur in the BNC). Even among these 626 words, however, there appeared to be little relationship between the two types of data: a regression analysis showed that mutual information scores explained very little variation in their strength of association in the EAT ($R^2=.07$). Mollin interpreted these findings as suggesting that pure lexical contiguity between words cannot explain WA behaviour, suggesting that the word association task “should [...] be seen as tapping into the semantic information of the mental lexicon only” (2009, p. 175).

One possible reason for the failure of studies such as these to identify close correspondences between corpus and WA data is that they may have selected too narrow a span of words through which to search corpora. While it may be the case that a narrow span, such as +/-4 words, is faithful to the concept of contiguity, a wider co-occurrence-based view, reflecting the possibility that the human mind can track and store co-occurrences between words at considerably larger distances, has been explored by Wettler, Rapp, & Sedlmeier (2005). Their study used a simple associative formula to generate co-occurrence-based predictions from the BNC. The model used a span of +/-20 words – far wider than in Bel Enguix et. al. (2014) – but used otherwise similar methodology. The model’s predictions were compared with responses from the 100-word Kent-Rosanoff list. The model was able to select the primary associate for 29 of these cues. By comparison, the average study participant selected the primary associate on 28 occasions. Furthermore, 64 of the model’s responses were selected by at least one participant, compared with 72 among human respondents. Wettler et. al. claimed that these results demonstrated that “the behaviour of participants in the free association task can be explained by associative learning of the contiguities between words” and suggested that semantic knowledge may not be necessary to explain WA findings. However, this conclusion is based on the similarity of the model’s response profile to that of a human respondent (i.e. the rate of overlap with other respondents – the same measure used by Bel Enguix et. al.). A goal better suited to this ambitious conclusion would be to predict *all* primary human WA responses, rather than to simulate

the response distribution of an average participant. The model clearly fails on this score. Most importantly, neither Wettler et. al. nor Bel Enguix et. al. have been able to explain why their models' predictions differ from human responses, such as on the 36% of occasions on which Wettler et. al.'s model selected a response which was never given by a human respondent.

One potential reason for this failure is hinted at in Wettler et. al.'s study. Their model was designed to compensate statistically for what the authors termed "word frequency bias" – i.e. the human preference for producing low-to-intermediate frequency WA responses. As discussed in Chapter 3, these response preferences have been measured in network models through the in-degree measure, which identifies the number of cues which yielded a given word as a response. Beyond frequency, human respondents also appear to prefer responses with low age-of-acquisition and high concreteness, resulting in high network centrality for these words (De Deyne & Storms, 2008, 2015). That this tendency was acknowledged in Wettler et. al.'s study reveals a tacit acceptance that human associative responses are not determined by co-occurrence alone. Instead, Wettler et. al. appear to have acknowledged that humans must make use of some other form of information or process, not reducible to co-occurrence, which accounts for their frequency preferences. This proposition is in fact the starting point for semantic views of WA, and will be further discussed in Section 6.3.2.

A cognitive element to WA was also suggested in another study comparing WA with corpus data, carried out by Kang (2018). He compared responses to selected cues from the Edinburgh Associative Thesaurus with the top collocates of those cues in the BNC. To do this, he used a textual co-occurrence method (Evert, 2008) which derives association measures (e.g. mutual information (MI), log likelihood (LL), t-score) from words within a "textual unit" – here defined as a single paragraph¹⁵. In line with a contiguity-based approach, Kang hypothesised that the top collocate selected by this method should coincide with the top associative response for each of the 3177 cues. The results showed that this was the case for 14.7% of cue-response pairs. Looking beyond the top collocate increased the likelihood of the most common WA response being identified by the model: in 25.1% of cases, it was present within the top 3 collocates, and this number rose to 52.3% of primary associates being present within the top 50 collocates. Furthermore, Kang found an overall correlation of .327 between the textual co-occurrence and associative strength of the word pairs in his study.

¹⁵ Kang also calculated these measures from single sentences, but the results showed consistently lower agreement with WA than did the paragraph-based measures.

These results are broadly in line with those presented for other co-occurrence-based studies above: they demonstrate similarity between co-occurrence and association, but fall far short of being able to explain all WA responses. Kang therefore suggested that while co-occurrence is “quite closely related to word association” (2018, p. 98), it may be the case that co-occurrence is “the starting point for full semantic knowledge” (*Ibid.*: 110). This suggestion is suggestive of the view that WA responses may not directly reflect the manner in which associations are learned – another key point made by proponents of a semantic view of word association (McRae et al., 2012). Kang further suggested the need for research looking into “the process of separating semantic relations from collocates” (*Ibid.*). These issues will be returned to in the following section.

One of the main problems emerging from the above research is that many human WA responses are not predicted by contiguity-based models. Schulte Im Walde et. al. (2008, p. 19) claim that most of these unexplained associations capture world knowledge which is not reflected in textual co-occurrence. They offer the examples of *drizzle-wet*, *munch-yummy*, and *magnet-physics* (among others). One way to explain these associations while remaining faithful to a co-occurrence-based view of WA is by suggesting that they may be derived from higher-order co-occurrence (i.e. the association of two words not through their own co-occurrence, but through their mutual co-occurrence with a third word; Lemaire & Denhière, 2006; Schulte Im Walde et al., 2008), or from knowledge of the shared contexts in which words occur. The latter method can be approximated using contextual models of lexical similarity, such as Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), which derive measures of the similarity of two words by comparing the similarity of the contexts in which they appear.

One attempt to compare such models with WA was made by Gruenenfelder, Recchia, Rubin, & Jones (2016). The authors constructed networks using three contextual models (the Topic model; Griffiths, Steyvers, & Tenenbaum, 2007; BEAGLE; Jones, Kintsch, & Mewhort, 2006; Latent Semantic Analysis (LSA); Landauer & Dumais, 1997), as well as from a simple co-occurrence-based model (similar to that of Wettler et. al., 2005, except that co-occurrence was defined as within a single text, rather than within a given window). The properties of these networks were then compared with those of an associative network built from the South Florida word association norms (D. L. Nelson et al., 2004). The authors calculated which (combination) of these networks provided the best fit with the properties of the WA network, as well as looking at the extent to which each model was able to predict WA responses. Their study found that none of the individual models, either contextual or co-occurrence-based, provided a close fit with the WA network. However, when hybrid models

containing both the co-occurrence-based component with one contextual component were created, the fit with the WA network's properties improved. In addition, the hybrid models tended to slightly outperform the individual models in terms of prediction of WA responses. Gruenenfelder et. al. therefore concluded that both co-occurrence and contextual information is drawn on in WA response production. However, none of the models (hybrid or otherwise) was particularly successful in predicting the 5 most common human WA responses. None was able to predict even 50% of these responses from within its top 56 predictions. This suggests that even when the predictions made by a context-based model are included in co-occurrence-based models, a large number of strong human associations are simply not predicted. While it is possible that these discrepancies are due to failures of the statistical aspects of the model (i.e. the manner in which the model predicts WA responses), the fact that more than 50% of top WA responses could not be predicted even in hybrid models strongly suggests that lexical co-occurrence and contextual similarity cannot explain all WA responses.

To summarise: although some strong claims have been made about contiguity- and co-occurrence-based models, the data presented here does not strongly suggest that human responses are determined by knowledge of contiguity or co-occurrence, or contextual similarity. Although it has been shown that such accounts are able to explain both syntagmatic and paradigmatic responses, there nevertheless appear to be many human WA responses which are not predicted by co-occurrence-based models, such as in the case of associations based on world knowledge. This discrepancy accounts, at least in part, for the low correspondences between human responses and contiguity-based associative models. More research is needed to address the issue of whether this discrepancy can be corrected using more advanced statistical models based on higher-order co-occurrence or contextual overlap.

Perhaps more importantly, the largely correlational nature of the research discussed above does not allow the investigation of any of the assumptions discussed at the beginning of the section: namely the plausibility of the cognitive capacities which underlie these models, and their parsimony; and whether or not such knowledge is accessed during WA response production. This suggests that evidence for a causative relationship between co-occurrence and WA responses remains far off.

6.3.2 Semantic models of word association

Semantic WA models assert that WA responses are determined not by the contiguity between two words, nor simply through the surface-level semantic similarity of two concepts, but by the manner and degree of a word's integration into semantic memory. Therefore, while these models do not deny that associations can be learned through contiguities between words in text (De Deyne & Storms, 142

2015; Deese, 1966; Santos et al., 2011), they assert that associations learned in this way are shallow, and will remain poorly integrated into semantic memory in the absence of semantic relations between their constituent word pairs (McRae et al., 2012; Prior & Bentin, 2008; Vivas, Manoiloff, Garcia, Lizarralde, & Vivas, 2018). This implies that cognitive work is carried out upon learned associates, resulting in an ongoing (re-)structuring of the networks of semantic knowledge from which WA responses are drawn. This cognitive work is consistent with both the predictions of usage-based models (e.g. Behrens, 2009) and with some reports of the structure of WA responses (e.g. Deese, 1966; see discussion below).

This basic description of semantic accounts of WA leads to a number of predictions which can be tested empirically. Firstly, if purely contiguity-based associations are not retained in semantic memory, they should show different patterns of priming in psycholinguistic tests. Such evidence has been provided in several studies. Pecher & Raaijmakers (1999), for example, taught participants new associations to semantically unrelated words using decontextualized paired associate learning and lexical decision methods. The authors then compared priming effects for these words with pairs which had strong pre-existing semantic connections, using both the lexical decision task and a perceptual identification task. The study found that while the newly learned associates showed a similar priming effect to the strongly associated pairs on the lexical decision task (i.e. the same task through which the associates had been learned), there was significantly less priming for these pairs on the (unfamiliar) perceptual identification task. These findings suggest that while it is possible to learn associations between unrelated word pairs, the effects of this learning may not generalise to new tasks, perhaps reflecting their incomplete integration into semantic memory. Furthermore, as McRae et. al. (2012) point out, the generation of these priming effects can require weeks of training (Dagenbach, Horst, & Carr, 1990; Schrijnemakers & Raaijmakers, 1997). McRae et. al. (2012, p. 48) conclude therefore that “meaningless associations are not retained precisely because they are meaningless”: without a semantic element, associative links do not persist in semantic memory.

Further evidence of this semantic imperative in long-term associative learning was provided by Prior & Bentin, in two studies exploring the learning of associations from sentential contexts. In the first of these studies (Prior & Bentin, 2003), the authors compared the learning of new associations presented either in sentential contexts or in decontextualized pairs. They found that the associations learned through sentential contexts were significantly easier to remember, as tested by cued recall and recognition tests. This suggests that the semantic context given to the two words aided the process of their integration into semantic memory. Two specific mechanisms potentially underlying this effect

were examined in the second study (Prior & Bentin, 2008): an *elaborative processing* mechanism, which states that depth of cognitive processing between words accounts for their association, and an *integrative processing* account, stating that the process of integrating words into a single, coherent semantic representation underlies the process of association. The authors tested these hypotheses by comparing retention of associations formed during processing of anomalous vs. coherent sentences. The authors argued that if the anomalous sentences resulted in greater learning of associations between words, this would support the elaborative processing account. This is because these sentences stimulate deep cognitive processing as a result of difficulty of generating a meaningful interpretation of the sentence. On the other hand, if the coherent sentences resulted in greater learning of new associates, the integrative processing account would be supported, since little elaborative processing should be needed to arrive at a coherent account of these sentences. Prior and Bentin's results supported the integrative processing account: more successful learning of associates was achieved from the coherent sentences. These results suggest that it is the process of deriving semantically coherent interpretations of sentences which accounts for the associations between the words in those sentences, not merely the co-occurrence of those words.

Further evidence for the incomplete learning of purely contiguity-based associations comes from studies researching "pure" associative priming – that is, priming effects between word pairs which are "associatively", but not semantically, related (Ferrand & New, 2003; Hutchison, 2003; Hutchison et al., 2008; Lucas, 2000; Murphy & Hunt, 2013). Hutchison, in a large-scale meta-analysis (Hutchison, 2003), has criticised much of this work on the grounds that it has inadequately delineated different types of semantic and associative relations. He argues that some researchers have defined semantic relatedness too narrowly (i.e. as semantic similarity), ignoring a wide range of semantic connections in cues selected as examples of non-semantic association¹⁶. Hutchison also levelled this criticism at studies investigating pure phrasal priming (i.e. priming between contiguous, but not semantically related, words). Such studies hypothetically provide support for a non-semantic interpretation of some WA responses by showing priming for word pairs which appear contiguously in discourse but have no semantic relationship (Hodgson, 1991; Perea, Gotor, & Nacher, 1997; Williams, 1996). However, Hutchison argues that these studies may not have adequately controlled for semantic relations. He gives numerous examples of word pairs used in these studies and therein categorised as contiguous and non-semantic: *spider-web*, *hammer-nail*, and *knife-fork* (Hutchison, 2003). It is not

¹⁶ Ferrand & New (2003), for example, defined semantically related words as those which were semantically similar, but categorised strongly associated words such as *needle-thread* and *aquarium-fish* as non-semantic.

difficult to see that, although these words are not highly similar, there is nevertheless a clear semantic relationship between them. Because of this problem, there is presently no reliable evidence available for existence of pure contiguity-based associative priming.

Indeed, as Deese (1966) has suggested, it will continue to be difficult to disentangle semantic from contiguity-based effects in studies such as these, because almost all contiguous, associated pairs share a semantic relationship of some sort. On the other hand, another prediction made by semantic models of WA is that it should be possible to find many examples of words which co-occur in discourse, but not in word association. Evidence for this comes from Mollin's (2009) finding that only 626 of twenty thousand frequently co-occurring bigrams (containing a given node) from the BNC were found in the Edinburgh Associative Thesaurus. As yet, however, too little research has looked in detail at the factors which differentiate those collocates reproduced in association from those which are not.

A third prediction made by semantic WA models is that associative networks should differ from those generated through co-occurrences. There is evidence both for and against this prediction. Firstly, as described in Section 6.3.1, a study by Gruenenfelder et. al. (2016) found that networks derived either from a contiguity-based associative algorithm or from contextual similarity models such as Latent Semantic Analysis failed to reproduce the properties of networks derived from WA responses. Specifically, these models were not able to emulate the clustering of WA networks: the associative, contiguity-based model produced sparser clusters than WA (perhaps reflecting the fact that many more words co-occur than are produced in WA), while the contextual models yielded denser clusters than WA networks (perhaps reflective of the semantic specificity of contextual models). These results support the prediction made by semantic accounts of WA that the properties of WA response networks will differ from co-occurrence-based networks. However, Gruenenfelder et. al. also found that hybrid models which combined the associative, contiguity-based network with a single contextual model corrected the poor fit with WA networks. These models were able to accurately reproduce the clustering of WA response-based networks, as well as emulating other aspects of their network distribution. These results are more favourable to the co-occurrence-based view of WA. It should be noted, however, that, as discussed in Section 6.3.1, these hybrid models still struggled to accurately predict specific WA responses.

A second, though less direct, source of evidence on the difference between co-occurrence based models and WA responses is provided by the work of Deese (Deese, 1962b, 1964, 1966). He investigated the semantic underpinnings of the lexicon by collecting associative responses to a set of words, and then (at a later date) presenting those responses as cues to new participants. For example,

the cue *butterfly* (Deese, 1962b) yielded responses such as *moth, insect, wing, yellow, and spring* in the initial study. Deese therefore presented these words as cues in a subsequent study. He investigated the extent to which the responses to these secondary cues overlapped. Matrixes of cue-response frequencies were created to measure this overlap, and factor analysis was applied to these matrixes in order to identify sets of words which shared the greatest overlap in their responses. In the example of the *butterfly* data, for instance, Deese found that factor analysis separated animate (*bees, fly, bug*) from inanimate concepts (*sky, nature, yellow*). Additional factors revealed bipolar splits between inanimate concepts: one factor contained *summer, sunshine, garden and flower*, while the other included *sky, blue, yellow, and colour*.

Deese demonstrated similarly logical semantic divisions of response profiles for numerous other concepts, including words related to music and religion, adjectives of size, and common adverbs, and attempted to collect these conceptual sets into an “associative dictionary” (Deese, 1966). This dictionary was deeply semantic in nature. Deese suggested that the patterns of associative responses given to a word reflect what the respondent perceives that word to mean: “the distribution of associates to any word provides the associative meaning of that word” (1962b, p. 174), where “associative meaning” was explained as “a subset of [...] *meaning*” (1966, p. 43; Deese’s italics). This characterisation of “associative meaning” as something distinct from “categorical meaning” and “dictionary meaning” (Ibid.) suggests that Deese viewed word association tests as capturing some property of words uniquely related to human perception and cognition. It might be said, then, that his conception of association was of a representation of links between words which are not reducible to any more specific general distinction than that they are simply “associated”.

While the concept of usage-based linguistics was yet to be developed at the time of Deese’s research, some of his conclusions are similar to contemporary UB views of lexical organisation. He suggested, for instance, that while associations can be formed through experience of contiguities between words or entities, or perception of their similarity, the structures into which they eventually organise themselves “derive in whole or part from the structures or categories of the human mind” (1962b, p. 174). Once experienced, words are subject to cognitive processes which result in their efficient semantic organisation. Deese would therefore have little difficulty explaining the failure of the mathematical, contiguity-based models of WA presented in Section 6.3.1 to explain a majority of primary WA associates: these models do not integrate processes of categorisation, abstraction, generalisation etc. on the input they receive, as human respondents do.

Nevertheless, a clear-minded view of Deese's work necessitates the conclusion that while Deese's findings are striking, they do not definitively demonstrate the semantic nature of associative structure, because they did not show that a contiguity-based network could not yield the same or similar clusters of associates, and because they were based too narrowly on the small sets of words most likely to support his hypotheses. Future research in this area, then, needs to provide clearer evidence both that contiguity-based networks differ from those based on WA responses, including when compared to responses to cues different from those used by Deese, such as abstract nouns or verbs.

One study which did hint at a psycholinguistic basis for Deese's concept of "associative meaning" was conducted by Vivas et. al. (2018). They tested the hypothesis that word associations reflect the defining, or most relevant, features of cues. They gathered responses to 199 concrete nouns, and calculated the correspondence of these responses with independently collected judgments of the defining semantic features of the same words. Overall, the study reported a relatively high overlap between these types of data. The authors compared the vectors produced by responses to the WA and the semantic features task, finding that 72.5% of WA distributions showed correlations of $r > .4$ with the semantic feature measures, including 29% of distributions with a correlation of between .6 and .8, and 8% with a correlation above .8. In addition, they found that 86.5% of primary responses corresponded with one of the verb's main semantic features. These results suggest that those features which most closely define a given word might be central to WA response production.

A final prediction made by semantic models of WA is that it should be possible to code a large majority of WA responses using semantic schemes. A number of studies have attempted to categorise WA responses according to their semantic properties. However, several of these studies have focused on paradigmatic responses only. For example, in a German language study¹⁷, Schulte Im Walde et. al. (2008) used the German language version of WordNet (Miller, Beckwith, Fellbaum, & August, 1993) to categorize the same-grammatical class relationships between more than 30,000 paradigmatic cue-response types, generated from approximately 600 noun and verb cues. The relationships encoded in WordNet are synonymy, antonymy, hypernymy, hyponymy, and co-hyponymy (for both nouns and verbs), plus holonymy and meronymy for nouns, and cause and entailment for verbs. They found that only around half of these cue-response pairs could be classified according to the WordNet scheme. This was due, in part, to the incomplete state of the GermaNet database, but also to the presence of

¹⁷ In both this study and those of Guida & Lenci (2007) and Vivas et. al. (n.d.) which follow, cues and responses were given in German, Italian, and Spanish, respectively. In the interests of space, only English translations are given here.

several semantic relationships not included in WordNet. These included consequences (*sweat-stink*) and causes (*sweat-run*), plus temporal relations (*address-send*) locations (*camel-desert; nurse-hospital*), and events (*goose-Christmas, sledge-snow*). This suggests that the WordNet scheme might be too narrow to capture the range of associations given in WA. Of those pairs which were successfully categorised, co-hyponymy was the most common relationship for both noun-noun (46.7% of all classifiable noun-noun responses) and verb-verb (42.3%) pairs.

Guida & Lenci (2007) took a similar approach to the analysis of WA responses to Italian verbs. However, they reported much higher levels of successful classification of verb-verb relations using the Italian version of WordNet. Only 6% of cue-response pairs in their study could not be classified, with synonymy (38.3%) and superordinate (22.8%) relations dominating. From the unclassified pairs, Guida and Lenci identified similar relations to those found by Schulte Im Walde et. al. (2008); temporal order (*cook-eat*), events (*conduct a test-give a lecture*), or purposes (*read-learn, record-remember*).

While these results are clearly more suggestive of a close semantic nature to WA, it should be remembered that they only concern paradigmatic relationships – those which should theoretically be easiest for semantic models to explain. However, Guida & Lenci (2007) also presented an investigation of the semantic properties of verb-noun relations in their study. They categorised all such pairs according to the sentential roles, such as *agent, patient, and experiencer*, as well as other relationships such as results (*disappoint-sadness*) and locations (*run-park*), shared by word pairs. The authors did not give exact statistics regarding how many pairs remained un-coded according to this scheme. However, while they commented that the scheme was “quite satisfactory in dealing with experiment responses” (Guida & Lenci, 2007, p. 313), it did not succeed in categorising all responses.

Guida and Lenci’s analysis of these results is of particular interest for the findings of Chapter 5. They found that while the *patient* role was by far the most common in general (35.1% of verb-noun responses), there was a tendency for *agent* associates to be given for more specific verbs. For example, they compared the highly polysemantic verb *go*, which did not yield agent-type associations, with the more specific verb *march*, for which 55% of associations were agents; these included *soldiers, troops, and army*. Similar patterns were observed for experiencers of verbs (*annoy-audience*). The authors argued that “agents and experiencers are usually produced as associations only when the stimulus verb strongly implies in its meaning a particular kind of agent/experiencer. In these cases, the verb lexically constrains the type of agent”. Similar lexical specificity effects were uncovered in interactions between cue frequency and cue-response semantic relationship type: high frequency verbs yielded significantly more subordinate verbs as responses, while low frequency verbs yielded more

superordinates (correlation between cue frequency and number of subordinate responses $r=.14$). Chapter 8 will look at this issue in more detail, with reference to the transitivity effects identified in Chapter 5.

These studies raise questions regarding the use of semantic coding schemes. Firstly, it may be that the differing results of the two studies (i.e. co-hyponymy relations dominating in Schulte Im Walde et. al., but superordinates and synonyms receiving most responses in those of Guida & Lenci) resulted from inconsistencies in the semantic categorisation of words by different groups of researchers, or within different-language versions of WordNet. The results may suggest that coding WA pairs semantically is not as easy as it may appear.

Secondly, both studies also reveal the inadequacy of WordNet's relatively narrow semantic scheme – most pointedly because WordNet does not encode syntagmatic relationships. Other researchers have suggested the use of far more detailed coding schemes. McRae et. al. (2012), for example, present a taxonomy containing 5 basic categories (*similar concepts*, which comprises several of the relations contained in WordNet, plus *entity*, *situation*, *introspective*, and *event*), each of which comprises numerous more detailed denotations, adding up to a total of 28 sub-categories. They have challenged researchers to “theoretically and empirically delineate among them” (2012, p. 43), and suggest that the process of doing so may lead to a clearer understanding of the types of items which are most strongly represented in semantic memory. Vivas et. al. (2018) applied this scheme to their Spanish WA responses, and found that the majority of responses were situational (e.g. *dog-bone*, *worm-soil*; 62.7%) in nature – a finding which appears to disagree with those of both Schulte Im Walde et. al. and Guida & Lenci, who found co-ordinate and synonym responses, respectively, to dominate in their studies.

This finding also contrasts with those of another set of studies (Santos et al., 2011; Simmons et al., 2008) which required the creation of a novel scheme for describing semantic relations. In these studies, the scheme was specifically devised to test the theory that language processing is handled by two distinct neurocognitive components – a linguistic system which stores surface-level knowledge, and a simulation-based system which generates novel embodied simulations of perceptual experiences in order to provide a detailed representation of a concept. This model is termed Language and Situated Simulation (LASS; Barsalou, Santos, Simmons, & Wilson, 2008). While both the linguistic and simulation-based systems are hypothesised to become active immediately upon perceiving a word, the linguistic system is assumed to peak first due to its relatively shallow nature. This leads to the prediction that semantic relations encoded within the linguistic system should be generated more quickly than those which derive from the simulation-based system, as well as being processed by

distinct areas of the brain. Cue-response relationships assumed to derive from the linguistic system were compound completions (*bat-man*), sound similarities (*hair-fair*), root similarities (*inject-injection*), plus synonyms and antonyms. Relationships assumed to derive from the situated simulation system were termed *object-situation* responses, which described either a property of a concept or an aspect of the situation in which it might be encountered. Examples included *bee-flower* and *golf-boring/sunshine* (Santos et al., 2011, p. 91). Other types of relationship – most notable hyponyms, hypernyms, and co-hyponyms, were not explicitly related to either system, since the authors claimed that either system could account for such responses. In both studies, the authors found support for the LASS model: in Santos et. al (2011) found that, in both word association and property generation tasks, relations derived from the linguistic system were produced more quickly than those from the situation simulation system. Simmons et. al. (2008) additionally found a neural dissociation between the two processing routes in fMRI examinations of neural activity during various word association, property generation, and situation generation tasks.

These ideas received further support from De Deyne et. al. (2008) who found that linguistic and taxonomic associations tended to be given as first responses in a three-response WA task; responses produced second and third were more likely to be entity and thematic relations, similar to those categorised by Santos et. al. as object-situation. As described above, however, Vivas et. al. (2018) found that situational responses were in fact the most common type of response. This was despite the fact that the time given for response generation in their study (7 seconds per cue) was less than the time which Simmons et. al. hypothesised to delineate linguistic and simulation-based responses (7.5 seconds) in their fMRI study.

It is possible that the significant variation in the type of semantic responses preferred by participants in these studies is due to the different languages used in these studies (De Deyne & Storms, 2008; Guida & Lenci, 2007; Santos et al., 2011; Schulte Im Walde et al., 2008; Vivas et al., 2018). Perhaps some aspect of the morphological make-up of these languages biases respondents towards a certain approach to the cue word. Nevertheless, it is also the case that problems with the variation in coding schemes used in the studies, or their implementation, make it difficult to confirm the prediction that semantic coding schemes should be able to categorise all WA responses. A particular problem for the consistency of these coding schemes is the possibility that individual research teams unconsciously reproduce their own coding preferences when categorising data. This possibility was hinted at in Chapter 2 with reference to Fitzpatrick's findings that WAT participants display individual preferences in the type of *responses* they give (Fitzpatrick, 2007, 2009, and see Chapter 2), and is particularly likely

in the case of detailed, multi-level schemes (e.g. McRae et al., 2012; Santos et al., 2011), since it is often possible to categorise a given response in more than one way. Given this challenge, it is somewhat surprising that WA studies rarely report detailed coding practices and inter-rater reliability ratings. These issues will be revisited in Chapter 7.

Nevertheless, it should also be noted that there is considerable research potential in deriving accurate, easy-to-use categorisation schemes for semantic relationships between WA responses. This point is emphasised by McRae et al. (2012), and also by Hutchison (Hutchison, 2003; Hutchison et al., 2008), who argues that the failure to properly delineate the semantic relations encoded within word association response norms lists has led to significant problems of interpretation regarding studies using WA norms as predictors of psycholinguistic performance. While he suggests (Hutchison et al., 2008) that word association remains one of the best methods for the extraction of semantic similarity estimates for use in psycholinguistic research, as well as one of the most accurate predictors of response times in psycholinguistic tasks such as lexical decision, the fact remains that many WA responses do not reflect direct similarity between concepts at all. As such, it is plausible that a clearer description of the semantic relations found in WA responses will further enhance its predictive powers, and shed new light on the types of information most salient to semantic memory.

In summary, semantic models of WA differ from contiguity-based ones in that they assert that different processes guide WA response production from those which determine the formation of associations. Specifically, they assert that semantic processing, resulting in integration of associates into semantic memory, is a prerequisite for production of word pairs in word association tests. Semantic memory is generally described as a network-based structure served by a search-and-retrieval mechanism. Furthermore, semantic models view cognitive processing as a key determinant of the manner and degree of integration into the lexicon.

Semantic models make three key predictions regarding WA. Firstly, there should be evidence that non-semantic associations are unstable and temporary: this prediction is well supported by data from a range of studies. Secondly, respondents should display selectivity in terms of the co-occurring words which they reproduce in WATs: this finding is also well supported, although more research is needed to examine the nature of this selectivity. Finally, it should be possible for all WA responses to be categorised according to semantic schemes. Most schemes defer slightly here, allowing responses to be categorised as form- or contiguity-based under specific circumstances. Nevertheless, it remains unclear to what extent all other responses can be classified as semantic. All current schemes have

reported at least some degree of failure here. In addition, the inconsistency of the various coding schemes (and perhaps also their application) is a barrier to progress in this area.

6.4 Discussion

At the beginning of this chapter, it was suggested that both contiguity-based and semantic models of WA could be accommodated within a usage-based framework. This is because UB frameworks highlight the importance of both statistical repetition learning (which suggests contiguity-based lexical knowledge; Hoey, 2005; Taylor, 2012), and subsequent cognitive processes involving the reorganisation of existing knowledge into efficient, semantically-ordered networks (which would be suggestive of semantic influences on word association; Behrens, 2009; De Deyne & Storms, 2008; Deese, 1966).

Neither of these accounts appears to give a fully coherent explanation for all WA responses. In the case of the contiguity-based models, this is both because they generally offer only a modest overlap with human WA responses, leaving a large number of responses unexplained even when supplemented by contextual data (Gruenenfelder et al., 2016); and because for those WA responses which do overlap strongly with corpus-based data, there is no evidence of a causative relationship. The same criticism can be addressed at semantic models, however: since frequently co-occurring words rarely fail to reflect a semantic relationship, it is difficult to assert a semantic locus for responses which co-occur. In addition, the semantic categorisation of responses has proven to be a difficult matter, with little agreement between research groups with regard to the most commonly occurring semantic categories in WA, perhaps reflecting inconsistencies in categorisation practices.

A further possibility, not explicitly explored above, is that multiple types of information could simultaneously influence word association. This is, in fact, generally acknowledged by most researchers – as noted above, even largely semantic schemes tend to allow categories for non-semantic responses such as form-based associations and compound completions (Santos et al., 2011; Schulte Im Walde et al., 2008; Vivas et al., 2018), while some contiguity-based studies also assert that lexical co-occurrence may only influence the early stages of associative knowledge (Kang, 2018).

More balanced categorization schemes explicitly refer to the relevance of a range of factors on WA responses (e.g. Fitzpatrick et al., 2015; K. Nelson, 1977). However, few studies have attempted to quantify the relative contributions of these types of knowledge. One such study was conducted by Fitzpatrick & Izura (2011), who collected WA responses to 95 cues in a first and second language (Spanish and English, respectively), along with response latencies. The responses were coded as

equivalent meaning, non-equivalent meaning, collocation, and form-based; however, responses which could fit more than one category were coded as dual-route responses (*form and meaning, and meaning and collocation*). The authors found that in both languages, responses of non-equivalent meaning dominated. However, these responses were also slower than other types of response. The fastest responses were to the *meaning and collocation* associations. The authors interpreted these findings as suggesting that, where available, different types of word knowledge might make relatively independent contributions to the activation and selection of a response, thus speeding its production: "if the word association event is conceptualized in terms of activations, then it might be suggested that there are a number of different kinds of potential activation triggers (meaning, collocation, form) and that the activation event is stronger, or faster, if more than one of these is present" (Fitzpatrick & Izura, 2011, p. 392).

Fitzpatrick & Izura's study lends some support to the view that different types of information contribute to WA responses. Gruenenfelder et. al. (2016) reach a similar conclusion (though through very different means) in their analysis of the superior performance of hybrid (co-occurrence plus contextual similarity) models over single-component models of WA. They suggest that their results point to different sources of knowledge being available during WA response production.

Further research is needed in order to verify these findings. One approach to this issue is to compare human response categorizations with corpus data, by analyzing co-occurrence scores for word pairs of different categorical types. A straightforward prediction here is that, if position-based responses are genuinely based on co-occurrence (rather than through some covert semantic relationship, such as a situational link), they should co-occur more frequently than purely semantic associations. In view of the transitivity effects reported in Chapter 5, it can further be predicted that co-occurrence scores for responses coded as position-based should be higher in the direction predicted by their coding. Such a study can therefore perform the dual roles of investigating the semantic and co-occurrence-based underpinnings of WA while simultaneously exploring the specific transitivity-related findings presented in Chapter 5. As such, the aim of the following chapter will be to test these predictions.

Chapter 7: Combining co-occurrence data with word association response categorisations

7.1 Introduction

In the previous chapter, generative and usage-based models of word association (WA) response production were discussed in the light of the findings of Chapter 5, which found categorical (position-based) effects of cue transitivity on WA responses. The discussion concluded that generative models of WA cannot explain these findings because they assert a left-to-right imperative which precludes the production of response-cue associates (H. H. Clark, 1970), and because they state that cue transitivity will influence only the proportion of noun to non-noun responses (Polzella & Rohrman, 1970). In addition, it was suggested that existing generative models of word association do not offer a satisfactory account of word association response data in general, since they fail to explain some common WA response patterns, such as noun-adjective pairs. The discussion then turned to usage-based (UB) models of WA, and similarly found that neither of the two UB models described in the chapter – contiguity-based models and semantic models – is able to independently capture *all* WA response variation: co-occurrence-based models appear to struggle to predict more than around half of WA responses, while semantic categorisation schemes may be unable to capture some types of associative relations.

One of the benefits of usage-based models of language processing is that they do not, however, require that co-occurrence-based and semantic models be considered mutually exclusive with regard to the explanation of word association responses. This is because usage-based models are maximalist (Behrens, 2009) – they assert that all forms of information available from a language user’s experience will influence their subsequent language use. As such, the main aim of the current chapter is to develop and describe a composite usage-based model of word association, which acknowledges the influence of several types of word knowledge on responses (e.g. semantic, co-occurrence-based, and form-based) while remaining fully compatible with an overarching usage-based framework.

Chapter 6 presented some evidence in support of such a model. For example, Fitzpatrick & Izura (2011) found that responses categorised as both meaning-based and collocational were produced more quickly than those coded into only one of these categories; while Gruenenfelder et. al. (2016) found evidence for a superior fit to WA data of models based on both co-occurrence and contextual similarity data, compared to those based on just one of these sources. Chapter 6 also highlighted the importance of both contiguity and semantic relatedness in the integration of words into semantic memory

(Hutchison, 2003; Hutchison et al., 2008; McRae et al., 2012; Prior & Bentin, 2003, 2008). The model to be developed in this chapter will take these results, as well as those found in the preceding chapters of this thesis, as its starting point. After describing this model, two experiments which aimed to look for further evidence for this view of WA response production will be described.

7.2 A composite usage-based WA model

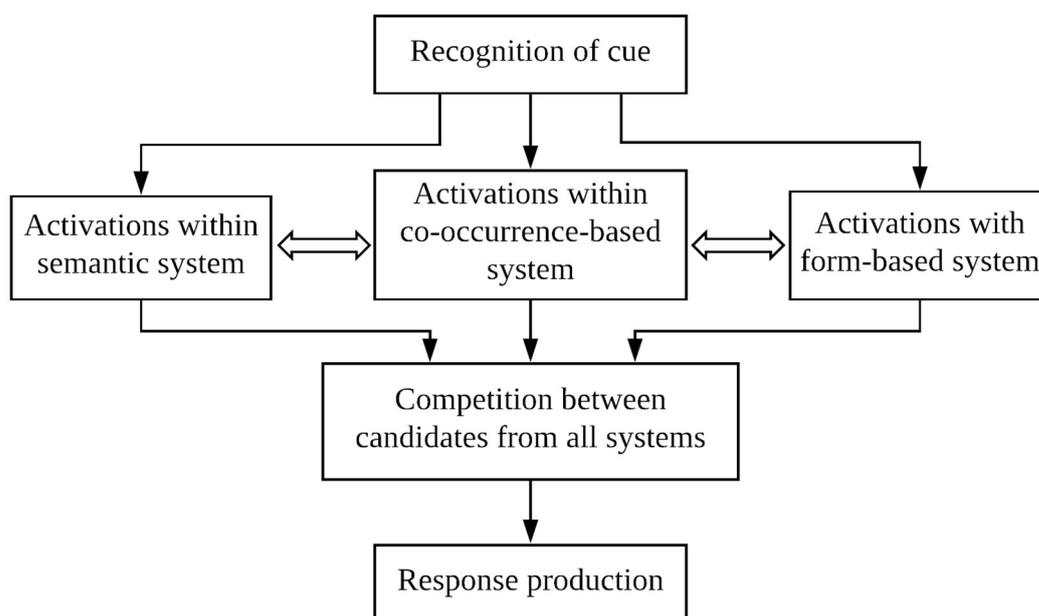
The model to be described here is explicitly usage-based in nature in that it views associative knowledge as emerging from experience of language in context, as well as from several layers of cognitive processing of that linguistic experience. These processes include the recognition of statistical patterns in language (e.g. n-gram frequencies, transitional probabilities etc.) as well as cognitive operations such as categorisation and generalisation.

The knowledge derived from this linguistic experience is viewed here as being stored in one of three components of memory. The three components are a co-occurrence-based structure similar in nature to the “mental concordance” put forward by Hoey (2005; see also Pace-Sigge, 2013, 2018; and Taylor, 2012), which stores both holistic examples of language in use, and statistical knowledge derived from these examples; a semantic component which stores knowledge of semantic relationships between words, derived from integrative processing during text comprehension (Prior & Bentin, 2008), and cognitive processing of words already stored in semantic memory, such as sorting, abstracting, equating, and categorizing (Behrens, 2009; N. C. Ellis et al., 2016); and finally a form-based component which is sensitive to orthographic, phonological, and morphological aspects of words. Each of these components corresponds to a particular type of word association response: position-based responses are viewed as emerging from activations in the co-occurrence-based component, meaning-based responses from the semantic one, and form-based responses from the phonological/orthographic sub-system.

Associations stored within each component are viewed as initially competing for activation with other response candidates within the same sub-system. For example, in response to *dog*, the semantic component might generate *cat*, *wolf*, or any other number of possible associates. These compete for activation (see Figure 7.1). Alternatively, one or more of the components may not generate any response candidates. This might happen, for example, in the case of the semantic and co-occurrence-based system when searching for responses to unfamiliar words, such as *sternutation*. In such cases, only the form-based system will be able to generate a response, based on orthographic similarity.

Response candidates generated within each component in a given timeframe are integrated, and then compete for final activation as the associative response. Competition with this integrative component is not necessarily strong, since it is unlikely that all three memory components will return strong response candidates in all cases. For example, if one system quickly outputs a candidate, while the other components do so only much more slowly, this candidate will be selected in the absence of competition. On the other hand, if all three components output different candidates at a similar point in time, these candidates will compete further for final selection.

Figure 7.1
The composite model of word association response generation.



For present purposes, the three components of the model are viewed as contributing to response generation with equal strength, and as being activated simultaneously: there is no hierarchy of operations. This means that none of the systems is preferred over the others, or that any operates more quickly than the others. The sole determinant of the speed at which a candidate is selected within a given component is the strength with which it is activated by processing of the cue¹⁸. The only exception to this rule is when two or more of the model’s components simultaneously activate the same response candidate. This interaction, which is viewed as being purely facilitative, is indicated by the bi-directional arrows between the three memory components in Figure 7.1. An example of this

¹⁸ It is acknowledged that certain findings from earlier WA research, such as the widely reported infrequency of form-based responses, argues against this equality of influence. These arguments will, however, be ignored for the time being, and returned to in Chapter 9.

interaction is the production of the association *cat-dog*. In order to quickly generate this response, activations occur in the semantic (on account of physical and contextual similarities between these animals), co-occurrence-based (because they co-occur in discourse) and form-based (because both are CVC words) components. These multiple activations facilitate one another, speeding the activation of *dog* in response to *cat*, and thereby making its selection more likely than words associated with *cat* in only one component, such as *jaguar*.

The composite model is a model of WA response *generation*; it does not (for the time being; see Chapter 9) model the acquisition and development of associative knowledge throughout the lifespan. Nevertheless, the three components of the composite model are assumed to derive from lifelong linguistic experience, as well as from the subsequent (also lifelong) cognitive processing of that experience. It is therefore not assumed (as McRae et al., 2012, argue has been the case in much previous WA research) that the same processes which guide the initial learning of associations (e.g. contiguity) are necessarily also those which govern the generation of responses in WATs. In other words, although experience of words in textual contiguity (e.g. *blue-sky*) results in the formation of associations in contiguity-based memory, it is also possible that subsequent processing will lead to the formation of associations between the two words in semantic and/or form-based memory. For example, this processing might lead to their association within semantic memory, as the experiencer reflects upon the blueness of the sky. As such, when either word (i.e. *blue* or *sky*) yields the other as an associative response, it cannot be assumed that the contiguity-based memory component is the sole source of the generation of the association.

This implies that the interactivity between word activations during response generation also exists at a deeper level throughout the lifespan, as words from (e.g.) the contiguity-based component are accessed and analysed by the semantic and form-based components. This reflects the dynamic, iterative process in which the linguistic system constantly re-adjusts to accommodate novel linguistic experiences (cf. Elman, 2004). Experiencing novel co-occurrence-based information, such as a familiar word being used in an unfamiliar collocational pairing, can prompt changes not only to the structure of the co-occurrence-based system, but also to the semantic one (e.g. if the new collocation brings out a novel sense of a word). These changes might include the strengthening or weakening of existing links between words, or the formation of new links. It is this process which results in the structure found within the three components.

Co-activation from more than one memory component can explain why some potential responses are more commonly selected than others, as found in some studies modelling WA responses in terms of

co-occurrence (recall the findings of Mollin, 2009, who found that far more words co-occur than are ever selected as WA responses). This is because associations from the contiguity-based component will not straight-forwardly be those which most frequently co-occur. Instead, associations derived from this sub-system will be more likely to achieve activation if they are also linked in the semantic system.

Many WA studies imply a similar model to that described above, in that they assume multiple potential processing routes. The coding scheme used in Chapters 1-5, for instance, is based on the assumption that these three basic forms of knowledge can influence WA (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick et al., 2015; Nation, 2001). Even explicitly semantic or co-occurrence-based approaches acknowledge these multiple influences in WA. For example, Vivas et. al. (Vivas et al., 2018) added form-based and compound completion categories to their otherwise entirely semantic coding scheme, while Kang (2018) acknowledges that co-occurrence-based information may only be a starting point for WA. Rather than being seen as diverging from previous research, then, it is largely only the explicit formalisation of these assumptions, with its hypothesised (provisional) equality between the three components, which sets this model apart from earlier work.

Two further issues are left undefined here. Firstly, no claims are made as to the neural separability of the model's components. This means that it should not be assumed that the model proposes discrete neural systems for semantic, co-occurrence-based, and form-based links. Instead, it is possible that a single distributed cognitive structure encodes relationships of these different types. This question is left for later research. Secondly, although it is assumed that conscious processes are more likely to be active in the later stages of response generation, no specific claims are made as to the timing, conditions, or effects of conscious processes in WA.

Having described these formal characteristics of the composite model, this chapter will now move on to a discussion of two experiments which were conducted in order to test some of its predictions.

7.3 Experiment 1a

7.3.1 Aims and predictions

As described above, the composite model assumes that meaning-based responses are derived from semantic memory, position-based associations from the co-occurrence-based system, and form-based responses result from activation of phonological and orthographic knowledge¹⁹. The nature of

¹⁹ This does not contradict the earlier statement that "is not assumed that the same processes which guide the initial learning of associations (e.g. contiguity) are necessarily also those which govern the generation of

the study described in Chapter 5 is well suited to further analysis of the co-occurrence-based component of the model. There are two reasons for this. Firstly, the study allows an investigation of the basic-level distinction between meaning- and position-based responses. The co-occurrence-based system contains both exemplars of words in discourse, and derived knowledge of the statistical probabilities of words being found together in text. This means that position-based responses should show a strong correlation with statistical measures of co-occurrence. Conversely, the semantic structure, while possibly influenced, in its early stages of its development, by co-occurrences between words (Kang, 2018; Prior & Bentin, 2008), is constantly being restructured through higher cognitive processes, such as categorising and sorting. This results in the reorganisation of words into semantic clusters which coincidentally reduce the system's connection to co-occurrence frequencies. As such, it is predicted that position-based responses will yield higher co-occurrence scores (e.g. mutual information, log-likelihood etc.: see Section 7.3.4) than meaning-based responses (although it is still anticipated that many meaning-based responses will co-occur, to a lesser degree, on account of their semantic similarity). Since there were no transitivity effects on the proportion of meaning- and position-based responses, cue transitivity is not important to this analysis.

Secondly, assuming that the directional divergence in position-based responses found for transitive vs. intransitive cues (i.e. transitive cues yield more cue-response associations, intransitives more response-cue ones) is derived from co-occurrence-based system rather than from semantic knowledge, it should be the case that the strongest collocates of a given word occur in the direction predicted by the response categorisation. For example, intransitive cues, given their greater tendency towards response-cue associates, should yield higher co-occurrence scores in their preceding than their following contexts (i.e. when a corpus search window of e.g. 0 to -4 is explored, compared to a window of 0 to +4). On the other hand, a non-composite, semantic view of WA suggests that the "position-based" category of responses is incorrectly defined: such associations are in fact derived from semantic memory, in spite of their superficial appearance as collocations.

Two experiments were conducted to test these hypotheses. The first investigated the following prediction:

responses in WATs" (p160). This is because the manner in which two words became associated is inconsequential to the system from which they are generated. An association learned through textual contiguity may subsequently achieve a stronger representation in the semantic system, by virtue of cognitive processing upon it. This does not mean that the processes which guide its production are not therefore contiguity-based.

- Associations categorised as position-based will have higher co-occurrence scores than responses coded as meaning-based, form-based, or erratic, because the co-occurrence system that activates and selects them is frequency-based.

7.3.2 Selection of data sources

The word association dataset examined in this study was the same as that used in Chapter 5, including the dual-coded categorizations used in the analysis of that data. Corpus data was from the British National Corpus (BNC; Consortium, 2007) and was analysed through the Sketch Engine (Kilgariff et al., 2014). The BNC was selected for a number of reasons. Firstly, several studies looking at corpus comparability with WA data have used this corpus (e.g. Bel Enguix et al., 2014a; Mollin, 2009; Wettler et al., 2005; Zareva & Wolter, 2012). This means that the findings of the current study will be comparable with those. Secondly, the BNC is sufficiently large (100 million words) to allow for the generation of co-occurrence scores for most cue-response pairs: since the lowest frequency words in the cue list had a frequency of around 1 occurrence per million words, this means that most words should occur at least 100 times in the corpus. Thirdly, the BNC samples from numerous discourse genres, and includes both spoken and written text. As such, it should be reasonably representative of most participants' linguistic experience. Finally, the BNC was collected from British language data, and therefore reflects the variety of English spoken by the (English L1, British university student) participants in the study reported in Chapter 5.

7.3.3 Corpus search parameters

To search this corpus for co-occurring associations, a window span of +/-4 words was selected. This is the same parameter used in earlier studies by Mollin (2009) and Zareva & Wolter (2012). It was selected, firstly, because a span of +/-4 words allows for verbs and their grammatical subjects/objects to be separated by several words (e.g. determiners, adjectives, adverbs) without being lost to the corpus analysis. This is important for the current study, since one explanation of WA transitivity effects is that they are attributable to knowledge of frequent subjects and objects of verbs. Secondly, it was assumed that the co-occurrence-based memory system is most sensitive to local relationships between words, rather than those separated by many words. While several studies have argued that wider spans, such as +/-20 words, offer better accounts of WA responses (e.g. Kang, 2018; Wettler et al., 2005), neither study has demonstrated this empirically, nor have they shown that the supposed superiority of these wide spans is not due to overlap with responses generated from the semantic system.

Corpus search queries also require the setting of parameters to determine, firstly, the minimum number of times which words (both the node and its collocates) must occur in the corpus in order to be returned by a search, and secondly the minimum number of times a word pair must co-occur in the corpus to be identified in a search. Recommended thresholds for these values (Evert (2004, 2008), for instance, suggests values of five and three respectively) are generally based on the needs of researchers aiming to identify important collocations in a corpus. The thresholds serve this purpose by eliminating noise generated by words and word pairs whose low frequency means that they are unlikely to be genuine collocations. For the present study, however, it was not so important to set strict thresholds, because the purpose of the corpus search was simply to locate those pre-specified pairs produced as associations in Chapter 5. Thresholds were therefore set a minimum of two occurrences of a word in the entire corpus, and one co-occurrence between word pairs. These values were set in order to allow all pairs observed in the WA data a chance to demonstrate a tendency to co-occur.

7.3.4 Selection of co-occurrence strength measure

Numerous methods exist for calculating co-occurrence strength scores, each of which handle distributional information in a slightly different way. This results in small biases towards certain types of words. Both mutual information (MI: Church & Hanks, 1990) and z-score, for example, tend to exaggerate the scores of low frequency words (Evert, 2008), while t-scores (see e.g. Hunston, 2002) can return inflated scores for grammatical words. The primary concern for the present experiment is to select a measure which demonstrates a close correspondence with WA responses. This is because the co-occurrence measure in the present study is being used as a proxy for human distributional knowledge – it will provide an estimate of how strongly associated a given word pair is likely to be within the human co-occurrence-based system. A close correspondence with existing WA data therefore gives the best indication possible that a given measure is effective in this capacity.

A further implication of this estimation of associative strength is that a measure of effect size, rather than of significance, is best suited to the current study. Measures of significance, such as t-score and log-likelihood, provide an estimate of how likely two words are to be associated. This allows researchers to deduce the statistical probability of a significant relationship between the words, but does not provide a measure of the strength of that relationship. Conversely, tests of effect size, such as mutual information, provide an estimate of how strongly associated two words are, rather than of the likelihood that two words are associated at all. Such measures are better suited to the present

study because of the focus on the strength of association between words, rather than the probability of their association.

Both Kang (2018) and Mollin (2009) have conducted correlation studies to test which co-occurrence score demonstrates the highest agreement with word association response data. The results of these two studies differ, with simple log-likelihood providing the most successful predictions of WA responses in Kang's study, and mutual information (MI) coming out on top in Mollin's work. Based on the data provided in these studies, and of the benefits of selecting a measure of effect size, MI was chosen for use in this chapter.

Two comments on Kang's findings should serve to justify this selection. Firstly, in identifying simple log-likelihood as providing the best fit with human WA data, Kang's study estimated co-occurrence strengths based on word co-occurrences across entire paragraphs. This is a far larger span than the window of +/-4 words used both in Mollin's study and in the present chapter. One of the effects of this difference in span is that a far higher number of co-occurring pairs are identified using the textual co-occurrence method. This is likely to influence the resulting co-occurrence measures, and may result in one method appearing superior to others. As such, Kang's results should be seen as demonstrating a superior fit for simple log-likelihood only where very wide spans of co-occurrence are being used; his findings should not be generalised to the closer contiguities explored in the present study. Secondly, Kang (2018, p. 104) found that while MI did not provide a close fit for WA data for the highest frequency words in his dataset, at frequency levels of around 50 occurrences per million words (fpmw) and below, the results for MI were comparable with those of other methods. Since the words in the current study have an average frequency of around 3 fpmw, with only a small handful of words exceeding 50, Kang's results do not point to a weaker performance of MI in human distributional knowledge in the current study. It is also worth noting that, in a large-scale review of numerous co-occurrence statistics, Evert (2008; 22) also recommended MI for use in psycholinguistic experiments due to its mathematical clarity and popularity in existing studies.

A brief description of MI will be given here. The statistic compares a word pair's observed frequency of co-occurrence with their likelihood of co-occurrence by chance. MI is formalised in the equation:

$$MI = \log_2 \frac{O}{E}$$

Where O is the observed frequency of the two words together, and E is their expected frequency – i.e. the likelihood that the two words will co-occur by chance given their individual frequencies and the size of the entire corpus. The above equation can be further elaborated as:

$$MI(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)}$$

where x and y are individual words in a corpus, $P(x, y)$ is their probability of occurrence together, and $P(x)$ and $P(y)$ are their probabilities of occurrence in isolation. Two words which have high probabilities of independence in a corpus (i.e. they are frequent words; e.g. *the* and *man*) will also have a relatively high probability of co-occurring simply by chance. MI therefore provides a measure of how frequently the two words actually co-occur in a corpus, compared with the probability of their co-occurrence by chance. An MI of zero reflects two words which occur exactly as often as predicted by chance. Higher MI scores reflect stronger co-occurrence, while negative values for MI indicate “repulsion” between two words – i.e. the presence of one predicts a greater than chance probability that the other will *not* be found in the surrounding context (Evert, 2008).

7.3.5 Procedure

Mutual information scores were calculated for all of the 1935 cue-response types produced during the experiment described in Chapter 5, using a span of +/-4 words. All data was then entered into a statistical package for purposes of analysis.

7.3.6 Results

The cue response pairs produced in Chapter 5 were grouped according to their basic-level categorisation (i.e. meaning-based, position-based, form-based, and erratic). Mutual information values for the word pairs within each of these categories were then compared. A statistically significant difference in MI values by category was revealed by ANOVA (mean MI for meaning-based responses = 1.17, SD=2.75; position-based mean = 2.22, SD=3.5, form-based mean = .56, SD=2.21, erratic mean = .58, SD=2.69; $F(3:1967) = 78.67, p < .001$). Post-hoc Tukey HSD tests revealed significant differences in MI values for all pairwise category comparisons, except for form-based vs erratic responses. These results are shown in Table 7.1.

Position-based responses had significantly higher MI than other response types, as predicted by the composite model. Meaning-based responses also yielded higher MI scores than either erratic or form-based associations, reflecting anticipated co-occurrences between semantically similar words. These results will be discussed further below.

Table 7.1

Post-hoc Tukey HSD tests of mean difference in MI values for four types of associative relation.

	Meaning	Position	Form	Erratic
Meaning		-2.607*	.946*	1.173*
Position			3.553*	3.78*
Form				.227
Erratic				

* = $p < .05$

7.4 Experiment 1b

The second prediction made of the composite model is that where cue-response pairs were coded as position-based, they should have higher MI values in the direction in which they were categorised (i.e. as cue-response or response-cue). Because the directionality of these categorisations was significantly influenced by the cue's transitivity, this prediction can be extended to refer to cue transitivity:

- Transitive cues will yield higher co-occurrence scores in the post-modifying range (i.e. following the node word in a corpus), while intransitive verbs will have higher values in the pre-modifying range (i.e. preceding the node).

The corpus, its parameters, and the co-occurrence measure used in this experiment were the same as those above, with the exception that different window span settings had to be calculated to allow the investigation of directionality effects. Two MI scores were calculated for each cue-response type. These were directional calculations measuring the frequency with which a response occurred in the context preceding its cue (i.e. within a window span of 0 to -4 words), and its frequency of occurrence in the cue's following context (i.e. 0 to +4 words). These latter calculations were used in the directional analysis pertaining to cue transitivity effects.

7.4.1 Results

The new directional MI values were compared as a function of cue transitivity. Because tests for normality of distribution showed that the data was negatively skewed (Kolmogorov-Smirnov test, $p < .001$), Mann-Whitney U tests were used as a non-parametric alternative to t-tests. The test revealed differences in MI values approaching significance as a function of cue transitivity in the pre-modifying

range (MI to transitive position-based pairs mean rank = 212.19, intransitive mean rank = 231.45; $U = 21191$, $p=.075$), and significant differences in the post-modifying range (MI to transitive position-based pairs in post-modifying context mean rank = 229.53, intransitive mean rank = 206.03; $U = 20741.5$, $p=.043$). All of these differences were in the direction predicted above.

One possible reason for the non-significant nature of the difference in the pre-modifying range is that, as discussed in Chapter 5, some associations appeared to be based on interpretations of cues as nouns (e.g. the cue *chuck* received responses including *berry/bass/Rugrats; glow-worm/stick*), while others were completions of frequent phrasal verbs (e.g. *fall-down; rise-up/above/again*). This latter type of response was particularly common for intransitive cues, and may have distorted the MI distributions of responses because it bucked the general trend for pre-modifying responses to intransitives. For that reason, a new Mann-Whitney U test was conducted after the removal of these cue-response pairs. These tests revealed statistically significant differences in MI in both of these new tests (MI to transitive position-based pairs in pre-modifying range mean rank = 189.17, intransitive mean rank = 209.69; $U = 16322$, $p=.048$; MI to transitive position-based pairs in post-modifying context mean rank = 206.11, intransitive mean rank = 182.24; $U = 16011$, $p=.031$).

These results again support the predictions of the composite model: any response which is both transitive and position-based in nature should be more likely to occur in the context immediately following its cue (with the opposite pattern for intransitives), reflecting activations from co-occurrence-based memory.

7.4.2 Discussion

The results of Experiment 1 provided evidence in favour of a composite view of WA, in which WA responses are viewed as deriving from three interacting cognitive systems – semantic, co-occurrence-based, and form-based. Since it is assumed that responses coded as position-based are the result of activations in the co-occurrence-based system, which draws on information hypothetically similar in nature to statistical co-occurrence measures such as MI, and generates activations for the word(s) with the strongest relationship of co-occurrence with the cue, it was predicted that position-based responses would have higher MI scores than responses from other categories. This prediction was supported in the first part of Experiment 1. In addition, it was further predicted that the directional differences in WA responses to transitive and intransitive cues would also be reflected in higher MI scores in the direction predicted by a word pair's coding. This is because these responses are drawn from the same co-occurrence-based system, which is assumed to be sensitive to information not only about the frequency of co-occurrence between words, but also their directionality. For example, if the

most reliable collocational relationships for intransitive verbs are words from their preceding context (e.g. their grammatical subject) rather than from the following context, it follows that these words will achieve activation most quickly, because of the strength of their relationship in the co-occurrence-based system. This should be reflected in higher MI scores in the direction predicted by the word pair's coding. Again, the results of Experiment 1 supported this prediction.

Initial inspection of these results, then provides support for the composite model. A closer inspection of the data presented above, however, reveals a problem: the standard deviation figures from Experiment 1a (reproduced in Table 7.2) are, in all cases, much higher than the mean. Indeed, it is only for the position-based results that the SD is less than double the mean. What this means in practical terms is that there will have been numerous meaning-based responses, as well as some form-based and erratic ones, which had higher MI values than some of the position-based responses. To some extent, this is anticipated in the composite model, since it is assumed that an association between two words may be represented in more than one memory component, and that convergence of more two or more memory components on a single response makes their production more likely. However, the scope of the problem suggests that high MI values overlap with non-position-based coding much more than might be predicted if we are assuming that frequency of co-occurrence is the main determinant of position-based responding. In total, 56 of the 100 word pairs with the highest MI values in the dataset were not coded as position-based responses (some examples of meaning-based pairs with very high MI values are *kneel-genuflect*, *devour-gluttony*, and *dawdle-dally*).

At the opposite end of the scale, 178 associations coded as position-based received MI values of zero (of a total of 1295 such pairs). These 178 responses cannot have originated from co-occurrence-based memory if their true probability of co-occurrence is zero, since they would not be represented within that system (and much less would they be able to compete for activation therein). These problems call into question the extent to which the above findings can be seen as supporting the composite

Table 7.2
Mean and standard deviation values for MI scores across 4 basic categories

	Mean	SD
Meaning-based	1.17	2.75
Position-based	2.22	3.5
Form-based	.56	2.21
Erratic	.58	2.69

model because they reveal that responses apparently (i.e. according to their categorisation) originating from the co-occurrence-based system may not in fact have done so. Instead, it appears that although *some* WA responses have higher MI than others, it can scarcely be said that

it is their position-based nature that differentiates these responses from others, since the mapping of position-based categorisations to MI is unclear.

This lack of clarity can be explained in one of two ways. Firstly, it may be that the mutual information values returned by the above analysis are not, in fact, similar to the statistical knowledge contained in human co-occurrence-based memory. However, given the extent to which previous researchers have supported the use of MI in existing psycholinguistic experiments (Evert, 2008; Mollin, 2009), this is considered the unlikelier of the two possible explanations. Secondly, it may be that the human codings produced in Chapter 5 do not map onto the processing route followed by the participants in the present study as closely as had been assumed. If this turns out to be the case, there could be significant implications for interpreting all WA results.

In order to investigate this latter possibility, a new experiment was carried out. This experiment provides a more sensitive measure of the relationship between position-based coding and mutual information. It is hypothesised that, if (a) position-based response generation and position-based response categorisation originate from the same co-occurrence-based cognitive structure as each other; and (b) MI provides a reasonable proxy for the information used in this system (as was assumed in Experiment 1), then cue-response pairs coded as position-based should, for the most part, be predictable from their mutual information scores.

7.5 Experiment 2

7.5.1 Aims and predictions

This experiment aims to test the prediction that binary position-based vs. non-position-based codings can be accurately predicted using their mutual information scores. In doing so, the experiment tests the assumption that both generation and categorisation of position-based responses is derived from a single cognitive structure which draws on MI-like information in order to generate word activations during the WA task.

7.5.2 Method and results

The same WA and corpus-derived data as used above were entered into a binomial logistic regression analysis. This test works by using one or more predictor variable to determine which of two categories of a single dependent variable each trial should be assigned to. In the current analysis, the predictor variable was a cue-response pair's MI, and the two categories were position-based and non-position-based (containing meaning-based, form-based, and erratic associations). Each trial consisted of a single cue-response pair being entered into the analysis. The model analysed the MI score for each

pair, and generated a categorisation prediction. The success of the model was judged on the extent to which these predictions corresponded with the human categorizations discussed in Chapter 5.

The logistic regression model was statistically significant (BNC4 $\chi^2(1) = 197.75, p < .001$). However, the amount of variation in human codings explained by MI proved to be small (BNC4 = .138), corresponding to just under 14% of variance in human codings. Further description of the model's performance is given in Table 6.26, which shows the various ways in which binomial logistic regression results can be broken down. The correct classification statistic shows that the model arrived at the same classification as the human coders around 70% of the time. While this indicates a high level of overlap between human and MI-based coding, other measures give a more nuanced view. Note that the total number of position-based predictions made by the model will be equal to the number produced by the human coders (438). This means, in effect, that the model chooses the 438 pairs with the highest MI for coding as position-based. The specificity statistic shows high agreement between

	BNC +/-4
Correct classification	73.5
Sensitivity	22.5
Specificity	92.8
Positive Predictive Value	54.22
Negative predictive Value	75.95

human and MI-based codings for cue-response pairs coded as non-position-based. This suggests that the relatively high correct classification score was largely down to high agreement on these pairs. This was probably the result of the many cue-response pairs which

received MI scores of zero: both the human coders and the MI-based model typically categorized these as non-position-based. However, the Sensitivity statistic demonstrates that the model showed far less overlap with human categorisations of pairs which *were* coded position-based, with only 22.5% of such responses being coded the same way by the MI-based system. This suggests that the human coders placed many associations into the position-based category in spite of the fact that they did not have MI values inside the top 438 values returned by the MI analysis.

This picture of higher correspondence between human and regression-based coding of non-position-based pairs is also apparent in the positive and negative predictive value statistics. These show that, of all cases predicted (according to the model's reading of the MI data) to be non-position-based, almost 76% were in fact coded this way by the human coders. However, for the pairs predicted to be coded as position-based, only 54% were actually coded this way. This again demonstrates the coders'

apparent lack of reliance on co-occurrence-based knowledge in order to determine position-based coding.

7.5.3 Discussion

Experiment 2 aimed to assess the extent to which position-based categorisations can be explained by mutual information scores, which were taken to be a proxy for the co-occurrence-based knowledge used by human coders. The results suggest that MI-based models do not offer a good fit with human categorisation data because the two methods frequently differ in the pairs which they categorise as position-based. The findings call into question those of Experiment 1, because they suggest that the underlying assumption that human codings provided a realistic estimate of the psycholinguistic origin of associations is unsound.

These results can be incorporated into the composite model of WA, since the model anticipates that many responses will be attributable to activations in more than one component of memory (semantic, co-occurrence-based, and form-based). On the other hand, these findings suggest that empirical tests of the predictions of WA models such as the composite model cannot safely be based on human codings of position-based responses, since the psycholinguistic underpinnings of the codings themselves now appear to be unclear.

It is possible, however, that methodological issues contributed to the low predictive power of the model. For example, the choice of corpus and/or its parameters might have given a poor match with the co-occurrence-based knowledge drawn on by coders. In order to test this hypothesis, the experiment was re-conducted using three new conditions. The first re-test addressed the window span issue by using the same corpus (i.e. the BNC) with a different window span to that used above. The other two used a new corpus, with both previously used spans, in order to test this factor. A span of +/-2 words was selected for the new analysis, since it was hypothesised that human coders may be most likely to select cue-response categorisation for those responses which co-occur in near-immediate contiguity with the cue. A span of 2 words allows only one word, such as an interposed determiner, to separate the cue from the response. The corpus selected was the TenTen English corpus (TT; Jakubiček et al., 2013). This corpus was chosen for two reasons. Firstly, its great size (around 22 billion words) allows for the testing of the possibility that the BNC was not large enough to reflect the linguistic experience of the test participants; and secondly its recency (it is created through web-based text-extraction, meaning that it will include very recently written texts in addition

to older ones) may offer a closer fit to the age of the coders in the study, who were not born at the time that much of the BNC text was created.

The same binomial logistic regression method was used to test the capacity of these new corpus conditions to predict human codings. In all corpus conditions, the logistic regression model was statistically significant²⁰, (British National Corpus, span +/-2 (BNC2) $\chi^2(1) = 184.34, p < .001$; TT2 $\chi^2(1) = 136.38, p < .001$; BNC4 $\chi^2(1) = 197.75, p < .001$; TT4 $\chi^2(1) = 114.40, p < .001$). However, none of the new models was able to explain as much variation as the BNC +/-4 model presented above (BNC2 Nagelkerke $R^2 = .129$, TT2 = .097, BNC4 = .138, TT4 = .082).

Specific details of the performance of each corpus condition are set out in Table 7.4. These suggest, firstly, that in spite of the lower amount of variation in categorization achieved by the BNC2 model, it may have in fact slightly outperformed the BNC4 model in terms of its ability to predict human classifications. In particular, the BNC2 model demonstrated noticeably greater positive predictive value (reflecting a greater proportion of occasions on which the MI-based predictions of the model agreed with the human codings with regard to position-based responses) than the other models. Together with the similarly higher positive predictive value of the 2-word span than the 4-word span in the TT-based analysis, these results suggest that human coders may be more likely to categorise words as position-based if they are closely contiguous than if they tend to co-occur at a greater distance.

Table 7.4:
Correct classification (position-based vs non-position based) of cue-response pairs by corpus condition (%)

	BNC-2	TT-2	BNC-4	TT-4
Correct classification	74.8	69.9	73.5	70.5
Sensitivity	24	7.7	22.5	4.8
Specificity	94.1	93.5	92.8	95.5
Positive Predictive Value	60.47	31.11	54.22	28.57
Negative Predictive Value	76.54	72.77	75.95	72.55

The results also suggest that using a larger and more modern corpus (i.e. the TenTen corpus) did not result in a better fit with the data. While the TenTen corpus performed comparably with the BNC with regard to the classification of words coded as non-position-based (as per the specificity and negative predictive value statistics), the percentage of overall correct (i.e. overlapping with human codings) classifications was lower than those of the BNC models. The reason for this is (as per the sensitivity

²⁰ Note that the data for the BNC4 condition is the same as that presented above.

and positive predictive value statistics) that the TT-based models performed poorly with regard to identifying pairs which were human-coded as position-based. This may be because the large size of the TT corpus, rather than providing an improved fit for the data as suggested above, actually meant that it yielded a larger number of non-zero MI values than the BNC, and subsequently (incorrectly) predicted that these would be coded as position-based.

In sum, however, this data does not suggest that further changes in corpus or corpus parameter settings will produce a markedly better fit with human categorizations. This is because none of the new models made anything more than marginal gains in predictive power.

Another possible explanation for the low predictive power of the model may be the use of MI to measure co-occurrence frequency. As discussed in Section 7.3.4, Kang (2018) found that simple log-likelihood provided a better fit for his WA data than did MI. While this possibility is worthy of further research, however, it is unlikely that such changes will result in markedly different results. There are two reasons for this. Firstly, the differences in fit with WA data provided by different co-occurrence measures were, in Kang's research, very small. For example, in his sentence-level analysis (the parameter closest to the current study), the difference in the percentage of top WA responses found within the 100 most strongly co-occurring words, as selected by five different co-occurrence measures was just 2.1% (varying from 41.6% to 43.7%). Secondly, as discussed in Section 7.3.4, Mollin's analysis suggested that other co-occurrence measures would be unlikely to provide a better fit for WA data than mutual information. As such, it appears unlikely that the data extraction methods used in this chapter are solely responsible for the poor performance of the MI-based models in predicting human position-based categorisations.

These results, then, point to two possibilities. Firstly, it may be that the composite model is simply incorrect in its predictions. Alternatively, it may be that the model is correct, but the assumption that human codings reveal the psycholinguistic basis of position-based responses is not valid. This implies that the results of Experiment 1 cannot be seen as evidence for the use of co-occurrence-based knowledge in position-based response generation because there appears not to be a sufficiently reliable relationship between co-occurrence score and position-based coding. In other words, position-based categorisations of word pairs as captured by human coders do not reveal the underlying psycholinguistic processes used by respondents.

It is important to note that these results do not call into question the existence of the transitivity effects found in Chapter 5. These effects were found at the detailed level pertaining to their

directionality. While the findings of Experiment 1b, which probed the relationship between directional coding and directional MI, made use of the same human codings which now appear unsuitable as a proxy for response generation processes, this does not call into question the fact that human coders perceived a difference in responses to transitive vs. intransitive verbs: it only questions the origin of this difference. Alternative interpretations of these transitivity effects will be discussed in the remainder of this thesis.

An important implication of the results of Experiment 2 is that support for the composite model will need to be sought using methodologies which do not rely on subjective coding of responses. This goal will be further pursued in the following chapter. Before moving on, however, it is worthwhile to explore some of the implications of the results of Experiment 2 for interpretations of the way in which WA responses are categorised. This discussion entails identifying similarities and differences between the processes which guide response production, and those used by coders as they categorise word pairs.

7.5.3.1 Implications for the coding of WA responses

If position-based responses are not coded based on statistical probabilities of co-occurrence, exactly how are they coded? In addressing this question, it is useful to explicate the view of position-based WA categorisation implied in the composite model described above. This process was assumed to be analogous to the processes used in response generation: coders would, upon perceiving a word pair, automatically recreate the processes by which the original association was made. They would then deduce the component of memory chiefly responsible for the association.

Here, it is helpful to develop the assumptions of the composite model with regard to the automaticity of processing. In Section 7.2, it was suggested that conscious processing may come online at some unspecified point during the latter stages of response selection. This activation of conscious processing is helpful in attempting to explain the differences between response generation and categorisation. It is assumed here that response generation and categorisation are, at least in their initial stages, automatic processes. This automaticity is viewed as being contingent upon processing speed: if a response is not generated within a given time period, conscious processes become active, and aid in the selection of a response. At this point, an interesting (hypothetical) divergence between response generation and response categorisation occurs. In the case of response generation, fully automatic processing may be most likely when multiple forms of word knowledge (semantic, co-occurrence-based, form-based) converge on a single word (Fitzpatrick & Izura, 2011). This is because a response reaches a selection threshold before conscious processing comes online. In the case of categorisation,

however, such responses are likely to be the most challenging to code, and thus the most likely to require conscious processing, since categorisation requires a decision to be made as to the most likely processing route for each pair. In psycholinguistic terms, this is a very unnatural decision to be faced with, since it was likely the convergence of several routes, not one of them alone, which yielded the association.

A converse phenomenon occurs for responses which originate from activation in only one memory system. In such cases, the generation of a response would be slower than when converging activations occur. This would be partly due to weaker activation of each candidate word, and partly also to competition for selection: if the three memory systems yield activations of three (or more) different words, then conscious processing will be needed to select from among them. For the coder, on the other hand, the categorisation of these responses is likely to be less complicated, since information of only one type has been used by the test-taker, and this should be apparent to the coder.

In summary, then, this theory would predict fast generation of pairs such as *grumble-mumble/groan*, and *glisten-gleam* (see Section 7.4.7), as well as similar pairs such as *waddle-dawdle* and *dawdle-dally*. However, it would predict slow and unreliable *categorisation* of these pairs. These contrasting predictions both arise, paradoxically, from the same multiple sources of similarity between the words – i.e. that they all share form-based and semantic features, and all share high mutual information. Anecdotal support for this interpretation in the current data comes from the coding of the above examples: the first three of these pairs were coded as form-based, and the other two meaning-based, despite their similarities.

Conversely, slow generation of responses should occur when the three processing routes generate different (but similarly strong) candidates for selection. An example is the cue *relieve*, to which the top responses were *pain* (position-based) and *relief* (form-based). These words were produced 8 and 6 times respectively, suggesting similar strength of association. As such, it could be expected that the competition for activation between them would slow the production of the final response. For the coder, however, neither word should be too difficult to categorise (although both have semantic relations to the cue).

However, such an explanation is not without problems. Firstly, this view makes the same assumptions as the composite model of WA generation, discussed above, regarding the (equal) influence of semantic, co-occurrence-based, and form-based systems on responses. As such, the present explanation is open to many of the same criticisms, such as the fact that far fewer form-based

responses tend to be given than other type of response. Secondly, since the explanation of coding given here involves the coder attempting to psycholinguistically retrace the footsteps of the respondent, it implies that coders can deduce the processes by which they arrived at an association. There are numerous reasons to doubt this. One WA-specific piece of evidence is provided by studies which have included post-WAT interviews with participants to aid coders in the categorisation process (e.g. Fitzpatrick, 2006; Higginbotham, 2010; Wolter, 2001). This task is similar to that faced by coders, although it differs in that it follows the actual generation of a response, rather than only a decision on an existing pair. One study which has reported the findings of such interviews is Wilks (2009). Her respondents claimed that they made use only of semantic information in generating their responses, describing their associations as synonyms, definitions, or opposites (Wilks, 2009, p. 33). When asked about apparently position-based responses, some respondents stated that they sometimes chose words which “go together”, or “nouns that go with a particular verb” (*Ibid.*); however, they remained insistent that these associations had come from thinking about the meaning of the cue. The vagueness of these descriptions is problematic for a view of response coding based on conscious awareness of lexical processes (although the respondents’ assertions that their responses were semantic in origin is tantalising - Wilks (*Ibid.*) suggested that respondents appeared to consider position-based responses “as a sub-category of meaning” – something which jars with the assumptions of the composite model). Finally, it is also problematic to assume that the process of categorising a cue-response pair can follow the same processing route as its generation. Since both cue and response are immediately available to the coder, it is unlikely that any process of activation analogous to that experienced during response generation – particularly pertaining to competition between candidates for selection – can occur.

This discussion suggests that further research into the differences between the generation and categorisation of responses might provide a fruitful way of looking at the processes involved in WA. A starting point for such research might be to compare response times for response generation vs. categorisation. If the above theory is in any way accurate, there should be negative correlations in RT between these tasks.

Two further points can be made with regard to potential coding effects. Firstly, some models of response categorisation have attempted to alleviate the difficulty of coding associations which have multiple relations by allowing coders to use dual-route categories, such as the “meaning and position” and “meaning and form” categories used by Fitzpatrick and Izura (2011). However, the analysis of inter-rater reliability ratings presented in Chapter 5 suggested that such categories result in

particularly inconsistent coding. In that chapter, 52 associations were placed into a dual-route category by at least one of the coders, but only 10 of these were placed into that category by both of them. In total, dual-route associations accounted for 20% of all coding disagreements, despite comprising only 6% of all codings. One reason for this (Fitzpatrick, personal communication) could be that noticing one possible link between words may impair coders' ability to see another. This may be due to some process of psycholinguistic inhibition, or possibly due to coders simply wishing to categorise their data quickly, and consequently coding a word pair using the first link which comes to mind. Whatever the basis of this difficulty, however, it suggests that the use of dual-route categories does not straightforwardly resolve the issue of coding difficult associations.

This latter point also raises the issue of potential coder bias. Fitzpatrick's work on individual WA response preferences (Fitzpatrick, 2007, 2009, Higginbotham, 2010, 2014; and see Chapter 2) has strongly suggested that individual respondents consistently select some types of response more frequently than others. The present study, with its conclusion that position-based responses do not appear to be straightforwardly related to word distributions, does not rule out the possibility that such preferences also apply when individuals are categorising responses. Further research is needed, firstly, to establish whether such preferences apply during coding, and secondly to explore the extent to which they can be attributed either to automatic processes (i.e. the dominance of one cognitive system over another) or conscious selectional ones (such as a habitual sense that some types of response are more likely than others).

7.6 Conclusion

This chapter has developed and described a model of response generation in the word association task. This model is broadly usage-based in nature, and proposes that three interacting memory components – semantic, co-occurrence-based, and form-based – independently generate association candidates which then compete for activation.

Support for this model was then sought through an experiment in which corpus-derived co-occurrence measures were calculated for responses assumed to originate from each of these three memory components. However, while the initial results of this experiment appeared to support the predictions of the model, the second half of the chapter explored one problematic issue raised by the first experiment: the equating of psycholinguistic processing routes during response generation with the later coding of those responses by independent coders. A second experiment found that the relationship between mutual information scores and position-based categorisations, though

statistically significant, did not capture a large amount of variation in coders decisions. This was found to be because coders did not always categorise those word pairs with the highest MI as position-based associates. Alternative MI-based models suggested that larger corpora do not improve the fit to human coding data. However, there was a suggestion that smaller window sizes result in greater similarity between MI and human codings, suggesting that coders may consider very closely contiguous responses as better candidates for position-based coding.

The results presented in this section do not directly question the main assumption of the composite model (i.e. that three independent but interacting processing routes contribute to WA). However, because of the confound revealed above, the results of Experiment 1, above, cannot be taken as evidence for the model. While it is clear that some cue-response pairs do share higher statistical co-occurrence than others, the work presented here does not constitute evidence that such associations originate from a co-occurrence-based memory component. Future experimental studies aiming to demonstrate this processing route will need to do so without recourse to human categorisation data.

In general, the findings presented in this chapter are somewhat reminiscent of those of other investigations into the fit between co-occurrence data and word association responses (e.g. Bel Enguix et al., 2014a; Kang, 2018; Mollin, 2009; Wettler & Rapp, 1993; Wettler et al., 2005) in that they suggest some overlap between these types of data, but ultimately fail to capture a majority of variation. They imply that a closer examination of the differences between corpus and WA data is needed. This will be another of the aims of the next chapter.

Finally, although neither the composite model, nor the existence of WA transitivity effects, have been directly challenged by this data, it is worth looking at how both fully co-occurrence-based and purely semantic models would explain the findings presented above. Firstly, it would be difficult to explain the results of this chapter using a strongly co-occurrence-based account. This is because, as mentioned in Section 7.3.2, a large number of cue-response pairs shared a mutual information score of zero, even when the very large TenTen corpus was used to generate MI values. This is not predicted by a co-occurrence-based model, and continues the trend, discussed in Chapter 6, for such models to fail to adequately account for substantial proportions of WA data.

A semantic model (e.g. McRae et al., 2012; Mollin, 2009), on the other hand, would explain these findings by arguing that the overlap between mutual information and position-based response categorisations is not, in fact, related to human knowledge of co-occurrence, but is instead symptomatic of semantic differences in response types. It would hypothesise that the types of

semantic relations which diverge as a function of transitivity (e.g. verb agent, patient, and experiencer relations) are more likely to co-occur within the spans investigated in this study than other spans (which might nevertheless co-occur within wider spans – e.g. synonyms, antonyms, situational and event-related associations etc.). The opposing predictions of fully contiguity-based or semantic models of WA will be returned to in Chapter 9.

Chapter 8: What can usage-based statistics reveal about word association?

8.1 Introduction

In Chapter 5, an influence of verb cue transitivity on word association responses was discovered: transitive cues yielded more cue-response-type position-based associations than intransitives, while intransitives elicited more response-cue links. In Chapter 6, it was suggested that a usage-based approach offered a better fit for this data than earlier generative models. Chapter 7 made such an approach explicit. It described a model of WA response generation, termed the composite model, in which three components of memory – semantic, co-occurrence-based, and form-based – all generate activations of words related to the cue. As such, responses of a given type were viewed as originating from the corresponding cognitive system. However, initial support for this theory suggesting that position-based responses are derived from co-occurrence-based knowledge, described in Experiment 1, was subsequently found to be unreliable due to the unpredictability of the human codings upon which it depended.

The current chapter will therefore investigate alternative methods for testing word association (WA) models. Firstly, since the psycholinguistic process of categorising responses seems to be independent, or at least different, from those used in generating them, it is probable that categorisations by human coders are not an accurate reflection of what the test takers were doing. As such, an attempt will be made here to test the composite model without relying on human response categorisations. This will involve the concept of *entrenchment* and the related statistical measure of *relative entropy*, both of which are important in usage-based theories of language.

Secondly, Chapters 6 and 7 both suggested that strongly co-occurrence-based models of WA are unable to explain large amounts of WA response data. Few studies, however, have looked into the qualitative differences between these types of data – for example by looking at the nature of those collocating word pairs which are not frequently selected as WA responses. A secondary aim of this chapter is therefore to make use of corpus analysis to identify dissimilarities between corpora and WA responses.

8.1.1 Entrenchment, relative entropy, and word association

The failure of human codings to provide a reliable estimate of WA response generation processes discussed in the previous chapter necessitates the development of alternative methods for analysing word association responses. One potential source for such methods is usage-based (UB) models of

language, since in many cases these models make detailed proposals about the way that linguistic experience should structure linguistic knowledge.

Usage-based theories were described in Chapter 6, where it was suggested that they assume linguistic knowledge and competence to be the result of the combination of two factors: a person's linguistic experience, and the domain-general processes which act upon it. These processes include the perception and production of linguistic units, the subsequent recognition of constructions (i.e. form-function pairings of any size, including morphemes, words, and phrases; Fillmore, 1988; Goldberg, 1995, 2003), and higher processes of categorisation, generalisation, and abstraction of these units (Behrens, 2009).

One specific linguistic phenomenon explored by UB theorists – that of type and token frequency – is particularly interesting from the perspective of WA research. The starting point for an understanding of these measures is the view that individual constructions can vary in their lexico-syntactic realisation, even while they remain manifestations of the same construction. This is because constructions have a prototypical structure in that they admit variation in form without compromising their semantic and pragmatic unity (Hanks, 2013, p. 399). As such, several form-meaning pairings (e.g. *Have you got a minute?* *Do you have a minute?* *Have you got a moment?* etc.) can be viewed as slightly different manifestations of a single prototypical construction, which retains a single semantic and pragmatic function across its various orthographic and phonological manifestations.

Type and token frequencies play an important role in the formation of these flexible constructions. Token frequency refers to the number of times a construction occurs in one specific form in the linguistic experience of a given individual. For example, *Have you got a minute?* occurs (in exactly this form) 5 times in the British National Corpus, and therefore has a token frequency (in this corpus) of 5. On the other hand, type frequency refers to the number of forms a construction can take without ceasing to be manifestation of that construction. All attested variations of a construction (such as *Do you have a minute?* or *Have you got a moment?*) which do not change its semantic or pragmatic function therefore sum to the type frequency of the construction. For example, if a corpus search revealed that the *Have you got a minute?* construction was manifested in only the three ways suggested above, it could be said to have a type frequency of three²¹.

²¹ It is not altogether straightforward to say whether two similar forms are exemplars of the same construction, or altogether separate constructions – at some point a boundary must exist between forms which are exemplars of the same construction and those which manifest the same pragmatic function but are not part of the same

These two features of linguistic distribution have different effects on the way that language is stored in the mind. Token frequency is a key aspect of the entrenchment of *exemplars*, which are cognitive representations of specific linguistic forms. Exemplars are created through linguistic experience, such that perceiving a given construction for the first time creates an exemplar which may subsequently be strengthened through repeated encounters with an identical form (Bybee, 2006; Bybee & Beckner, 2010). As such, high token frequency (i.e. frequent encounters with an identical construction) leads to *entrenchment*, which refers to the increasing stabilization of a given form in the mind (Behrens, 2009; Bybee, 2006; Tomasello, 2003). Encounters with a construction which has a single manifestation, with one meaning, for one pragmatic function, and in only one context, would lead to strong entrenchment between this form and its meaning, pragmatic function, and context of use.

High type variation counteracts this process of entrenchment, leading instead to the *abstraction* of a general category, or schema, over a range of similar forms. When the mind encounters a phrase such as *Have you got a minute?*, an exemplar is created, and subsequent encounters with exactly this form will strengthen its representation as described above. When comparable forms such as *Do you have a minute?* are encountered, they are stored in close proximity to the earlier construction. Repeated exposure to all comparable forms will provide evidence which the mind will use to determine the similarity of the constructions. In the case of the examples regarding *Have you got a minute?*, the result will be the grouping of these exemplars into a single schema which compiles the general features of these utterances (Behrens, 2009). Abstraction is therefore the process of the assimilation of various formal manifestations of a construction into a single schema. A side-effect of this process is the weakening of the association between any one form (e.g. *Have you got a minute?*) and its meaning, pragmatic function, or context of use, as the language user comes to see that these functions can be realised in more than one way (Bybee, 2003). This abstraction is both predictive and semantic: a word which has a strong, exclusive collocational relationship to another word (e.g. *pestle*, in *mortar and pestle*) will additionally become semantically entrenched with that word, potentially to the extent that the two words come to represent a single semantic unit (it is not easy to determine which is the mortar and which the pestle!).

The present chapter will investigate the extent to which the effects of type/token distribution on the psycholinguistic processes of entrenchment and abstraction influence word association responses. The subject of this investigation will be the 98 verb cues used in Chapter 5. It will be assumed here

construction. This complex matter will not, however, be discussed here, since the technicalities of what is and is not a construction are of only minor concern to the discussion which follows.

that each of these verbs is a construction (e.g. Bybee, 2010; Evans, 2007; Goldberg, 2003) whose meanings and syntagmatic associations are shaped (at least in part) by the type/token distributions of the nouns with which they co-occur. While such a view is rarely explicitly stated in the literature on construction grammar and usage-based linguistics, it is implied by the description of type frequency given by Gries & Ellis (2015, p. 234): “the type frequency of an element is the number of different types that the element co-occurs with”. That such an effect could pertain to collocational pairs is supported by Hanks (2013, p. 113), who claims that “the semantics of each word in a language is determined by the totality of its complementation patterns”. This view can be seen in turn as supporting the theory that the type/token distribution of a verb’s collocates will influence its psycholinguistic representation, since it implies that verbs which have very narrow patterns of complementation (i.e. low type variation) should also have narrow semantic ranges and relatively entrenched relationships – semantic and collocational – with each word with which they co-occur. This situation can be contrasted with that of “light” verbs, such as *take* and *make* (*Ibid.*; 287), which have very broad type variation, and therefore abstract semantic representations and low levels of association with any specific collocational complement²².

The hypothesis explored in this chapter is therefore that the type/token distributions of each cue’s collocational patterns will influence various aspects of that verb’s WA responses. Before describing existing research of relevance to this view, however, it is useful to consider how type/token distributions can be described statistically.

8.1.2 Statistical approaches to entrenchment and abstraction

In the previous chapter, statistical measures of co-occurrence such as mutual information (Church & Hanks, 1990) were used to give an indication of how likely two words are to co-occur in a corpus, as a function of the total corpus size. A high mutual information score shows that two words are much more likely to co-occur in a given corpus than would be predicted by chance. What this measure does not reveal, however, is how many other collocational pairings a given word is also likely to occur in: to say that Word A shares high mutual information with Word B does not preclude similarly strong relations between Word A and Words C, D, E etc. In other words, MI and related statistics (e.g. log-likelihood, *t*-score etc.) provide measures of the strength of association between two words in text,

²² This is not to deny that individual verbs (e.g. *run*) are polysemantic: many verbs can form the basis of multiple constructions (e.g. *run a business/company* is a different construction to *run a race*). This is because the formation of a construction is based on the perception of semantic similarity between phonologically or orthographically similar forms.

but do not give a precise indication of the *exclusivity* of that relationship. It is this latter measure which is required to facilitate the investigation of type/token distributions on linguistic knowledge.

A measure which can provide this holistic viewpoint is relative entropy (H_{rel} : Gries, 2010; Gries & Ellis, 2015), which provides a measure of the entropy (orderliness, or predictability) of a distribution. The raw material for this measure is token counts for all types of interest. For example, assuming that we are interested in the noun collocates of the verb *eat*, we would search a corpus for all of the words which collocate with *eat* (within a given span, and with a predetermined minimum frequency). The total number of these collocates is the type frequency for *eat*'s verb-noun pairings. Additionally, we would record the token frequency for each collocate. The relative entropy measure compares these type and token distributions. It outputs a score between zero and one. A score of 1 would imply that the distribution is maximally disordered. This occurs when each type in the distribution has exactly the same token count, and is therefore equally likely to co-occur with the node. In the terms set out in Section 8.1.1, this would lead to psycholinguistic abstraction, rather than the entrenchment of any one pairing. On the other hand, a score approaching zero indicates that the distribution is very orderly and predictable. This occurs when the distribution is dominated by a single type whose token count far exceeds any others. This type of distribution would encourage entrenchment, rather than abstraction.

Relative entropy is derived from the entropy statistic, which is formalised by Gries (2010) as:

$$H = \sum_{i=1}^n (p(x) \cdot \log_2 p(x)) \text{ with } 0 \cdot \log_2 0 = 0$$

Relative entropy normalises this measurement along a scale of 0 to 1, according to the following formula:

$$H_{rel} = H/H_{max} = H/\log_2 n$$

8.1.3 Type/token distributions and word association

Few existing studies have looked into the relationship between the entropy of word distributions and word association responses. One such study was, however, carried out by Hahn & Sivley (2011), who, in a series of experiments, explored correlations between the predictability of a word's collocational relationships and aspects of its forward word association distribution. Hahn and Sivley collected recurring word sequences (i.e. *n-grams*) of between 2 and 5 words in length from a dataset of 95

billion sentences extracted from the Internet. In order to qualify for inclusion in the study, each n-gram had to occur a minimum of 40 times in the dataset. Since the total size of the resulting dataset was too large to be computed (more than 4 billion n-grams), the authors limited their investigation to those n-grams which contained one of 62,474 word pairs found in the South Florida word association norms (D. L. Nelson et al., 2004).

In their first experiment, Hahn and Sivley (2011: Experiment 1) calculated the probability that a given word (each cue from the South Florida WA norms), when occurring as the first word in an n-gram, would be followed by another specific word (i.e. that cue's response from the WA norms) somewhere within the same n-gram. The authors then calculated correlations between these probabilities and the forward association strength of each cue-response pair. Across the entire South Florida norms list, they found a weak ($r = .151$) but significant ($p < .0001$) correlation between these measures – a finding not dissimilar to earlier results relating to the overlap between word distributions and WA responses, which (as discussed in Chapter 6) have tended to suggest that corpus data overlaps to only a moderate extent with WA response data.

In a follow-up experiment, Hahn & Sivley (2011: Experiment 4) investigated the relationship between word distributions and word association responses in a more holistic manner. They examined the extent to which the presence of a given word (again, a cue from the South Florida norms) constrains the subsequent context within an n-gram. This was calculated using conditional entropy, which is similar to the entropy measure described above, but is calculated as conditional on the specified word being the first word in an n-gram. The measure describes the entropy of each n-gram in which one of 9,917 words drawn from the South Florida database was the first word. The difference between this and Experiment 1 is that it concerns not the direct probability of co-occurrence between two specific words, but the orderliness of the word distributions following a specified word in an n-gram. As such, the difference between Experiment 1 and Experiment 4 is somewhat similar to the difference between mutual information and relative entropy statistics discussed above.

Hahn & Sivley then calculated correlations between the conditional entropy of each target word's n-gram distributions and the number of responses yielded to each cue in the WA norms. As in Experiment 1, Hahn & Sivley found a significant but modest correlation between the two statistics ($r = .195$, $p < .0001$). Cues which had a low conditional entropy (i.e. relatively predictable sets of collocates) also received fewer WA response types; higher entropy resulted in larger numbers of response types. This is in line with the usage-based description of language development and word association given above.

Hahn & Sivley do not directly compare the results of their first and second experiments. However, in interpreting the results of Experiment 4, they briefly sketch a model, similar in nature to the composite model, which consists of a predictive component (which exists to aid on-line language processing) derived from word distributions; and a semantic network which links words related by meaning. This component is described as being similar to the semantic networks hypothesised by other researchers (e.g. De Deyne & Storms, 2008; Morais, Olsson, & Schooler, 2013; Steyvers & Tenenbaum, 2005). Hahn & Sivley further suggest (Hahn & Sivley, 2011, p. 754) that the correlation, found in their Experiment 4, between predictive and semantic networks may imply a neural representation shared between the two systems. On this final point, the composite model makes no claims. It does, however, suggest that distributional data influences the semantic system in that it facilitates either entrenchment or abstraction of co-occurring items. It is hard to tell, however, whether this interaction is what is meant in Hahn & Sivley's discussion of "shared neural representation".

Two further studies with potential relevance to the impact of type/token distributions on WA were conducted by Schulte Im Walde & Melinger (Schulte im Walde & Melinger, 2005), and Guida & Lenci (2007). Both studies explored the semantic relationships between cue-response pairs for verb cues, and found negative correlations of around $r=-.3$ between cue frequency and response grammatical class (GC): as cue frequency decreased, noun responses became more frequent. Guida & Lenci (2007) attributed this phenomenon to the increased lexical specificity of infrequent verbs. They suggested that, as verbs decline in frequency, they become more closely associated with specific meanings, contexts, and functions; this semantic transparency leads to them yielding, in WATs, "nouns referring to entities typically participating in the event" denoted by the verb. They use the example of *cycle-bike* to illustrate this (*Ibid.*: p. 305). More frequent verbs do not, according to Guida & Lenci, generally result in these syntactically-related responses.

While these studies do not explicitly discuss type/token distributions, they nevertheless offer a hint that entrenchment of verb-noun pairings may lead to a larger number of noun responses. Assuming that less frequent verbs are also those which have low type variation, then the entropy of the verb's distribution may provide an alternative explanation for the proportion of noun responses given to cues. However, further study is required to demonstrate that type/token distributions are indeed correlated with differences in response grammatical class before this possibility can be discussed.

The experimental findings presented above therefore suggest two potential effects of type/token distributions: low type variation correlates with small associative set sizes (i.e. the number of response

types given to a cue; see Table 3.1) and a high proportion of noun responses, while high type variation appears to correspond to higher set sizes and more frequent verb responses.

8.1.4 Testing the composite model using type/token distributions

The present chapter will test these predictions regarding the effect of the type/token distributions of words co-occurring with verb cues on WA responses. These investigations can be viewed as testing the usage-based aspect of the composite model (i.e. its sensitivity to word distributions) against generative-transformational models of WA, such as those described in Chapter 6 (H. H. Clark, 1970; Polzella & Rohrman, 1970), which do not predict any effect of distributional factors on WA responses. These models posit that the same rule-based system governs response generation for all words, regardless of their frequency and distributions (at least once mature word-feature lists have been developed: McNeill, 1966). Indeed, some generative theories directly reject the influence of these factors on the cognitive representation of language (Chomsky, 1957; see also Pace-Sigge, 2018; Taylor, 2012). As such, if word association is essentially a rule-governed process involving the minimal transformation of cues into satisfactory responses, there should be no significant influence of type/token distributions on responses. This outcome is not predicted, since earlier studies have, as shown above, already given evidence of such effects. Furthermore, neither the basic findings of Chapter 4 (i.e. that noun and verb cues yield different types of response), nor of Chapter 5 (i.e. that transitive and intransitive cues yield responses which differ in terms of directionality) are predicted by generative models.

8.2 Experiment

8.2.1 Hypotheses

Usage-based models of word association, such as the composite model, predict a sensitivity of WA response types and distributions to the type/token distributions of WA cues. Generative models predict no such effects. The current chapter will test these models. If the usage-based models are correct, then:

1. The relative entropy of each verb's collocational distribution (which reflects its type/token distribution) will be correlated with the number of responses of each grammatical class given to that cue in a WAT: low type variation (i.e. low relative entropy) will lead to increased noun responses, while high type variation (high relative entropy) will lead to a higher number of verb responses (cf. Guida & Lenci, 2007).

2. Type variation will positively correlate with associative set size (cf. Hahn & Sivley, 2013): more entropic textual distributions will lead to a greater number of WA response types.

8.2.2 Procedure

In order to test the above predictions, an experiment was devised using the WA data generated in Chapter 5, pertaining to transitive and intransitive verb cues. For each cue in the dataset, all nouns which co-occurred with the cue within a span of 4 words on either side of the node were retrieved from the British National Corpus (BNC). This search was conducted using cue lemmas, rather than word forms. This was because it was felt that lemma forms would provide a more representative sample of the totality of patterns into which a verb enters.

The initial data revealed an imbalance in the number of noun collocates found for the cue verbs: far more noun collocates were returned for transitives than intransitives. This was because transitive verbs have both subject-verb and verb-object relationships with nouns, whereas intransitives have only the former. Since this imbalance can strongly influence the entropy of the verbs' distributions (such that transitives will return higher relative entropy scores, because the increased number of collocates will make their distributions appear less orderly), it was decided to analyse only the nouns occurring in the pre-modifying context for intransitives (e.g. *roar-lion*), and in the post-modifying context for transitives (e.g. *avoid-danger*). While it is acknowledged that this approach is slightly problematic because type variation in both the preceding and following context may influence the psycholinguistic representations of verbs (particularly transitives), the results of Chapter 5 suggest that the strongest influence of such variation is likely to be in the direction selected above. While it may be preferable to include data from both directions, the resulting imbalance in the number of collocates entered into relative entropy calculations for transitives and intransitives is likely to be a more significant confound on the data than the use of the limited but principled approach taken here.

Having identified these verb-noun relationships and their frequencies of occurrence, the relative entropy of each cue's distribution among its noun collocates was calculated, according to the formula given in Section 8.1.2. Associative set size calculations had already been made as part of the study reported in Chapter 5. These two measures were then statistically compared in order to test Prediction 2, above. In addition, the distributional entropy measures were compared with the number of verb and noun responses given to each cue, in a test of Prediction 1.

8.2.3 Results

Looking firstly at the test for correlation between distributional entropy and the proportion of noun and verb responses, an initial test for normality of distribution showed that the distributional entropy values were negatively skewed, with most values peaking close to the maximum value of 1. A Kolmogorov-Smirnov test for normality of distribution confirmed that the distribution was not normal ($p < .0001$). As such a Spearman test was used to determine the rank-order correlation of the variables. This test found that distributional entropy was negatively correlated with the number of noun responses generated to a cue ($r = -.460, p < .001$) and positively correlated with verb responses ($r = .349, p = .001$). As suggested by Prediction 1, high type variation therefore appears to have resulted in an increase in verb responses, while low type variation correlated with increased noun responses.

These correlations are slightly higher than those reported by Guida & Lenci (Guida & Lenci, 2007; Schulte im Walde & Melinger, 2005), who found significant correlations of $r = .24$ between cue frequency and verb responses and $r = -.28$ between cue frequency and noun responses. This offers a hint that type/token distributions may offer a better fit for explaining response grammatical class differences than cue frequency does. In order to test this further, correlation analyses were run between cue frequency and noun/verb responses. Both of these tests returned non-significant values (frequency vs. noun responses Spearman's $r = .022, p > .1$; frequency vs verb responses $r = -.034, p > .1$). This appears to confirm that, for the cues used in the present study, the relative entropy of a verb's distribution is a better predictor of noun and verb responses than is cue frequency. This may suggest that lexical specificity is a consequence of low type variation, as hinted at by Hanks (2013).

Prediction 2 states that low type/token variation should lead to narrow response distributions. As discussed above, this effect may manifest in correlations between distributional entropy and associative set size (i.e. the number of response types given to each cue). In order to test the first of these possibilities, a Spearman correlation was calculated between distributional entropy and associative set size. A modest but significant correlation was found between these variables ($r = .215, p = .035$). It is worth noting that the size of this correlation was very similar to that reported by Hahn & Sivley (2011: Experiment 4) in their study of the relationship between conditional entropy and associative set size ($r = .195, p < .0001$). The smaller p value in that study can be attributed to the far larger sample size used (more than 9000 cues, as opposed to the 98 used in the present experiment), while the slightly stronger correlation in the present study may be due to the exclusive use of verb-noun pairs here, as opposed to more syntactically diverse n-grams studied by Hahn & Sivley. This result

suggests that in addition to influencing the proportion of noun and verb responses, a word's distributional entropy impacts on the number of response types which are typically given to a cue.

8.3 Discussion

The results presented above confirm both of the predictions made in Section 8.1.4. This suggests an influence of the textual type/token distributions of verb cues and their noun collocates on WA responses. In the terms introduced in Section 8.1.1, the results suggest that verbs which have relatively entrenched relationships with their most frequent noun collocates are more likely than verbs with less orderly distributions to yield noun responses and comparatively small associative set sizes. Conversely, verbs whose varied and disorderly noun collocate distributions lead to greater abstraction away from small numbers of entrenched exemplars tend to yield verb responses and larger set sizes than verbs with more orderly distributions.

These results support a usage-based view of psycholinguistic representation in which key features of a word's distribution exert an influence on the manner in which words are stored and processed, including the semantic representations which are held for words, and are fully compatible with the predictions of the composite model. The wider psycholinguistic implications of these results will be discussed in Chapter 9.

One aspect of the current results which deserves a fuller treatment, however, is weakness of the correlations reported above, particularly between type variation and associative set size. Similarly weak relationships between distributional data and WA have been reported elsewhere. For example, Hahn & Sivley (2011) reported a correlation of $r = .195$ between conditional entropy and set size, while a slightly stronger correlation ($r = .327$) was found between cue-response MI and association strength in a study by Kang (2018). This will be further investigated below.

8.3.1 Comparing corpora and WA responses

The collection of frequently occurring verb-noun pairs for the present study provides an opportunity for the side-by-side comparison of WA responses with this corpus data. The main aim of this qualitative analysis is to identify dissimilarities which can potentially account for the generally modest correlations found between these data sources. A guiding question is, "In what way(s) did the WA responses in this dataset differ from those words which were frequent collocates of WA cues, but did not appear in the WA data?".

A relatively shallow (though informative) set of answers to this question can be generated by considering the contexts of both the corpus and the WA test described in Chapter 5. For example, the two data collections took place at an interval of around 25 years. This is reflected in some of the data. The cue *meditate*, for instance, received numerous WA responses which would have been less likely in the 1990s or earlier, such as *yoga* and *Buddhism*. The BNC collocates appeared to reflect an earlier conception of meditation – *study* and *prayer* were amongst its top collocates. In other cases, the technical nature of some of the BNCs texts is likely to account for some differences in word pairs. Collocations such as *oscillate-neutrino/radiation* did not occur in the list of WA responses (a point which will be returned to below).

Table 8.1

Associative and distributional properties of 5 cues with the largest difference between associative and distributional entropy

	H_{rel} Dist.	H_{rel} Assoc.	Difference	Primary Response	Assoc. freq.	Set size
<i>Mean</i> (all cues)	.94	.77	.17	N/A	18.61	19.76
<i>purr</i>	.9	.23	.67	cat	49	5
<i>erupt</i>	.91	.29	.63	volcano	47	7
<i>roar</i>	.96	.33	.62	lion	46	6
<i>extinguish</i>	.89	.31	.59	fire	47	5
<i>shiver</i>	.96	.46	.5	cold	42	6

H_{rel} Dist. = Distributional Relative Entropy; H_{rel} Assoc. = Associative Relative Entropy; Assoc. Freq. = Associative frequency of the primary response.

It is possible to identify deeper differences between WA and corpus data, however, which may point to potential psycholinguistic preferences which respondents bring to the WA task. One way to identify these is to compare relative entropy values for each cues associative distribution (henceforth to be referred to as associative entropy) with the distributional entropy values discussed above²³. The five cues which showed the largest difference between these values are set out in Table 8.1, along with their statistical profiles. It can be seen that all of these cues had small associative set sizes, with distributions dominated by a single highly frequent primary response. That the textual distributions for these words were not similarly dominated by a single frequently co-occurring collocate is evident from their comparatively high distributional entropy. Comparing these collocational pairings with the cues' associative responses can therefore provide an indication of the type of words which are

²³ The calculation of associative entropy is done by applying the same formula as for distributional entropy, described in Section 8.1.2, above, to the set of cue-response pairs collected for each cue.

preferred by WA respondents. The associations provided to these five cues are therefore set out together with the corresponding number of collocational pairs in Table 8.2.

Table 8.2
Association responses and most frequent collocations for the 5 cues presented above

	Association responses	Assoc. Frequency	Collocates	Co-occurrences with cue
<i>purr</i>	cat	49	cat	16
	growl	1	car	13
	noise	1	engine	9
	meow	1	voice	6
	stroke	1	hear	5
<i>erupt</i>	volcano	47	violence	31
	mood	1	row	30
	anger	1	lava	26
	bomb	1	volcano	22
	splurge	1	riot	19
	earth	1	fight	17
	explode	1	war	16
<i>roar</i>	lion	46	engine	28
	shout	2	car	26
	tiger	2	head	18
	growl	1	wind	16
	bear	1	fire	15
	loud	1	crowd	14
<i>extinguish</i>	fire	47	fire	28
	put out	3	system	14
	foam	1	flame	10
	eliminate	1	light	7
	flames	1	right	7
<i>shiver</i>	cold	42	cold	32
	shake	4	eye	12
	timbers	3	night	12
	spine	2	body	11
	warm up	1	voice	10
	me timbers	1	Ruth	9

As per the discussion in Section 8.2.1, only noun collocates occurring within a span of 4 words were used in the study. The collocates for intransitive cues (i.e. *purr*, *erupt*, *roar*, and *shiver*) were drawn only from the preceding textual context (i.e. -4 words), while those for transitives (above, only *extinguish*) were taken from the subsequent context

The first thing to note about this data is that the frequency distributions amongst the collocates are far more even (i.e. entropic) than the associative frequencies, as would be expected given their higher distributional than associative entropy. A second point of note is that, in most cases, the collocations identify a broader range of semantic referents for the verb cues than are found in the associates. The cue *purr*, for example, appears to have strongly activated the sense of *cat* (as evidenced by the

responses *cat*, *meow*, and *stroke*), rather than the more metaphorical senses which emerge in corpus analysis: *car*, *engine*, *voice*. The collocates of *roar* reveal a similar pattern: this concept was largely associated with a single concept (a fierce animal, such as a lion, or perhaps a tiger or bear), but the corpus again revealed multiple referents, including *car*, *engine*, *wind*, a *crowd*, and *fire*. A similar situation is found for *erupt*: respondents associated this word only with volcanoes, but the corpus suggests that words relating to conflict (*violence*, *row*, *riot*, *fight*, *war*) co-occur with *erupt* at least as frequently. These examples suggest a selectivity on the part of WAT respondents with regard to their knowledge of co-occurrence, as well as a uniformity of conceptual content in WA responses.

One possible explanation for this selectivity is that, for words with low type variation, respondents may simply choose the most frequent collocation of the cue. This would, in fact, be predicted by the composite model if it were assumed that these responses originated solely from the co-occurrence-based component. However, such an explanation could not account for all responses (at least if the BNC gives an accurate picture of co-occurrence frequencies). While in the cases of *purr*, *extinguish*, and *shiver*, for example, the top collocate and most frequent WA response are the same word (i.e. *cat*, *fire*, and *cold*, respectively), a look at the entire dataset reveals only eight occasions (from all 98 cues) on which this occurs (in addition to the three above, these were *bark-dog*, *glow-light*, *recede-hair*, *seethe-anger*, and *waddle-duck*). Moreover, even in the above examples of cues with large differences between associative entropy and type variation, there are two occasions on which the top collocate and the dominant WA response do not coincide (*erupt* and *roar*).

There is, on the other hand, some evidence that response selectivity may have been more strongly influenced by semantic than collocational factors. For example, the primary responses to all five of the cues presented in Table 8.2 all capture core semantic content for their cue in a way which is not true of the word's other collocates. For example, the concept of a *volcano* is far more integral to the semantics of *erupt* than is *violence* (despite the latter being a more frequent collocate of *erupt*), in the same way that *cat* is integral to *purr*, but *car* is not. Furthermore, it is notable that even when a given respondent did not select the primary response to a word, they nevertheless generally chose a word conceptually related to it, as in *stroke* and *meow* (both conceptually related to *cat*) for *purr*; *tiger* or *bear* (rather than *engine* or *car*) to *roar*; and *flames* to *extinguish*. That these responses appeared to be selected from amongst the same lexical set as the primary response suggests that most respondents accessed the same conceptual information for that cue: they are apparently not selecting from amongst collocational candidates.

This conceptual selectivity, then, may be one reason why WA and corpus-derived data diverge. There is a danger, however, in drawing such a conclusion from the above data, which represents the extreme low end of the associative entropy spectrum. Another way to investigate the balance of semantic and co-occurrence-based information on WA responses is to look at responses to cues with *low* distributional entropy. Following the predictions of the composite model, it should be expected, for these cues, that WA responses should generally coincide with the most frequent collocate. This is because low distributional entropy should indicate a strongly entrenched word pairing dominating the cue's distribution, leading to strong association between those words in both the semantic and co-occurrence-based systems.

In order to test this, the ten cues with the lowest distributional entropy values in the dataset were examined. In general, the results did not support the above predictions. Of the most common responses to these ten cues, several were in fact synonyms or troponyms (*happen-occur*, *tremble-shake*, *sparkle-shine*, *hibernate-sleep*) rather than syntagmatic responses. Amongst cues responded to syntagmatically (and therefore potentially based on verb-noun distributions), it was notable that only *bark* yielded a frequent collocate of the cue (*dog*). This word is similar to *roar-lion* and *erupt-volcano* in that it captures core semantic content of the cue while also being a frequent collocate. Other syntagmatic responses included *irrigate-water*, *lick-tongue*, and *ache-pain*. None of these responses constitutes a frequent collocational pairing – they are better described as (semantic) entailments of the cue verb. In fact, in most cases, each cue's most frequent collocate was only very rarely given as a response. Only in the cases of *bark*, *ache*, and *hibernate* was the most frequent collocate produced by more than one respondent (*bark-dog* 38 times, *ache-head* 15 times, *hibernate-hedgehog* 7 times).

In fact, of these ten cues, it was only in the case of *bark-dog* that the primary response coincided with the top collocate. This compares with four such coinciding responses/collocates among the ten cues with the lowest *associative* entropy. This suggests strongly that a relatively orderly corpus distribution, dominated by a single frequent collocate, does not guarantee that this collocational pairing will also dominate the cue's associational profile. A similar point was made by Hahn & Sivley (2011, p. 756), who noted that neither low distributional entropy nor a strong associative relationship with a specific word was a guarantor of a predictive (distributional) relationship between two words. They gave the example of *Marxism*, which has a low distributional entropy and only one semantic associate – *fascism* – but the predictive relationship between these words is close to zero.

The cue *happen* is a particularly telling case. While this cue has many noun collocates (in the BNC, almost 2000 nouns occurred at least twice in the four-word span preceding *happen*), its textual distribution is dominated by *thing* (*nothing* and *everything* also both occurred within its five most common collocates), leading to its relatively low distributional entropy ($H_{rel} = .809$ – the third lowest in the dataset). However, *thing* appeared only once as WA response to *happen*. Instead, more than half of respondents chose the synonym *occur* as their response. This provides a concrete example of the fact that while type/token distributions do appear to influence WA, they do not necessarily determine specific responses. Instead, the near total lack of semantic content in both *happen* and its dispreferred response, *thing*, suggests that semantic relatedness between words is a stronger influence on response selection.

This strongly semantic aspect to response selection is further suggested by analysis of responses to those cues with the most entropic associative distributions. As suggested by the correlation between distributional entropy and response grammatical class, these cues largely yielded verb responses. As in Guida & Lenci (2007), the majority of such responses were synonyms and troponyms. For example, the most common responses to *avoid* (associative $H_{rel} = .954$) include *move*, *ignore*, *dodge*, *run*, and *evade*; only one potentially co-occurrence-based association (*avoid-people*) appeared among the top 6 responses to this cue. Similarly, *dwindle* (associative $H_{rel} = .944$) yielded *lessen*, *down*, *reduce*, and *decrease* among its top 5 responses.

The examples above appear to suggest a view of WA in which response selection is largely dependent upon shared semantic content between cue and response. This conclusion is in line with those models of WA which view syntagmatic responses to be a form of semantic relation (Guida & Lenci, 2007; McRae et al., 2012; Mollin, 2009; Schulte Im Walde et al., 2008; Vivas et al., 2018). One way to explain this is to suggest that where semantic relevance can be best supplied through a collocate, as in the cases of cues such as *bark*, *roar*, *erupt*, *extinguish*, and *shiver*, then this type of response will be selected. The collocating word providing this core semantic content may or may not coincide with one of the most frequent collocates of the cue. However, in the case that it does, the above analysis suggests that this should not be taken as evidence of a causative relationship between the two (i.e. that the response emerged from co-occurrence-based processing).

In cases where no single collocate captures this core semantic information, a synonym or lexically related word will be preferred to any frequent but semantically uninformative collocation, even where that collocation has a relatively dominant relationship with the cue (as in the case of *happen-thing*). This hypothesis is in line with the results of Vivas et. al. (Vivas et al., 2018), who found that the majority

of responses to the 199 concrete nouns in their WA study were words previously identified as amongst the defining features of the cue. It also implies that several aspects of a cue's distributional-semantic representation influence WA, including not only the availability of frequent (semantically related) collocates, but also of synonyms, troponyms, etc. For example, in the case of the cue *erupt*, available paradigmatic relations such as *explode* or *flare up* may not share sufficient semantic content to achieve activation. Future research could therefore profitably look at the determinants of syntagmatic and paradigmatic responses by looking closely at the distributional properties and semantic neighbourhoods of cues. The implications of this view of WA will be returned to in Chapter 9.

One question which arises from this discussion concerns the nature of the influence of type/token distributions: if shared semantic content is the key component of WA responses, how should distributional influences on the WA process, demonstrated in section 8.2, above, be interpreted? One possible answer to this question is that type/token distributions may be confounded with the effects of cue lexical specificity (cf. Guida & Lenci, 2007). That is to say, it is probable that words with low type variation also have high lexical specificity (i.e. these two variables are likely to be negatively correlated); and if so, it may be that the latter variable is a better explanation of WA response patterns than the former. Perhaps, given a suitable psycholinguistic measure, lexical specificity would explain greater variation in both associative set size and response grammatical class than does distributional relative entropy.

One obstacle to such research is the difficulty of disentangling these two variables. Indeed, usage-based theories of language suggest that distributional and semantic aspects of words are mutually reinforcing (see Section 8.1.1). As such, these two variables are likely to be highly correlated – i.e. words which are highly specific will generally have low type variation, and vice versa. Nevertheless, there are some cues in the present dataset which suggest the separability of these variables. For example, *happen* has low lexical specificity, but quite low distributional entropy (due to its co-occurrence with *thing*). As has already been seen, this cue yielded mostly synonym responses (i.e. *occur*). The production of synonyms is more common for lexically non-specific, high type-variation cues. In this example, then, it appears to be the low lexical specificity of *happen* which results in the production of synonym responses, since the low distributional entropy did not result in frequent production of *thing*.

On the other hand, there were also several examples of cues which followed the converse pattern – i.e. high lexical specificity but also high distributional entropy. Such cues include *fidget*, *grovel*, *trounce*, *grumble*, and *snooze*. The textual distributions of these cues were characterised by a small number of

collocates of roughly equal frequency. This accounted for their very high distributional entropy (between .995 and 1). Their associative entropy varied widely, from .569 for *snooze* to .889 for *trounce*. The most frequent responses to these words were synonyms or troponyms. The primary responses, for example, were *fidget-move*, *grovel-beg*, *trounce-beat*, *grumble-moan*, and *snooze-sleep*. For these cues, high lexical specificity did not lead to noun responses, as would be predicted if lexical specificity were a better explanation of response grammatical class than type/token distributions. As such, these examples suggest that high distributional entropy provides a better explanation for the production of verb responses than does lexical specificity. This conclusion is, however, the opposite of that suggested in the example of *happen*, where it appeared that low lexical specificity was more likely to account for the high level of verb responses, rather than low distributional entropy. These examples therefore point to considerable further complexity in the explanation of WA responses, but do not in general argue for the possibility that the type/token influences on WA discussed in this chapter are in fact the result of differences in cue lexical specificity.

Another way to understand cue type/token effects (if WA responses are assumed to be largely semantic in nature) is to view them as an influence on the development of the semantic system underlying WA response generation, rather than on the process of generating responses itself. This process was outlined in Section 8.1.1., wherein it was hypothesised that low type variation would lead to the entrenchment of given word pairings, both in terms of their predictive textual relationship and their semantic association. What the above data may suggest is that only these semantic reflections of type/token variation are visible in WA; and furthermore that they are obscured by the nature of a response generation system which prioritizes semantic relatedness over exclusivity of predictive relations.

This possibility has the implication, within the composite model, that while co-occurrence-based, semantic, and form-based knowledge exist in the mind, these systems are not drawn on to the same extent in WA. Instead, the semantic system is chiefly responsible for response generation, with form- and co-occurrence-based knowledge playing a lesser role. This suggests that the assumption of equality between these three components of the model posited in Chapter 7 may need to be revised. These issues will be returned to in Chapter 9.

Returning finally to the earlier question of the ways in which corpus- and WA data differ, the above discussion suggests that WA responses are substantially closer to the core semantic features of a given word than are collocational distributions. One further difference between these two types of data can be observed: corpus searches appear to return collocates which are much more specific than those

generally given as WA responses. One example of this was given above: the technical terms contained in some corpus texts meant that pairs such as *oscillate-neutrino/circuit* were returned by the corpus search; these pairs were never given as WA responses. A similar example is *migrate*, which returned *cell* and *larva*. It is therefore helpful to ask whether these examples are indicative of a wider trend towards higher lexical specificity in corpus than WA response data. Several further (non-technical) examples suggest that this might be the case. For example, *swoop* frequently yields *bird* as both a collocate (9 times) and a WA response (20 times); but the collocation list included specific bird species much more frequently than did the WA response list: *gull* (7 times, compared with zero WA responses), *eagle* (6 co-occurrences, zero WA responses, and *hawk* (6 co-occurrences, one WA response; *swoop*, incidentally, provided further evidence of the centrality to WA of core semantics: the most frequent collocates were *police* and *officer*, neither of which occurred in the WA dataset). Other examples include *detect*, which returned numerous chemicals and ailments from a corpus search, but only the much more general *metal* and *virus* in WA; and several cues (e.g. *bury*) which yielded non-specific human entities (e.g. *body*) where the corpus search returned the names of specific people. This suggests that a certain level of lexical generality might be another feature of WA which sets it apart from corpus data. This possibility is supported (and could be further researched) by reference to network models of WA, which suggest that words which are high in frequency, early acquired, and perhaps also relatively concrete, are more central to the lexicon than words which do not fit this profile (De Deyne & Storms, 2008; Steyvers & Tenenbaum, 2005, and see Chapter 3).

8.4 Conclusion

The present chapter has uncovered significant effects of cue type/token distributions on word association response patterns. Cues whose noun collocate distributions show low type variation (i.e. are dominated by one or more highly frequent collocates, measured through their relative entropy) are more likely than cues with higher type variation to yield verb responses and small associative set sizes. Cues with high type variation (i.e. many different noun collocates which co-occur with the cue with similar frequency) are more likely to yield noun responses and have larger set sizes.

These effects support the usage-based model discussed in Section 8.1: if a verb's noun collocates show low type variation, these verb-noun pairings will become entrenched in the lexicon, leading to strong association between the cue and its most frequent collocate(s) on both predictive and semantic levels. On the other hand, if the verb frequently co-occurs with numerous other nouns, this will lead to abstraction away from a small number of similar exemplars, resulting in a bleaching of the verb's semantics and a weakening of its associations with its collocates.

This process suggests the interactivity of co-occurrence-based and semantic knowledge on the construction of lexical knowledge, since a verb's distribution contributes to the specificity of its mental (semantic) representation. For this reason, efforts to disentangle the effects of lexical specificity from those of type/token distributions are likely to be challenging, although worthy of further research. This interactivity also raises the possibility that co-occurrence-based knowledge is not directly involved in the process of WA response generation – something which was further suggested by the post-hoc comparison of corpus data with WA responses conducted above. This analysis suggested that WA respondents may not be drawing on their knowledge of collocations in producing responses, even in the cases where those responses are classically syntagmatic. This is evidenced by the fact that many response sets are more similar to conceptual category groups than lists of collocational candidates (cf. Vivas et al., 2018).

The current chapter has focused on two aspects of WA response generation – grammatical class and associative distribution. Future research into the effects of type/token distributions on WA could focus on other aspects of response generation. For example, it is possible that cue type/token distributions may affect the speed of response generation. It could be predicted that, in cases of low type variation, responses should be relatively fast, because (a) assuming that the co-occurrence-based system is activated during WA, there will be relatively little competition for activation within this system, and (b) there should (regardless of activations in other systems) also be relatively little competition within the semantic system, due to the strong semantic association between a given word and its most frequent collocate(s). On the other hand, responses to words with high type variation will face greater competition for activation in both systems, since this system will potentially activate many predictive and semantic associates of the cue.

The current chapter has raised numerous issues regarding the relationship between a word's textual distribution, its semantic representation in the mind, and its associative profile. The following chapter will discuss these issues in the context of the findings of the accumulated results of Chapters 2-8, and integrate them into a wider view of the lexicon.

Chapter 9: General Discussion

9.1 Revisiting some initial questions

This thesis began by posing a simple question: why, given wide agreement on the psycholinguistic promise of the word association (WA) task, has research using this method generally failed to produce clear and informative findings on the structure and processing of lexical knowledge? While several answers can be given to this question, one in particular has been pursued in the preceding chapters: that existing studies have paid too little attention to unexplained variation in WA response patterns, resulting in untested assumptions about the nature of word association, its conceptual unity, and the factors which influence it (Fitzpatrick, 2007, 2009; Fitzpatrick et al., 2015; Hutchison, 2003; Hutchison et al., 2008; McRae et al., 2012).

The work in this thesis has demonstrated that the critical investigation of these areas of variation (in this case, the investigation of lexical variables including grammatical class and transitivity) can lead not only to a clarification of some of the methodological options available to word association researchers, but also to new insights into the structure and development of the mental lexicon. Two examples of the former are the insights gained into the different influences of semantic and distributional variables on associative network properties (see Chapter 3 and below), and the uncertainty surrounding the nature of WA response categorization highlighted in Chapter 7.

The present chapter will focus not on these methodological factors, however, but on the light shed by the experimental work in this thesis on the nature of the mental lexicon. Building in particular on the work presented in Chapters 7 and 8, it will be argued that several aspects of the composite model presented in Chapter 7 require revision. A starting point for this discussion is the debate, raised by the apparently position-based differences between responses to transitive and intransitive verbs in Chapter 5, regarding the role of distributional, contiguity-based word knowledge in word association. Classical discussions of associative thought (Warren, 1916) tended to assert that associations are learned through textual contiguities, and responses are mere recollections of these contiguities. More recent debates have, however, suggested that such assumptions are unreliable.

9.2 Deeper investigations into word variable influence

McRae et. al. (2012), highlighting of the problem of untested assumptions in WA, have argued that it cannot be taken for granted that the factors leading to the formation of association are also those which govern their retention and/or production in WATs:

there are a number of discontinuities between the definition of association and its operationalization. Association proper is learning-based; word association is retrieval or production-based. Association proper is based heavily on sensory information; word association is linguistically based. Association is based on contiguity, accidental or otherwise; word associations are [...] almost always meaningful.

Consequently, any role of distributional knowledge in WA response generation must be demonstrated empirically.

Chapter 6 explored how distributional influences on WA might be explained by two contrasting theoretical models. The first of these was a generative-transformational model developed by Clark (1970) and Polzella & Rohrman (1970). This highly detailed account is capable of explaining many paradigmatic responses using minimal contrast rules, and can account for associations between words of the same lexical set. However, as argued in Chapter 6 and beyond, the model fails to account for a wide range of other types of response, including the majority of paradigmatic responses. Given this major limitation, it will not be considered further here.

A second set of accounts of WA responses is based on usage-based models of language (Bøyum, 2016). These models suggest that language acquisition, knowledge, and use can be explained by the combination of an individual's experience of language in context and the domain-general cognitive processes applied to linguistic experience. Because these models incorporate linguistic knowledge on numerous levels, including semantic, distributional, and form-based knowledge, they appear well placed to explain a wide range of WA response effects, including those established in Chapters 2-5:

1. Individual WA respondents show preferences which are consistent over extended periods of time and with cues of different grammatical classes (Fitzpatrick, 2007, 2009, Higginbotham, 2010, 2014, and see Chapter 2)
2. Responses are influenced by a range of factors:
 - a. distributional variables (particularly AoA) appear to influence the extent to which a given cue acquires a central position in the lexicon (De Deyne & Storms, 2008, and see Chapter 3)

- b. semantic variables are more likely to affect the number and type of outgoing links from a cue (see Chapter 3).
- c. Grammatical factors influence the type of responses given to cues; in particular, the transitivity of verb cues influences the directionality of position-based responses. These effects may indicate an influence of distributional knowledge, such as a word's patterns of collocation and complementation, on WA (see Chapters 4 and 5).

All usage-based models are capable of explaining findings 1 and 2a. Their recognition of the impact of unique linguistic experiences and cognitive preferences on individuals' linguistic structure and processing provides a ready-made explanation for both the individual WA response preferences described in Chapter 2 (first discussed by Fitzpatrick, 2007), and the influence of distributional variables on associative network centrality. Indeed, this latter finding, which receives support from numerous empirical and theoretical sources including novel WA data (De Deyne & Storms, 2008; Morais et al., 2013; Steyvers & Tenenbaum, 2005), is inherently usage-based in that it suggests that lexical structure is derived from the linguistic experience of each individual.

UB models are also capable of explaining the influence of semantic and grammatical factors on WA responses. Researchers have, however, tended to fall into one of two camps concerning how this influence happens. The first view is a distributional position: the formation of associations is driven by contiguities between words or phenomena, and WA responses reflect this type of knowledge. The second is that WA responses are underpinned by semantic memory; even if associations are *learned* through contiguity, it is semantic processing which governs their *retention* and *production*.

These accounts differ in the manner and extent to which they can explain findings 2b and 2c. According to the semantic model, words are organised in the mind in a semantically structured network, such that lexical items with related meanings are interconnected and therefore become activated during WA response generation. Semantic influences on outgoing links are therefore easily explained by this model, since semantics is seen as guiding the processes of lexical access, lexical selection, and retention of associations. In the case of grammatical influences on word associations, associations are understood as responses to cue semantics; responses to noun cues (for example) will be similar to the extent that nouns share semantic features, such as their denotation of entities rather than events (Croft, 1990, 1991, 2001). In cases where grammatical factors appear to influence the number of position-based responses, such as with verb transitivity, it remains possible to offer a semantically-based explanation: semantic WA categorization schemes code apparently position-based responses

in terms of thematic roles or conceptual features of cues (rather than as cue-response or response-cue associations; Guida & Lenci, 2007; Vivas et al., 2018).

Because distributional accounts of WA assert that responses reflect the learning of contiguities between words in text, their ability to explain the above effects is essentially determined by the extent to which words can be shown to co-occur in text. The explanation of verb transitivity effects is therefore relatively straightforward: since intransitives do not take objects, their most frequent co-occurrences (and therefore their strongest associative relationships) will be with their grammatical subject. They will, as such, yield response-cue associations. For transitives, the more predictable relationship is with the grammatical object, thus resulting in cue-response links. It is less clear, however, how contiguity-based accounts of WA could explain cue semantic effects, such as the finding that concrete words yield more homogeneous sets of responses than do less concrete ones. Explaining such findings would require evidence either that textual distributions of concrete words are more predictable than those of abstract words, or that they occur in more contextually similar settings (thus enabling the process of association through substitutability: Ervin, 1961). Some support for this possibility is provided by Gruenenfelder et. al. (2016), who found that by combining associative and contextual similarity-based models of word knowledge, key characteristics of human associative networks could be replicated using corpus-based modelling. These models do not, however, provide evidence of specific concreteness or affective strength-related effects.

To summarise: while usage-based models appear able to explain the most prevalent WA response patterns, it is unclear precisely how these models function with regard to the specific aspects of lexical knowledge utilised during WA response generation.

In Chapters 7 and 8, a potential solution to this problem was put forward: could a composite UB model, in which WA response processing is served by simultaneous, interactive, and approximately equal activation of semantic, contiguity-based, and form-based components of lexical knowledge, provide a more coherent account than either of these components alone?

This composite usage-based model posited that word associations are learned from perception and understanding of semantic, distributional, and form-based aspects of linguistic input, and are stored in corresponding memory subsystems. Their organisation within these components is strongly influenced by general cognitive operations such as perception, comparison, and categorisation. There is much interaction between the three memory subsystems, allowing, for example, the emergence of phonaesthetic knowledge through the gradual association of form-based and semantic aspects of

words such as *glisten*, *gleam*, and *glimmer*. Associations within each of the three memory components are simultaneously activated by the presentation of WA cues. The first word to reach an activation threshold is selected. This activation can be facilitated by the simultaneous activation of a single word by more than one system (as would hypothetically happen for *cat* and *dog*, which are semantically and distributionally related, as well as both being CVC words, thus sharing associations of all three types).

The composite model explains most important WA effects by drawing on both the semantic and distributional subsystems: where semantic effects occur, they do so because of responses which have emerged from the semantic system. Likewise, distributional effects such as those pertaining to grammatical class and transitivity are explained by greater activation of the distributional system during response generation for (e.g.) verb than noun cues.

The composite model was intended to represent a kind of null hypothesis regarding the influence of different components of linguistic knowledge on WA: until it can be demonstrated that equality of influence from each component does not exist, it is experimentally useful to assume that it does. The research in Chapters 7 and 8 (as well as several findings from wider WA research), however, provide several hints that an inequality of influence can now be demonstrated. It will be argued below that the system chiefly responsible for the generation of WA responses is in fact semantic. A revised composite model of word association will then be presented, before a number of additional theoretical and empirical challenges are discussed.

9.3 Re-evaluating the composite model

Word association responses demonstrate a tendency towards meaningfulness. This is a fact evidenced by the frequent finding that paradigmatic/meaning-based WA responses are more common than syntagmatic/position-based ones (De Deyne & Storms, 2008; Deese, 1962a; Entwisle, 1966a; Fitzpatrick, 2007, 2009; Fitzpatrick et al., 2015; Higginbotham, 2010, 2014; K. Nelson, 1977; Wolter, 2001; and see Chapters 2, 4, and 5 of the present thesis); and that both of these are far more common than form-based responses. More importantly, however, the meaning-based/position-based system of response categorisation (and perhaps also the paradigmatic/syntagmatic one) blurs this semantic imperative. This is because it can easily lead to the misconception that position-based responses are less meaningful than “meaning-based” ones – something which, upon examination, is not obviously the case. In fact, it appears that position-based responses are skewed away from frequent collocations that do not signal a semantic relationship, such as determiners and pronouns (Mollin, 2009), in favour of those that are. Examples of the meaningful nature of position-based responses abound: see, for 202

instance, the data described in Chapter 5, which yielded several position-based responses to the verb *thrive*, such as *plants*, *economy*, *bees* and *environment*, which constitute statements regarding the type of things which thrive; and the same semantic relationship pertaining between the cue *migrate* and its responses *migrant*, *birds* and *animals*.

Numerous researchers have chosen not to adopt the meaning- vs. position-based distinction, favouring instead schemes which reflect the view that all WA responses contain semantic content. For example, McRae et. al. (2012, p. 44) give one detailed (though largely noun-cue focused) description of the types of semantic relationships which can exist between cue-response pairs. This includes several categories of response which others might categorise as position-based, such as object properties (*apple-red*, *fridge-cold*), systemic features (*dolphin-intelligent*) and evaluations (*gown-fancy*). Applying a scheme created by combining McRae et. al.'s description with additional semantic and form-based categories suggested by Santos et. al. (2011), Vivas et. al. (2018) found that “the vast majority of responses in the WA task were governed by some form of semantic relation with their corresponding cue word”. A similar conclusion was reached by Guida & Lenci (2007), who used an alternative scheme, based on the semantic roles which hold between words, to explain associations to verb cues. This scheme also included many semantic associations which could be defined as position-based: agent (*kill-assassin*), patient (*eat-food*), instrument (*pen-write*), cause (*tremble-cold*) and result (*fall-pain*), for instance. These schemes suggest that the meaning-based/position-based distinction, though informative, may itself feed assumptions as to the origin or nature of responses.

The coding schemes described above demonstrate the plausibility of a predominantly semantic basis to word association. They do not, however, preclude the possibility that activation of a contiguity-based cognitive component is a factor in the production of WA responses. There are, however, numerous further justifications for the assertion that distributional knowledge plays only a minor role in the generation of WA responses. A starting point for this discussion are the findings of Mollin (2009, p. 185, and see above), which reveal, in line with several other comparisons of corpus and WA data (including that presented in Chapter 8 of the present thesis), “that many word pairs that collocate in the BNC [i.e. the British National Corpus] are not produced as stimulus-response pairs in the word association task (and vice versa)”. Mollin argued that the low level of correspondence between these two data types is because the types of knowledge underlying corpus and WA response data are fundamentally different. Language production of the type recorded in a corpus, while essentially meaning-focused, is the result of contextualised communication which begins with a “pre-verbal proposition” in the mind of the speaker, something that is not present in word association (Mollin,

2009, p. 197). Furthermore, the production of fluent language in communicative contexts is facilitated by an intimate knowledge of meaningful patterns of lexical co-occurrence (Hoey, 2005; Pace-Sigge, 2013; Taylor, 2012). This knowledge is probabilistic and derived from prior experience of linguistic communication: Mollin argues that it "is not consciously and artificially reproducible in elicitation" tasks such as word association (p197). As a result, word association response sets rarely contain the most common of words from corpora.

Mollin's view is supported by research into the extent to which psycholinguistic priming effects occur between collocating word pairs in the lexical decision task (LDT). An absence of "pure associative priming" effects [i.e. priming between words associated through contiguity but not by semantics] in tasks such as lexical decision would suggest that contiguity-based knowledge does not become activated in decontextualized tasks, such as word association. In a test studying priming effects between collocating pairs which additionally were or were not also word associates, Durrant & Docherty (2010) found that, once strategic responding had been ruled out using a masked priming technique, only associated pairs yielded a significant priming effect; non-associated collocates did not significantly prime.

This finding supports the interpretation of McRae et. al. (2012), who note that while "no study has conclusively shown that pure associative priming exists", numerous studies have demonstrated pure semantic priming effects (i.e. semantic priming in the absence of collocational relations; McRae et al., 2012, p. 57). A similar conclusion is reached by Hutchison (Hutchison, 2003), who in a review found that the few studies to have looked into pure associative priming did not support the existence of a pure associative effect. This was in part because these studies (e.g. Ferrand & New, 2003; Hodgson, 1991; Perea et al., 1997; Williams, 1996) did not adequately control for semantic relationships between words (e.g. Ferrand & New, 2003, arguing (p26) that "associative relations [...] reflect word use rather than word meaning", include pairs such as *astronaut-space*, *aquarium-fish*, and *needle-thread* in their list of non-semantic associates), or that they use unmasked methods which allow respondents to develop strategic approaches to their responses.

That WA respondents apparently do not apply distributional word knowledge to the WA task provides an explanation for the generally underwhelming performance of studies aiming to replicate human WA response patterns using contiguity-based algorithms (see Sections 6.3.1), as well as the low correlations between WA and corpus data reported elsewhere (e.g. Kang, 2018). Rather than viewing such correlations as evidence of modest contributions of distributional knowledge to WA response generation, the semantic view of word association being argued for here suggests that these results

are simply a consequence of the occasional overlap between co-occurring words and their most semantically salient features or associations, as in the case of cue-response pairs such as *shiver-cold* and *extinguish-fire* (see Chapter 8).

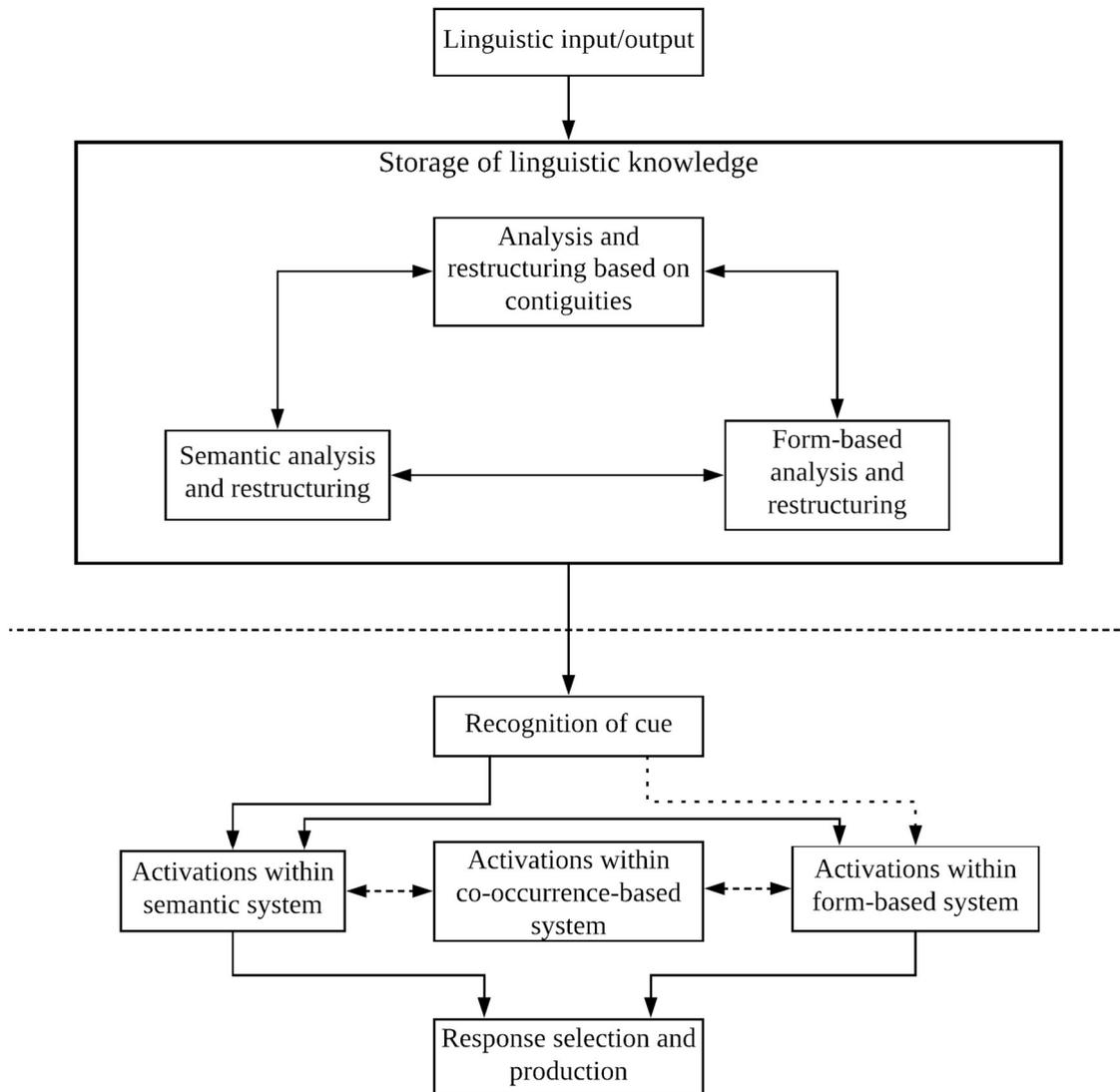
One of the reasons that a semantics-first usage-based model of word association is able to explain apparently position-based responses is that it posits that the semantic system underlying WA is itself derived in part from the interaction between semantic and distributional subsystems – something which occurs throughout the lifespan. This is evidenced by the influence of type/token information as presented in Chapter 8: while there was a significant correlation between the entropy of a word's textual distribution and its associative set size, the evidence from the brief qualitative comparison of corpus and WA data presented at the end of Chapter 8 did not support the view that this influence was caused by direct access to distributional knowledge.

Several examples from Chapter 8 show why a semantic interpretation of this data is preferable to a distributional one. Firstly, while several primary WA responses coincided with a very frequent collocation of the cue, as might be expected if distributional knowledge were a frequent source of WA responses, this in fact tended to occur only when the frequent collocates exemplified an aspect of the cue's core semantic content. For example, in primary cue-response pairs such as *erupt-volcano* and *purr-cat*, the response is semantically implied by the cue. Similarly, frequent collocates of these words which did not contain such core semantic entailments, such as *violence* and *riot* (for *erupt*) or *engine*, *car* and *voice* (for *purr*), were not present in the WA data.

Secondly, even amongst cues which had the most orderly verb-noun distributions (and were therefore hypothetically most likely to yield responses corresponding to the cue's most frequent collocates), there was little evidence of the application of distributional knowledge. The most frequent WA responses to these cues were paradigmatic, and only very rarely was the most frequent collocates of the cue produced by more than one WA participant. Moreover, the most frequent syntagmatic responses to these cues were again words entailed by the cue (e.g. *lick-tongue*, *irrigate-water*, *limp-leg*, *dawdle-slow* and *ache-pain*) rather than being frequent collocates. Indeed, in some cases these responses did not appear at all on the cue's list of verb-noun collocates.

What these examples suggest is that it is not sufficient for two words to collocate, however frequently, in order for them to become associated in the manner relevant to production of word association

Figure 9.1
 The second iteration of the composite model of word association response generation.



responses. A clear example, discussed in Chapter 8, is the relationship between the cue *happen* and its most frequent collocate, *thing*. Both of these words are semantically non-specific; neither contains core conceptual content which naturally entails the other. As such, despite the strong predictive relationship between these words in the corpus analysis, *thing* was only given as a WA response by a single participant, and is absent from the 150 responses to *happen* in the South Florida word association norms (D. L. Nelson et al., 2004; though the deletion of *hapax legomena* from this database means that *thing* may have been produced by one respondent). The interpretation of this phenomenon suggested by the above discussion is that *thing* did not appear among the associates of *happen* because it does not reflect the cue's semantics.

Support for the view that semantic relatedness is essential to WA response generation is provided by Prior & Bentin (2003), who investigated the processes by which words come to be associated. They found that, in an associative learning task, decontextualized word pairs were more difficult to remember, and were retained for less time, than those presented in sentential contexts. In a follow-up study (Prior & Bentin, 2008), they found that this sentential advantage was explained by integrative processing: the attempt to generate coherent semantic interpretations between words in the course of sentence processing led to stronger and more stable associations between the words in those sentences, compared with sentences which discouraged such processing.

Further studies support this suggestion that non-semantic associations are learned in only a superficial manner. For example, at least two associative learning studies (Dagenbach et al., 1990; Schrijnemakers & Raaijmakers, 1997) have found that while non-semantic associations can be learned, it takes considerable time to do so, and the resulting associative knowledge does not generalise to different psycholinguistic tasks from those by which the associations are learned. This suggests that such associations do not become deeply integrated into semantic memory.

The evidence presented above suggests that no direct access to distributional knowledge occurs during WA response generation. This in turn suggests that correlations between type/token distributions and associative set sizes are best explained as reflecting interactions between distributional and semantic knowledge which occurs throughout the lifespan, but *prior to*, not simultaneous with, the presentation of word association cues. This is explained by the processes of entrenchment with, or abstraction away from, a limited number of specific semantic relations between words (discussed in Chapter 7). Words which have narrow textual distributions become associated narrowly not only with individual collocations, but with the semantics implied by those collocations: thus *purr* is associated with not only *cat*, but also with cat-related words like *stroke* and *meow*; but is not associated with collocates such as *voice* or *car*. *Extinguish* is associated with not only *fire*, but also *flames* and *foam*; but not with collocates such as *rights*, *life*, or *hope*. Similarly, words which activate two or more distinct semantic senses tend to elicit multiple (often syntagmatic) responses related to each of those senses: *eject* yielded an “ejector seat” set (*seat*, *plane*, *cockpit*, *rocket*, *car*) and a distinct “media player” set: *button*, *CD*, *DVD*, *tape*, *video*, *disc*, *remove*, suggesting the activation not of collocations but of a semantic schema (see also Deese, 1962b, 1964, 1966). Words which possess few semantically informative collocations instead yield paradigmatic responses such as synonyms or members of the same lexical set, since no strong semantic relations have formed

between these words and their collocates. These effects can all be seen as evidence of a semantic imperative to word association responses.

9.4 A revised composite model

It has been argued above that the equality, hypothesised in Chapter 7, between the composite model's three subsystems (semantic, contiguity-based, and form-based) cannot be maintained in the face of available evidence. This does not imply, however, that other foundational aspects of the model are inaccurate. It suggests, instead, that while the model's three components do interact, and do all influence WA responses, the distributional component does so only indirectly, by influencing the semantic system. This is also the conclusion reached by McRae et. al. (2012, p. 57), who suggest that "higher than chance levels of co-occurrence between words are meaningful, and the semantic system takes advantage of this systematicity" by using it to construct semantic knowledge; semantically uninformative links between words are not retained.

These arguments necessitate a revision of the composite model as presented in Chapter 7, since it now appears likely to overestimate the contribution made to WA response generation by contiguity-based processing. This revision is presented in Figure 9.1. The new model shares with its previous iteration the division of linguistic knowledge into semantic, contiguity-based, and form-based components. Several changes have occurred, however. Firstly, the model has been split into two sections, with the upper segment (above the dotted line) depicting the general, lifelong process of acquiring, storing, and processing linguistic experience, including both the language which we read or hear and that which we produce. The storage of linguistic knowledge is seen here as a dynamic process of continual restructuring which occurs as a function of processing both within each component, and through interaction between them. Processing occurs continually, including during the integration of new linguistic experience into existing structures: word distributions influence semantic representations (as in Chapter 8), connections are formed between consonant clusters and semantic classes (as in the case of phonaesthemes; Bergen, 2004), and subtle shifts in word senses give rise to new possibilities for collocation and affixation (Hanks, 2013).

It is important to stress that these interactions are both highly individual and highly social. At the simplest level, this acknowledges the fact that no two individuals have exactly the same linguistic experiences throughout their lives. On a cognitive level, however, individuals also display preferences in the manner in which they process language (Bergen, 2012; Blajenkova, Kozhevnikov, & Motes, 2006; Kozhevnikov, 2007; Peterson, Deary, & Austin, 2005; Riding & Cheema, 1991). Blajenkova & Kozhevnikov (2009) have suggested, for example, that individuals differ in the extent to which they

make use of different types of visual and verbal information when processing language. Coupled with the unique nature of each person's linguistic experience, these differences explain the individual differences in response patterns demonstrated by Fitzpatrick (2007, 2009) and Morais et. al. (2013).

At the social level, diachronic changes in language usage are also assumed to impact upon the structure of linguistic knowledge in the mind. Such differences in linguistic input have been offered as one explanation of age-related differences in WA response patterns (Bøyum, 2016; see also Fitzpatrick et al., 2015), which suggest that similarly-aged respondents produce more homogeneous sets of WA responses, both in terms of the specific words given as responses and of the categorical types of responses provided, than when responses from a range of age groups are pooled.

The upper level of the model addresses the arguments of McRae et. al. (2012), above, by allowing for a clear delineation of the processes underlying, firstly, the perception of associations in linguistic experience and their subsequent integration into associative memory systems; and secondly, the processes involved in the generation of responses in word association tests. According to the model, contiguities between words are processed and stored by a distributional memory component. This knowledge is retained in this system for use during the production and comprehension of contextualised communication, in a manner such as that described by Hoey (2005); importantly, it is not activated directly during the process of WA response generation. Remaining in the upper level of the model, this distributional information is additionally processed by the semantic and form-based systems, which analyse and store cases in which similarity or semantic relevance is detected²⁴. The chances of such storage in semantic memory will increase with the regularity of the relationship between two words (as evidenced by the type/token effects described in Chapter 8), since repeated co-occurrences increase the likelihood that semantic integration of the words' meanings will occur. There is also likely to be an age-of-acquisition effect: earlier acquired links between words will be more likely to become entrenched within semantic memory. This possibility may provide a further explanation (beyond appealing to "core semantics": see Chapter 8, and Section 9.6.1, below) for why cues such as *erupt* yield responses related to volcanoes, but not to equally frequent, but perhaps later acquired, senses of *erupt* pertaining to violence, riots, and war.

²⁴ This is not to deny that associations may be learned through non-linguistic means, such as by repeatedly encountering objects such as tables and chairs, in spatial contiguity. However, such processes are yet to be researched, meaning that, in addition to the complexity which they would add to the model being described here, any assertions made as to their nature would be impossible to support empirically.

The lower section of Figure 9.1 shows the functioning of the linguistic system when put to the task of generating word association responses. Initial recognition of the cue leads to activations in both the semantic and form-based systems. The latter is activated as a consequence of the form-based processing involved in word recognition and lexical access. However, in line with research suggesting that phonological and orthographic priming effects occur at short latencies but disappear at longer ones (Ferrand & Grainger, 1993, 1994; Grainger, Kiyonaga, & Holcomb, 2006), this access is viewed as being brief (as indicated by the dotted line) and subsumed, in its later stages, by semantic processes. This means that while the form-based system rarely yields responses independently, this can occur when semantic processes yield no response (for example in cases where the cue is not known to the participant), or when a cue activates, for a given participant, a particularly strong association within the form-based system.

Activations in the semantic system immediately trigger response candidates. These semantic activations also trigger activations (of the same word) in the co-occurrence-based and form-based systems, if present. These parallel activations are facilitative in nature: they speed the selection of words represented in more than one linguistic system. This allows the model to explain the faster production of cue-response pairs linked in both meaning- and position-based manners reported by Fitzpatrick & Izura (2011), and may also explain why many form-based responses are also semantically related to the cue, such as in the case of affix-manipulation responses, or those categorizable as both form- and meaning-based.

9.5 Advantages of the composite model

The discussion above has focused on the capacity for the composite model to explain numerous WA research findings. These include the bias towards meaningfulness in both paradigmatic and syntagmatic responses, the occasional appearance of entirely form-based responses, numerous semantic and distributional network effects, the existence of type/token effects on responses, and individual and age-related patterns of response behaviour (Bøyum, 2016; Fitzpatrick, 2007, 2009; Fitzpatrick et al., 2015) and associative network structure (Morais et al., 2013).

There are, however, numerous other benefits to viewing WA in the manner outlined above. Chief among these is that the composite model fits word association research into a wider view of linguistic knowledge and processing derived from usage-based accounts of language. This way of viewing language has received support, in recent years, from numerous areas of psycholinguistics which have argued that generative models of language do not offer parsimonious explanations of the evolution of linguistic ability in humans, cannot explain numerous aspects of child language development, and

offer no explanation of the processes involved in language change (e.g. Behrens, 2009; Bergen, 2012; Bybee, 2008; N. C. Ellis, 2002; N. C. Ellis et al., 2016; Evans, 2006b; Evans, Bergen, & Zinken, 2007; Goldberg, 2003; Hanks, 2013; Langacker, 2000; Lieven & Tomasello, 2008; Pace-Sigge, 2018; Taylor, 2012; Tomasello, 2003).

It is a significant strength of usage-based models that they provide a holistic account of these and other linguistic phenomena, viewing language “as a complex adaptive system where language structure, acquisition, processing and change are inextricably intertwined in rich, complex, and dynamic ways” (N. C. Ellis et al., 2016, p. 23). Indeed, given the holistic aims of the usage-based perspective on language, it should be assumed that psycholinguistic phenomena such as word association can be incorporated into such a viewpoint. The composite model outlines how this might happen.

A further benefit of the usage-based architecture of the composite model is that it does not require any language- or WA-specific processes to be invoked in order to explain existing response patterns. The model makes use of general cognitive and perceptual processes including input recognition and discrimination and conceptual categorization, generalisation, and abstraction.

More generally, there has been a conspicuous absence of fully-fledged, empirically-derived models of word association ever since the 1960’s, when Deese began to challenge the long-held assumption that the associative structure of thought is essentially contiguity-based. The generative-transformational model of Clark (1970), described in Chapter 6, represents probably the most coherent attempt at the development of such a model in the intervening years. The failure of this model to account for important WA findings was quickly acknowledged however (K. Nelson, 1977), and few further attempts have been made. This lack of detailed models is a key reason for the rather aimless, assumption-prone state of recent WA research described in the introduction to this thesis, and it ultimately explains the efforts of some researchers to highlight the need for new research into the nature of word association (Hutchison, 2003; Hutchison et al., 2008; McRae et al., 2012). The composite model might therefore offer a greater purposefulness to future WA research.

9.6 Some unanswered questions

While the model presented above is broad in terms of its integration of word association into a wider, usage-based model of language, there are important aspects of the model and its relationship with WA research findings which remain unaddressed thus far. Some of these issues will be addressed below.

9.6.1 Which semantic knowledge?

Up to the present point, the conception of “semantic knowledge” used in this thesis has been a vague one: no attempt has been made to define exactly what aspects of lexical semantics are accessed during WA response generation. However, in Chapter 8 it was suggested that only limited aspects of cue semantics are accessed during this process. This suggestion was made in view of the fact that, for numerous cues, only one of several possible senses was referred to by WA respondents. This situation obtains for cues such as *erupt* and *purr*, discussed above, as well as others including *swoop*. This cue yielded responses which were heavily biased towards birds, or flight more generally. There were no responses which referred to the sense of policing (as in *drug swoop*), despite the fact that this sense occurs more frequently than the former in the British National Corpus.

The phenomenon of selective access to the lexicon is not a new discovery. Deese also noted that only limited aspects of a word’s meaning are produced in WA, writing that “*associative meaning* is a subset of the set *meaning*” (Deese, 1966, p. 43; Deese’s italics), while more recently Vivas et. al. (Vivas et al., 2018) have suggested that “WA tasks may be said to favour the recruitment of one specific subset of relations within semantic memory”. Exactly what this subset contains is, however, an unanswered question. In Chapter 8, the term “core semantics” was used to describe the apparent centrality of the most important semantic aspects of words to WA response generation. This term is, however, vague and in need of elaboration. Two questions need to be asked: firstly, what is the evidence that lexical access in (psycholinguistic tasks such as) word association is partial? And secondly, what does existing research suggest about the specific aspects of word meaning accessed during WA response generation? A short overview of some key research findings pertaining to these questions will be given below.

One initial piece of evidence pointing to the partial access of semantic knowledge in WA is provided by Barsalou’s research on context-dependent and context-independent word meanings (Barsalou, 1982). Barsalou argued that only core aspects of a word’s meaning are accessed during decontextualized tasks (a group which would doubtless include the word association task). These core aspects were termed context independent (CI), and include a word’s most salient properties. Other properties, termed context dependent (CD), are only accessed when activated by a particular context. Barsalou’s research, which involved property verification and object comparison tasks, suggests that CD information is accessed quickly in situations where the context suggests it, but more slowly and with lower accuracy where it does not. CI information, on the other hand, is accessed quickly and accurately regardless of context.

One problem with Barsalou's hypothesis is that it is not entirely clear exactly which properties of a word should be CI or CD. There is therefore a risk of circularity – CI properties should be activated in all situations; we know they are CI because they are activated in all situations. Barsalou's answer to this problem is to suggest that context-independence is not a fixed property of concepts, but a fuzzy, flexible one. He suggests that CI properties hinge upon four factors: diagnosticity, human-concept interaction, and frequency and recency of activation. Diagnosticity refers to the extent to which a property is "useful for distinguishing instances of a concept from instances of other concepts" (Barsalou, 1982, p. 82). For example, *gills* is CI for *fish*; it is diagnostic, since other animals do not have *gills* (*ibid.*). Human-concept interaction refers to the manner in which people interact with the concept (e.g. by eating it, watching it, or going to sleep on it). For example, Barsalou suggest that *edible* is a CI property of apples because it describes the typical way in which people interact with an apple.

The third and fourth determinants of context-independence – "the frequency and recency of processing episodes" (Barsalou, 1982, p. 91) – suggest that the extent to which words are, at any one time, context (in)dependent is gradable and mutable. This gives Barsalou's theory a distinctly usage-based tone. He suggests that frequent activation of a novel or CD property of a word can result in that property becoming context-independent. Thus, an individual's unique experiences – linguistic and otherwise – will lead to different properties being CI and CD for each person. While this is somewhat helpful from the point of view of explaining certain word association results (such as individual and age-related differences in responses), it means that CI and CD properties of words are fuzzy: they can only be estimated, not listed in their totality, and may vary between individuals.

Context-dependent semantic access and conceptual fuzziness are also features of the view of lexical semantics presented by Patrick Hanks in his Theory of Norms and Exploitations (TNE; Hanks, 2013). TNE is a "lexically based, corpus-driven, bottom-up theory of language" (*ibid.*; p17) which posits that word meanings can be understood on the basis of their patterns of pragmatic and collocational usage. Hanks posits that patterns of word co-occurrence can be viewed either as conforming to accepted linguistic conventions, referred to as *norms*; or as (rarer) *exploitations*, which constitute occasions on which a norm is exploited in order to present a word in a novel or non-standard manner:

If an expression is reused recurrently by different members of a language community (even only rarely), it is a norm, even if its internal structure is odd or incoherent. On the other hand, if we can find no evidence that the phraseology recurs—that is, if we are reasonably sure that the phrase is a creative coinage by the writer or speaker—it is an exploitation (p292).

Norms can include not only the most central aspects of a word's meaning (e.g. the sense of *buy* in the sentence "Lily bought a ticket to the opera"), but also more figurative (but conventional) uses of language such as "Geoffrey doesn't buy generative grammar" (*Ibid.*; p2). The diagnosis of the latter as a norm is based largely on its frequency; since most people are likely to be familiar with this usage, generating such a sentence is likely to require little creative thought. An exploitation, on the other hand, requires some inventiveness on the part of the speaker. Hanks (*Ibid.*; p412) discusses the example of "She fired an opening smile across Celia's desk" (taken from the BNC), which exploits both a military norm ("fire an opening shot") and a naval one ("fire a shot across the bows") in order to bring out a novel sense of the intent with which someone smiles.

One of the implications of Hanks' theory is that "words in isolation have meaning potential rather than meaning" (*Ibid.*; p65). For example, the word *buy* possesses the potential to mean "acquire by exchange of money". However, since "actual meanings are best seen as events, only coming into existence when people use words, putting them together in clauses and texts" (*Ibid.*), even this basic sense of *buy* is only activated when used communicatively (and, as the above examples show, may not be part of a given meaning event at all when used figuratively).

Hanks goes on to argue that word meanings are determined by usage, suggesting that "the semantics of each word in a language is determined by the totality of its complementation patterns" (*Ibid.*; p113). Hanks suggests that the syntactic complexity of a word tends to correspond to its semantic complexity. For example, so-called "light" verbs such as *take*, *make*, and *come* occur in such diverse contexts that any description of their typical syntactic patterns would fill many pages; as a result, these words tend to possess broad semantic potentials.

These points imply a significant amount of fuzziness to the meaning of words in isolation, since words are viewed as not acquiring distinct semantic senses until placed within a communicative and sentential context. While this is also true of Barsalou's model, TNE is perhaps more helpful, in terms of understanding WA response patterns, because it suggests that textual distributions of words will influence the level of specificity with which semantic potentials can be accessed in decontextualized psycholinguistic tasks: words which have many patterns of conventional usage will generally have less well-defined semantic potentials than words with few frequent patterns (note the similarity of this suggestion to the explanations of WA type/token effects given above). Additional work by Hanks & Pustejovsky has yielded a Pattern Dictionary of English Verbs (Hanks & Pustejovsky, 2005) which can be used together with WA data to test the impact of pattern diversity on WA responses.

While such an approach might be useful as a way to understand the relationship between word distributions and word association responses, however, TNE gets us only a little closer to understanding the specific aspects of lexical semantics accessed in WA. Firstly, TNE (along with Barsalou's theory of context-independence) suggests that the specific concepts activated during WA response generation are likely to be influenced by frequent and/or recent linguistic and perceptual experience. Secondly, since WA responses generally reflect literal interpretations of words (e.g. *erupt* elicits volcano-related responses, *swoop* those pertaining to birds and flight; *purr* to cats), and more figurative senses of words rarely figure in WA, even when corpora reveal frequent norms of usage pertaining to these senses (e.g. in a corpus, but not in WA, *violence erupts*, *police officers swoop*, *engines purr*, *computers hibernate*, and people *relieve soldiers*), we can hypothesise that the WA typically accesses only these literal senses of words. However, TNE does not help us to explain why these figurative senses do not occur in WA. This is because, according to Hanks (2013), figurative uses of words are, given sufficient frequency, no less norms than are literal usages. This leaves unresolved the conundrum of why WA responses tend to be typically literal interpretations of words, we must look elsewhere.

One model of lexical semantics which can account for the fuzziness implied by both Barsalou and Hanks' theories (and compatible with both), as well as suggesting a reason why semantic access in WA is predominantly literal, is prototype theory (Geeraerts, 2006; Mervis & Rosch, 1981; Rosch, 1973, 1975, 1998; Rosch & Mervis, 1975; Taylor, 2008). This view of lexical semantics emerged from research into colour words by Rosch (Heider, 1972), who found that some colours are more salient, and therefore easier to remember and more frequently encoded in various languages, than others. This implies that conceptual category membership is a gradable phenomenon, with fuzzy boundaries: some fruits, for example, are more "fruitlike" than others (e.g. apples, oranges), and there will be disagreement about borderline cases (olives, tomatoes).

There is a neuro-cognitive element to these findings. Research into colour perception (Berlin & Kay, 1969) suggests that the human eye is sensitive to colours of certain wavelengths, such that "optimum" reds, yellows, and blues (for example) are simply those which correspond to the eye's innate sensitivities (Taylor, 2003). This perceptual basis to colour terms has been extended to other areas of prototype theory (Taylor, 2008), and has been an influence on the development of cognitive views of language in general, underlying research, for example, into the perceptual basis of noun and verb grammar (Langacker, 1987, 2000; Shibotani, 1985; Taylor, 1998) and first and second language acquisition (Bybee, 2008, 2010; Ibbotson & Tomasello, 2009).

One implication of prototype theory is that decontextualized semantic access, for example in word association, may begin with access to prototypical senses of words. This is also suggested by Rosch's research: when asked to provide exemplars of given categories (e.g. *furniture*), respondents tend to give prototypical category members (e.g. *table*, *chair*). A similar phenomenon may apply to word association: a prototypical eruption, for example, would almost inevitably involve a *volcano* (the top WA response for *erupt*), rather than a *war* or a *riot* (frequent in corpora, but absent from WA), while a prototypical injection would involve using *needles* to inject *vaccines* or *drugs* (all frequent responses to *inject*), rather than institutions or individuals injecting *money* or *energy* into something (both among the top 10 collocates of *inject*, but absent in WA).

Indeed, these examples offer the possibility that the findings of Chapter 4, pertaining to the differences in response patterns for noun and verb cues, could be explained by prototype theory. It may be that prototypical conceptual representations of the events represented by verbs include information about typical agents, patients, and experiencers of the action denoted by the verb. Noun prototypes, on the other hand, may be more likely to include information about the properties of the entity in question, rather than the events with which it is associated.

However, more research is required into the parallels between prototypical concepts and word association responses. Such research might involve comparing judgements of prototypical properties of events or entities with WA responses, using words which have more than one distinct usage. Nevertheless, in the absence of such research the examples above and several others presented in Chapter 8 suggest that semantic access during word association may involve access primarily to prototypical representations of word meaning.

Thus far, this exploration of semantic access in WA response generation has suggested the following:

- Barsalou's theory of context-independence, Hanks' TNE, and prototype theory all support the hypothesis that only partial semantic access occurs in WA. This conclusion is in line with so-called "good enough" conceptions of lexical access, which posit that the depth of access to semantic representations is only as deep as required for completion of a task (Ferreira, Bailey, & Ferraro, 2002; Ferreira, Engelhardt, & Jones, 2009; Ferreira & Patson, 2007; Karimi & Ferreira, 2016).
- TNE and context-independence theory suggest that frequency and recency of linguistic and perceptual experience will influence the specific information accessed in decontextualized tasks.

- TNE additionally suggests that diversity of syntactic patterns will influence the specificity of semantic information which can be accessed.
- Both theories also suggest that semantic access in WA is likely to be dominated by non-figurative senses of words.
- Prototype theory suggests that prototypical senses of words may be accessed first; good-enough models of lexical access suggest that no further semantic access will be required if initial activations are sufficient for the generation of a response.
- Prototype theory can also explain differences in response patterns to nouns and verbs by suggesting that prototypical representations of verbs include information about agents, patients, and/or experiencers of the verb; noun representations rarely include such information.

These points offer a slightly more focused view of the concept of “core” semantic access discussed above. They pertain largely to the depth of access required by word association. It may be possible, however, to develop a greater specificity by looking at those aspects of concepts most likely to be returned by WA respondents. That is to say, if a WA cue (e.g. *inject*) tends to activate prototypical aspects of a concept (i.e. injections involve doctors or nurses administering medicine or vaccinations to patients, using needles), which aspects of these concepts in particular are most likely to be selected by respondents?

The simplest answer to this question is: it depends on the cue. Two studies (Guida & Lenci, 2007; Schulte Im Walde et al., 2008) have suggested that cues of different semantic types (e.g. verbs of motion, perception, or weather) elicit responses of different types (i.e. grammatical classes, thematic roles, or conceptual relations) in varying proportions. Given these findings, it appears that any adequate description of semantic access in WA will have to consider the role of cue semantics in determining semantic activations. Few studies to date have attempted this sensitivity of analysis on a large scale: such an approach is deserving of further investigation.

Several studies have, however, attempted to more closely describe WA semantic access on a more general level. At least two such studies have drawn on the concept of feature lists as sources of WA responses. This concept is derived from classical, Aristotelian views of semantics, which assert that words and conceptual categories can be reduced to finite lists of criterial properties. For example, the concept “bird” can be defined using a list of features such as [+animate], [+has wings], and [+has a beak]. Such theories suffer from the problem that they cannot explain the phenomenon of graded class membership demonstrated by Rosch in her research on prototypes; in classical semantics, there

is a clear demarcation between members and non-members of a category, defined by whether or not the concept in question possesses all of the criterial properties of its category (Geeraerts, 2009; Taylor, 2003). Prototype theory, by contrast, argues that no such list of necessary and sufficient properties can cover all members of a class. Instead, categories display “family resemblances”: each item in the category possesses properties which overlap with one or more properties of other category members; but no properties are common to all items (Rosch & Mervis, 1975; Wittgenstein, 1953).

Perhaps the first WA study to have made explicit use of classical feature lists is the generative-transformational model of Clark, initially discussed in Chapter 6. From the point of view of lexical semantics, this model hypothesises that feature lists are called upon in WA response generation as the cognitive system applies transformations to cue words. For example, the minimal contrast rule consists of the transformation of a cue using the most specific of the word’s features (e.g. in the case of *man*, this feature is [+male]; its transformation results in *woman*). The feature addition/deletion rule similarly calls upon feature lists in order to perform transformations upon cues. As detailed above, however, Clark’s model fails to explain numerous key WA findings.

A second study to draw on conceptual features as a source of WA responses is that of Vivas et. al. (2018), who compared responses to 199 WA cues (all concrete nouns) with word feature lists collected separately. They found that that 86% of primary WA responses also occurred in the lists of cue conceptual features (e.g. *axe-wood*). Furthermore, 72.5% of all responses to these nouns were described by participants as medium or highly defining features of the cue (e.g. *apple-red*). These features included not only object properties but also situational relations (e.g. *dog-bone*, *worm-soil*). In total, 42.8% of primary responses corresponded to the most relevant feature of the cue.

This data suggests that criterial or prototypical cue features, including situational properties, are likely to be amongst the semantic aspects of words activated in WA. This important insight notwithstanding, however, Vivas et. al.’s research leaves numerous questions unanswered. Firstly, it does not allow us to distinguish between classical feature-list and prototype-based views of lexical access, since both of these models could account for the high overlap between word features and WA responses. Secondly, it does not elaborate the relationship between cue semantics and these aspects, described above and in the work of Guida & Lenci (2009). Finally, even if access to conceptual features explains the majority of WA responses, there remains a significant percentage of responses in Vivas et. al.’s study for which this explanation does not apply. As such, a model of semantic access in WA purely based on access to conceptual features is likely to be inadequate.

One final possibility will be discussed here: that WA responses may derive, at least in part, from access to embodied, modality specific semantic representations. Such views of linguistic representation derive from research suggesting that language is subserved by neural structures underlying perception and action (Bergen, 2012; Bergen & Feldman, 2006; Bergen, Lau, Narayan, Stojanovic, & Wheeler, 2010; Bergen, Lindsay, Matlock, & Narayanan, 2007; Borghi & Binkofski, 2014; A. Clark, 2006; Kemmerer, 2015; Kemmerer, Miller, Macpherson, Huber, & Tranel, 2013; M. Louwerse & Hutchinson, 2012; M. M. Louwerse & Jeuniaux, 2010; Lynott & Connell, 2009; Macgregor, Pulvermüller, Casteren, & Shtyrov, 2012; Pulvermüller, 2013; Pulvermüller, Härle, & Hummel, 2001; Pulvermüller, Shtyrov, & Ilmoniemi, 2005; Sidhu, Kwan, Pexman, & Siakaluk, 2014; Wheeler & Bergen, 2010; Wilson-Mendenhall et al., 2011; Zwaan, 2003).

Research into the precise mechanisms underlying modality-specific language processing are still the subject of much debate (see e.g. Jackson, Hoffman, Pobric, & Lambon Ralph, 2015; Vigliocco et al., 2011; Wilson, 2002). However, one attempt has been made to gather evidence in support of one particular model of embodied language processing through word association studies. This model was the Language and Situated Simulation (LASS) model developed by Barsalou (Barsalou, 2008a, 2008b). LASS proposes that two sub-systems underlie language processing. The first is a network derived from linguistic experience. This network contains not only information stored from contiguities between words, but also semantic knowledge derived by comparing the extent to which word distributions overlap. Several researchers have suggested that models derived from the combination of these types of information result in computational networks with similar properties to human associative networks (Andrews et al., 2009, 2008; Gruenenfelder et al., 2016).

The second element of the LASS model is an embodied, situational simulation system. The basis of this system is modality-specific memory: as concepts and entities are encountered in the real world (Barsalou et al., 2008, give the example of a dog), memories are formed within visual, olfactory, auditory, gustatory and somatosensory systems, as well as emotion centres in the brain; actions result in the creation of memories related to motor activity. Later linguistic experiences (such as reading the word *dog*) bring about a partial reactivation of these memories. This modality-specific information is then available for further cognitive processing of linguistic content. In psycholinguistic research settings, this processing can include the generation or verification of object properties of the sort not available through the linguistic system. This happens through the automatic generation of visual knowledge, such as the recollection of internal (e.g. *watermelon-red*) or situational (e.g. *golf-sunshine*)

properties of words (Barsalou, Solomon, & Wu, 1999; Solomon & Barsalou, 2004; Wu & Barsalou, 2009).

Two important aspects of the LASS model are that the linguistic system should work more quickly than the situated simulation system, since it requires access only to statistical knowledge of contextual word knowledge where the situated simulation system requires the generation of detailed mental reconstructions of previous experiences; and that different brain regions should subserve the two processes. In two experiments, Barsalou and colleagues (Santos et al., 2011; Simmons et al., 2008) have attempted to demonstrate these effects; their studies focused in part on word association experiments. In the first (Simmons et al., 2008), functional magnetic resonance imaging (fMRI) was used to demonstrate that in the early stages (i.e. the first 7.5 seconds) of a property generation task, active brain regions corresponded to those activated during word association (assumed here by the authors to indicate access largely to linguistic processing). However, during the later stages of the same task (i.e. the next 7.5 seconds), other brain regions, corresponding to those activated during situation generation tasks, became active. This result suggests the neural and temporal separability of linguistic and simulation-based processing.

In the second study (Santos et al., 2011, Experiment 1), it was assumed that both linguistic and simulation-based processes would be active during the generation of WA responses. The authors hypothesised, however, that responses originating from the linguistic system would be identifiable as such: they would be collocations, compound completions, sound similarity associations, and synonyms and antonyms. Responses generated through simulation, on the other hand, would constitute descriptions of objects or situations; this would include cue-response pairs such as *bee-wings/summer/flowers* and *golf-sunshine/boring/Jack Nicklaus* (*Ibid.*; p116). Consequently, it should be the case that such associations are produced more slowly than those underpinned by linguistic processes. Santos et al.'s results supported this hypothesis.

Taken together, the results of these two studies suggest that semantic access during word association occurs on two levels – a fast-access linguistic level derived solely from linguistic experience, and a slower, situated simulation-based system. As suggested above, however, not all embodied models of language support this two-way divide between linguistic and embodied conceptual knowledge. Rogers et al. (2004), for instance, have suggested that although both linguistic and perceptual experience are utilised in semantic memory, semantic processing is underpinned not by the two processes operating independently, but by a single semantic system which integrates knowledge from various different sources. This hypothesis, though untested on word association, is supported by research on semantic

dementia (Jackson et al., 2015). It is also possible to criticise the Santos et. al. (2011) study on the grounds of the replicability of their coding schemes, since some of the examples which they describe as likely to have originated from simulation-based processing are also likely to co-occur. For example, *bee-flowers*, given by Santos et. al. as an example of situational processing, is (if *flower* and *flowers* are treated as one) the most common collocational pairing of *bee* in the British National Corpus. It would thus, under Santos et. al.'s system, be very likely to originate from the linguistic system. As such, in spite of isolated studies supporting the claims of Barsalou and his colleagues (De Deyne & Storms, 2008), there remains insufficient evidence to accept LASS as a model of semantic access in WA.

There remains considerable potential, however, in using ideas from embodied psycholinguistics to investigate the nature of word association. One approach which may bear fruit is the use of perceptual strength measures to determine the types of word knowledge – visual, auditory, olfactory etc. – most likely to be accessed during WA response generation. These values have been shown to outperform similar measures such as concreteness and imageability in accounting for response times in priming tasks (Connell & Lynott, 2012), suggesting their importance to lexical processing. Furthermore, the emergence of large-scale WA networks (e.g. De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2016) means that the contribution of certain word properties to network centrality and associative assortativity (Van Rensbergen et al., 2015) is now a relatively uncomplicated procedure. These methods will, for instance, allow researchers to understand whether cue perceptual properties (e.g. the auditory content in *bang*) exert a probabilistic influence on the same properties of their responses.

The foregoing discussion reveals that, although much research remains to be done in the area of semantic access in word association, it is possible to draw an approximate sketch of some of the processes involved. Firstly, semantic access appears to be partial: not all aspects of a word's meaning are accessed every time the word is comprehended, including during WA tests. Secondly, the aspects of a word's conceptual representation accessed earliest are likely to be those which are most prototypical to the word's meaning. Figurative uses of words, even when frequently encountered in natural language use, are rarely (if ever) reproduced in word association. Thirdly, cue semantics influence response semantics. Words from different semantic classes vary in the extent to which they yield synonyms and antonyms, taxonomic associations, or thematic cue properties (Guida & Lenci, 2007; Schulte Im Walde et al., 2008); cue properties including concreteness and affective strength select for the same properties in their responses (Van Rensbergen et al., 2015). Finally, it is plausible that word associations reflect two discrete processing systems, one derived from linguistic knowledge

and the other from embodied, modality-specific world knowledge; but competing models argue against this view by proposing that a single semantic system integrates these forms of knowledge.

In view of the uncertainty regarding this latter aspect of semantic access, the composite model in its present iteration avoids making any assertions regarding the role of conceptual knowledge in WA. Instead, it views the knowledge from which WA responses are drawn as deriving from the combination of linguistic input and cognitive processing only. It will be left to future versions of the model to specify how modality-specific perceptual experience influences responses. As such, the model can be considered potentially compatible with either single- or dual-route theories of semantic processing. In the case of the former, perceptual information is viewed as being integrated with the linguistic knowledge described in the composite model; in the latter, it exists as a discrete, independently accessed system.

Future research will need to distinguish between these two possibilities. In the following section, it will be suggested that two key findings from earlier WA research can be used as theoretical tests of such models in order to establish their plausibility as explanations of WA response generation.

9.6.2 Two “litmus tests” for word association models

9.6.2.1 The syntagmatic-paradigmatic shift

One of the most important of the findings of the word association work of the 1960’s was the discovery of a shift from syntagmatic to paradigmatic responses in children responding to cues in their first language. Among the most comprehensive descriptions of this phenomenon was that given by Entwisle (Entwisle, 1966b, 1966a; Entwisle et al., 1964), who described a modest shift towards paradigmatic responding to nouns between kindergarten (61.2% paradigmatic) and their peak in fifth grade (78.1%), with a small drop-off at college age (77%). This was more pronounced for cues of other grammatical classes. For adjective cues, paradigmatic responding increased from 16.8% in kindergarten to 78.5% in fifth grade, before dropping to 65.8% in college. For verb cues, the increase was from 16.6% in kindergarten to 59.6% in fifth grade: there was no drop-off at college age. Entwisle suggested that the drop-off in paradigmatic responding, particularly to adjective cues, by college age was caused by “late” syntagmatic responses qualitatively different from those produced by younger children. Following McNeill (McNeill, 1964), she suggested that young children generate syntagmatic responses which have little underlying semantic relationship with their cues. Those of adults, on the other hand, reflect integration of semantic and co-occurrence-based knowledge.

Since the publication of studies such as these, numerous researchers have attempted to explain the syntagmatic-paradigmatic shift. A broad distinction can be made between theories which have seen the shift as reflecting the changing state of linguistic knowledge, and those which posit general maturational changes as its cause. Examples of the former category include the model of Ervin (1961), who suggested that increasing linguistic experience allows children to perceive the similarities of context in which words occur, and therefore use a substitutability judgement to generate responses; and that of Cronin (2002), who suggested that the shift could be explained by changes in the nature of word knowledge attributable to progress in reading proficiency and exposure to written text. Katherine Nelson (1977), on the other hand, suggests that general changes in cognitive preferences – particularly a shift towards increasingly logical, as opposed to associative – thinking is behind the shift.

The various competing theories of the syntagmatic-paradigmatic shift will not be described in detail here (K. Nelson, 1977, provides a comprehensive review of then-current ones). Instead, it is suggested that the syntagmatic-paradigmatic shift represents one of the two most well-established word association findings (the other is to be described below). New models of word association, including the composite model described in this thesis, therefore need to be capable of providing plausible explanations of these findings: the ability to do so may be considered a litmus test for models of word association.

While the capacity for the composite model to explain the syntagmatic-paradigmatic shift is a matter for future research, it is possible to sketch a brief hypothesis of how such an explanation might work, based on the processes of entrenchment and abstraction described above, and drawing on the properties of complex dynamic systems assumed to underlie linguistic representation in usage-based models of language (Beckner et al., 2009; N. C. Ellis, 2002; N. C. Ellis & Larsen-Freeman, 2009; N. C. Ellis et al., 2016; Larsen-Freeman, 2006; Schmid, Kopke, & de Bot, 2012). As children encounter language, numerous processes occur. These include the storage of contextual and contiguity-based knowledge of words, the discrimination of form-function relationships, and the integration of individual word meanings into coherent semantic representations. All of these processes contribute to the developing linguistic system which is drawn upon in word association. As the composite model makes clear, this system is in a state of constant development throughout the lifespan.

From this point of view, one of the key aspects of the development of the semantic system is the extent and diversity of input received. Small amounts of repetitive input (such as that which children might receive, or be able to comprehend) will lead to the entrenchment of word pairings and their semantics. In terms of word association, this would likely lead to the production of syntagmatic

responses (since, as Chapter 8 showed, there is a statistically reliable relationship between low type variation and the production of noun responses to verb cues). However, as children grow older and are exposed (e.g. through reading, cf. Cronin, 2002) to increased quantity and diversity of linguistic experience, the semantic system responds by developing increasingly abstracted representations of language, since type variation increases as a result of this broadening linguistic experience. This would in turn result in a greater number of paradigmatic responses, since more abstract word representations tend to result in the production of synonyms and antonyms.

This is only a very brief sketch of how the composite model could explain the syntagmatic-paradigmatic shift: it is not intended as anything but an invitation to future research. It might be possible to test the above hypothesis, however, by investigating the impact of increases in word type variation as children mature. It may be possible, for example, to identify words which are used relatively narrowly in early childhood (e.g. *hold hands*), but which broaden markedly in later childhood (e.g. *hold* becomes increasingly associated with other objects and events, and perhaps also increasingly dissociated with *hands*). Such cases should show marked syntagmatic responding in young children (i.e. the production of *hold-hands*), but a shift towards synonym, antonym, or taxonomic responses later, as a function of increasing abstraction away from earlier, entrenched, representations.

9.6.2.2 Individual response profiles

A second “litmus test” for word association concerns the explanation of Fitzpatrick’s discovery of individual response preferences in word association, already described extensively in this thesis and replicated in Chapter 2. This finding has now been demonstrated in three published studies (Fitzpatrick, 2007, 2009; Higginbotham, 2010), as well as in Chapter 2 of this thesis. These studies have involved both L1 respondents and L2 participants at varying levels of proficiency, as well as cue words of varied frequencies and grammatical classes (Higginbotham, 2014), was established by coding responses into detailed categories showing the types of semantic and position-based relations between words (e.g. synonyms, members of the same lexical set, and directional relationships between position-based responses), rather than only semantic responses (as would be expected according to the composite model).

Simply put, if such a scheme had no psycholinguistic validity, it would be incapable of identifying the individual preference patterns revealed by Fitzpatrick. As such, insofar as these patterns are robust and replicable, new models of WA will fall short if they are unable to apply coding schemes which are capable of revealing individual response preferences at least as well as Fitzpatrick’s categories.

Without such a finding, it could be argued that the scheme does not capture the psycholinguistic reality of word association, since it does not reflect the preferences which individual respondents are now known to bring to the WA task.

In theory, it should not be difficult for a scheme to be devised in compliance with the semantic imperative asserted in the composite model. In fact, as described above, predominantly semantic schemes of coding word association responses have been developed by numerous scholars (Guida & Lenci, 2007; McRae et al., 2012; Vivas et al., 2018). Such schemes largely preserve Fitzpatrick's meaning-based categories, while reassigning position-based responses to new categories defined as, for example, semantic features of words (e.g. *fire-hot*) or in terms of thematic roles such as agent, patient, and experiencer (Guida & Lenci, 2007). The most appropriate way of testing these schemes, however, is not to determine the extent to which they can encompass the very great majority of responses (as has been done in previous studies), but to determine (as described above) the extent to which they reveal individual response preferences. This imperative applies not only to the semantic scheme required of the composite model, but other schemes as well. This includes the LASS-based coding model devised by Santos et. al. (2011), which was criticised above on the grounds that it may be difficult for coders to apply.

9.7 Implications for further research

In describing these unanswered questions regarding the composite model, the above discussion has also provided a rough sketch of some directions from which future research can contribute to the development of understanding about word association and its contribution to psycholinguistics in general. These possible avenues of research include:

- Investigating the nature of semantic access in word association, for example by:
 - Identifying the semantic cue-response relationships most common to cues of different grammatical classes (or other properties, such as concreteness),
 - Further investigating the semantic properties of words most likely to be produced in WA,
 - Further investigating how the properties of cue words, such as their concreteness and affective strength, influence response patterns,
 - Investigating the determinants of equivalent- (i.e. synonyms) and non-equivalent-meaning associations (taxonomic or syntagmatic responses), particularly with regard to type/token distributions and lexical specificity,

- Investigating the relationship between type/token distribution, syntactic pattern diversity (Hanks, 2013; Hanks & Pustejovsky, 2005) and lexical specificity (Guida & Lenci, 2007). As suggested in Chapter 8, usage-based models of language see these properties as being strongly related, and they are likely to be highly correlated. Nevertheless, it may be possible to assert that one is more important than the other in the determination of WA responses,
 - Developing a deeper understanding of when, how, and why cue semantics influence the properties of responses,
 - Using network models to explore the properties of cues which do and do not develop centrality to the associative network,
 - Alternatively, using WA networks to explore co-variation between cue and response variables, including perceptual properties.
- Investigating potential age-of-acquisition effects in the entrenchment of associations in semantic memory – e.g. by looking at the age at which *associations*, rather than individual words, are acquired, and the consequences for adult association.
 - Investigating the syntagmatic-paradigmatic shift from a usage-based angle, with particular regard to the way in which type/token frequency influences lexical representation in young children.
 - Investigating the extent to which fully semantic methods of categorizing word association responses can account for individual WA response preferences.

9.7.1 Methodology and the rehabilitation of the WA method for psycholinguistic research

It is possible to express these discrete research aims as part of a more general aim for WA research, based on the appeals of several researchers to elaborate the nature of word association before making use of it as an independent variable in psycholinguistic studies (Hutchison, 2003; Hutchison et al., 2008; McRae et al., 2012). That aim is to rehabilitate word association for use in psycholinguistic research by demonstrating the underlying nature of the measures which it provides.

Such a research program would be very large and encompass a very wide range of empirical questions. It would require the use of numerous research methods, including large-scale associative network analysis, both on group and individual levels, large-scale collection and analysis of WA response-time data; and a large number of more focused, small scale studies investigating specific aspects of new WA models. The diversity of measurements available to WA researchers (see Table 3.1) will be a

significant challenge to this research program, since it will massively increase the amount of data to be collected and analysed; it must, however, be viewed as a benefit to the WA method, since the analysis of associative networks is now beginning to show that each measure reveals different aspects of the linguistic system underlying the generation of WA responses.

Though this goal for future research is undoubtedly very large, it would have many benefits. Firstly, in the field of psycholinguistics, priming studies frequently make use of word association norms to reference the semantic relatedness between words, without clearly delineating exactly what this semantic relation entails. A better understanding of word association would therefore help researchers to better understand the nature of priming effects. This would in turn be beneficial for research which makes use of priming methods, such as the study of various forms of dementia and the nature of lexical knowledge in general.

A second benefit of a rehabilitated perception of word association is that it may facilitate renewed research into the second language (L2) opportunities presented by the method. Numerous attempts have been made to find ways of testing second language lexical knowledge through the use of WA-based tests (Bahar et al., 1999; Dronjic & Helms-Park, 2014; Meara & Fitzpatrick, 2000; Munby, 2011; Read, 1993, 1995, 2000; Wolter, 2002), but they have suffered from problems related to the unpredictability of responses and unclear relationships between proficiency and association. A deeper understanding of the determinants of, and processes underlying, L1 word association might aid in this endeavour.

Lastly, a deeper understanding of what is meant when words are referred to as being “associated” is an interesting and valuable research project in its own regard, facilitating (as demonstrated by this thesis) discussion and hypothesis development regarding the manner in which cognitive functioning and linguistic experience drive the organisation, access, and production of lexical knowledge.

Chapter 10: Conclusion

This thesis identified, and has sought to address, a perceived failure of previous research to explore untested assumptions and unexplained variation in word association (WA) responses. In doing so, the thesis has made several original contributions to both methodological and psycholinguistic aspects of the word association task, and to our understanding of the nature of the mental lexicon. In this final chapter, these original findings will be summarised from the perspective of the contribution they make to the state of research on word association. Subsequently, a number of limitations of the work presented here will be discussed.

The majority of the original findings within this thesis relate to the role of lexical variables in the word association task. Previous research into the influence on word association responses of these variables, which include word frequency, concreteness, affective strength, grammatical class, and age-of-acquisition (AoA), was reviewed in Chapter 3. Three main findings stand out from that chapter. Firstly, there is evidence that all of the above variables influence the WA task. Secondly, however, the influence of each of them is not visible to every method of measuring WA (e.g. response time, homogeneity; see Table 3.1). The influence of some variables can only be discerned through a single measure, pointing to a specific locus for its effects. For example, word frequency effects are negligible or inconsistent on all measures except those which gauge network centrality, such as in-degree (that is, the number of WA cues which yielded a given word as a response; see Table 3.1); this points to an influence on the system underlying how words become associated, rather than on the process of generating a WA response. In the case of other variables, however, broad effects can be discerned across numerous variables (as in the case of grammatical class; see Chapter 3). As such, a range of measures needs to be tested in order to discern the full nature of the effect of any given variable.

Thirdly, approaching WA responses from a network perspective reveals two distinct types of lexical variable effect. The first refers to the influence of the lexical properties of cue words upon the pattern of their responses. For instance, given a cue with high concreteness, it is likely that responses will be relatively homogeneous, produced relatively quickly, and be higher in concreteness, than if the cue had been low in concreteness. The cue properties with the strongest influence upon responses are largely semantic ones: in addition to concreteness (and imageability, since these have largely been treated in WA research as synonymous), significant effects upon responses have been reported for cue grammatical class and affective strength. The influence of distributional variables such as frequency and AoA upon response patterns is, by comparison, weak and inconsistent.

Viewing WA responses from the perspective of local relations between cue and response words, as above, has been termed a *microscopic* level of analysis (De Deyne & Storms, 2015). The second type of word variable influence refers the properties of words frequently given as responses, independently of the cues which elicited them. Such properties can be revealed only through large-scale network models; this approach is termed *macroscopic* (Ibid.). By contrast to the microscopic, this type of analysis reveals that distributional factors (i.e. frequency and AoA) are key determinants of network centrality. For example, early acquired and highly frequent words are produced as responses to a wider range of cues than are later acquired or less frequent ones (that is, they have a higher “in-degree”; De Deyne & Storms, 2008, and see Chapter 3).

One caveat to these findings is that the research which supports them is rather fragmentary – not nearly enough work has attempted to systematically explore such variation. As such, it is important that more research is conducted into the relationship between distributional and semantic information within the mental lexicon. It is also important to note that without the statistical and theoretical power provided by network models of WA, the importance of distributional variables on WA would be very difficult to discern. As such, the wider use of network methods, supported by methodologies which collect a broad range of dependent measures, is highly desirable in future WA work.

Another problem identified in Chapter 3 with regard to the state of research into lexical variable influence on WA was that very few studies have investigated these variables in a way which demonstrates their independence from related properties. In Chapter 4, an attempt was made to address this problem as it applies to research into grammatical class influence on WA. A study was conducted to explore differences in response patterns to verb and noun cues after potentially confounding variables such as frequency, concreteness, and AoA have been controlled. The results suggest that cue grammatical class is an independent determinant of the categorical type of responses, with nouns yielding more meaning-based responses than verbs, and verbs more position-based responses than nouns. One particularly interesting finding was that while noun cues were equally likely to yield position-based associations in either direction (i.e. cue-response or response-cue associations), verb cues yielded significantly more cue-response ones. In Chapter 5, it was found that verb transitivity is a key element of this pattern – transitive cues yield more cue-response associations; intransitives a greater proportion of response-cue ones. In addition to these basic findings, Chapters 4 and 5 show that relatively small-scale factorial studies expand upon previous empirical results have the potential to yield important findings for future WA research.

The attempt to explain this transitivity effect theoretically, begun in Chapter 6, was a key driver of the eventual development of the new model of word association presented in its final iteration in Chapter 9. Since no existing model of is capable of explaining directional effects of cue transitivity (see Chapter 6), a usage-based (UB) model was developed with this aim. The first iteration of this model hypothesised equality of influence from each of three memory systems – semantic, distributional, and form-based. However, an attempt to provide evidence for this theory by using a comparison of corpus and WA data to demonstrate a distributional influence on position-based responses was successful only in casting doubt on the assumption that human WA response codings accurately reflect the psycholinguistic processes through which WA responses are generated.

As a result, a second attempt at comparing WA and corpus data (Chapter 8) avoided the use of human coding in favour of comparing cue type/token distributions with WA response patterns. The results of this final study, which showed that the textual distribution of noun collocates of cue verbs influences the verbs' WA response patterns, was the basis for the development of a second iteration of the usage-based model of WA. A key feature of this model, which was described in Chapter 9, is that it asserts a distinction between the processes which govern the gradual, lifelong formation of associations in the mental lexicon from those which handle the generation of responses in word association tests. According to the model, although the associative network underlying WA is essentially semantic in nature, it is formed through the combination of linguistic experience and cognitive processing. It is therefore sensitive to distributional information such as word frequency and complementation patterns. Empirical and theoretical research has established at least two ways in which this can happen. Firstly, Steyvers & Tenenbaum (2005) have used computational modelling to suggest that early acquired and frequent words take on the role, within the mental lexicon, of anchors to which newly acquired words attach (see also De Deyne & Storms, 2008; Morais et al., 2013, for empirical views of this theory). Secondly, the work presented in Chapter 8 of the current thesis suggests that a word's textual type/token distributions influence the extent to which word meaning is narrowly associated with a limited range of semantic potentials, on one hand, or abstracted away from any fixed semantic representation, on the other. These findings together suggest a usage-based, distributionally-influenced nature to the mental lexicon.

That this network obtains part of its structure from distributional information does not, however, mean that the process of WA response generation is also distributionally influenced. In addition to the above results, it was suggested in Chapter 8 that it is not the case that frequent collocates are necessarily also frequent WA responses, since the most common collocates of WA cues only appear

to correspond to frequent WA responses when there are additionally strong semantic relationships between the words. Instead, the model suggests that the process of generating WA responses from this network is semantic in nature and does not directly draw upon distributional, predictive word knowledge. This explains the strong influence of semantic variables on WA response patterns, and the negligible direct effect of distributional ones, described above as well as in Chapter 3. According to this model, the apparently position-based transitivity effects revealed in Chapter 5 are therefore actually semantic: transitives are more likely to yield words which denote the patient or experiencer of the event denoted by the verb, while intransitives are more likely to receive responses corresponding to their agent.

In summary, the research presented in this thesis has contributed not only a number of relatively small-scale empirical findings (i.e. the independence of grammatical class effects from other variables, the discovery of verb transitivity effects, and the finding of type/token influence on associative distributions and response types) and methodological insights (i.e. the contribution of various WA measures to understanding of the lexical network, the importance of a network perspective to WA research, the need for control of multiple variables, and the value of investigating underexplored variation in WA findings), but also an important and powerful new model which can be used to generate and test hypotheses about word association and the mental lexicon. An agenda for future research, based on this new model, was set out in Section 9.6.

There are, inevitably, limitations to the work described above. Some of these, such as the need for future work to elaborate the specific details of the type of semantic representations accessed in WA response generation, have already been discussed in Chapter 9. A number of limitations remain, however. Firstly, while the model described in Chapter 9 is based on a wide research base not limited to the studies conducted as part of this thesis, it must be noted that the findings presented in Chapters 5 and 8, used as support for the model, are based on findings from a single set of WA data, collected using cues selected on the grounds of their transitivity. While these cues were carefully chosen for the purposes of investigating the research questions formulated in Chapter 5, it is possible that the need to select equal numbers of transitive and intransitive verbs made them somewhat atypical with regard to verbs in general. As such, although similar findings presented by Hahn & Sivley (2011) support the findings of Chapter 8, it will nevertheless be important for future studies to verify these findings of using cues of more varied grammatical profiles.

Secondly, it is a limitation of all empirical work that only a limited number of participants can take part in research tasks, and only a limited number of test items can be used in those tasks. This leaves all

studies, including the original experiments presented in Chapters 2, 4, 5, 7, and 8, vulnerable to the impact of individual response preferences. While effort has been made to ensure that the conclusions drawn from these experiments are theoretically justified and empirically supported as far as possible, there nevertheless always remains a possibility that different participants, on a different occasion, would produce results which would not support these conclusions.

Thirdly, it has, throughout this thesis, been difficult to present accounts of WA response generation that would stand as a credible alternative to the usage-based model described here. This is because very few models more recent than the generative-transformational account given by Clark (1970) are available. It is not, however, the intention of the present thesis to suggest that generative models of language in general, including those more recent than Clark's WA model, are incapable of explaining current WA findings. Instead, it is viewed as unfortunate that no up-to-date non-UB account of word association is available for comparison with the model developed in Chapter 9, since such a possibility for comparison would almost certainly help to sharpen the account given here.

Finally, all empirical work requires a commitment to a chosen research path. One example is the decision, earlier in this thesis, to pursue a comparison between WA and corpus-derived data, rather than following another direction, such as exploring other lexical variables. The commitment to one pathway inevitably closes off another: it may be the case that the paths not chosen in this thesis could have led to more useful research outcomes than the ones which have emerged. In addition, both the usefulness of these chosen pathways, and the research studies designed to explore them, are vulnerable to unrecognised biases and assumptions held by the researcher.

One specific way in which these points relate to the present thesis is through the decision not to collect a broader range of data – particularly WA response times – for each study. This decision inevitably cuts off a potentially interesting research pathway. On the other hand, the aim of this thesis has been to test specific research questions rather than to collect normative data; there is virtually no limit to the data which could potentially be collected, and which might reveal important patterns, but effective research requires the discipline of testing clear research questions and interpreting the outcomes with confidence that the methods chosen are those which will yield the most valuable results, and with circumspection about what those choices leave untested.

With this said, it is hoped that the research presented here offers an opportunity for future researchers to approach word association, if not with more certainty about its nature, then with the benefit of a principled, theoretically justified account through which new hypotheses may be

generated and tested. From this model, new accounts may be arrived at which are less vulnerable to the methodological limitations and unconscious biases which may have unwittingly influenced the work presented here.

Bibliography

- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review*, 23(3), 275–290.
- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times. *Psychological Science*, 17(9), 814–823.
- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (Fourth). Chichester: John Wiley & Sons.
- Akhtar, N. (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26(2), 339–356.
- Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and image ability ratings and word associations for abstract, concrete, and emotion words. *Behaviour Research Methods, Instruments, & Computers*, 31(4), 578–602.
- Andrews, M., Vigliocco, G., & Vinson, D. P. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116(3), 463–498.
- Andrews, M., Vinson, D. P., & Vigliocco, G. (2008). Inferring a Probabilistic Model of Semantic Memory from Word Association Norms. In *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1941–1946).
- Ares, G., & Deliza, R. (2010). Studying the influence of package shape and colour on consumer expectations of milk desserts using word association and conjoint analysis. *Food Quality and Preference*, 21(8), 930–937.
- Asch, S. E., & Ebenholtz, S. M. (1962). The Principle of Associative Symmetry. *Proceedings of the American Philosophical Society*, 106(2), 135–163.
- Bahar, M., Johnstone, a. H., & Sutcliffe, R. G. (1999). Investigation of students' cognitive structure in elementary genetics through word association tests. *Journal of Biological Education*, 33(3), 134–141.
- Balota, D. A., Cortese, M., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental ...*, 133(2), 283–316.
- Balota, D. A., Yap, M. J., & Hutchison, K. A. (2007). The English lexicon project. *Behavior Research ...*, 39(3), 445–459.
- Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? *Visual Word Recognition: Models and Methods, Orthography and Phonology*, 2, 90–115.
- Barber, H. A., Otten, L. J., Kousta, S.-T. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, 125(1), 47–53.
- Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10(1), 82–93.
- Barsalou, L. W. (2008a). Grounded Cognition. *Annual Review of Psychology*, 59, 617–645.
- Barsalou, L. W. (2008b). Grounding Symbolic Operations in the Brain's Modal Systems. In G. R. Semin

- & E. R. Smith (Eds.), *Embodied Grounding* (pp. 9–42). Cambridge: Cambridge University Press.
- 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 (Eds.), *Symbols, Embodiment, and Meaning* (pp. 245–283). Oxford: Oxford University Press.
- Barsalou, L. W., Solomon, K. O., & Wu, L. L. (1999). Perceptual simulation in conceptual tasks. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, Psychological and Typological Issues in Cognitive Linguistics* (pp. 209–228). Amsterdam: John Benjamins Publishing Company.
- Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Christiansen, M. H., ... Schoenemann, T. (2009). Language Is a Complex Adaptive System. *Language Learning*, 59(Supplement 1), 1–26.
- Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, 47(2), 383–411.
- Bel Enguix, G., Rapp, R., & Zock, M. (2014a). A Graph-Based Approach for Computing Free Word Associations. *Proceedings of the 9th Edition of the Language, Resources and Evaluation Conference (LREC 2014)*, 3027–3033.
- Bel Enguix, G., Rapp, R., & Zock, M. (2014b). How Well Can a Corpus-Derived Co-Occurrence Network Simulate Human Associative Behavior? *Proceedings of 5th Workshop on Cognitive Aspects of Computational Language Learning (CogACL)@EACL 2014*, 43–48.
- Benedek, M., & Neubauer, A. C. (2013). Revisiting Mednick's Model on Creativity-Related Differences in Associative Hierarchies. Evidence for a Common Path to Uncommon Thought. *The Journal of Creative Behavior*, 47(4), 273–289.
- Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80(2), 290–311.
- Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. New York: Basic Books.
- Bergen, B. K., & Feldman, J. (2006). It's the body, stupid! concept learning according to cognitive science. ... , *Stupid: Concept of Learning According ...*, 1–18.
- Bergen, B. K., Lau, T.-T. C., Narayan, S., Stojanovic, D., & Wheeler, K. (2010). Body part representations in verbal semantics. *Memory & Cognition*, 38(7), 969–981.
- Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, 31(5), 733–764.
- Berlin, B., & Kay, P. (1969). Basic Colour Terms: Their Universality and Evolution. *University of California Press*. Berkeley: University of California Press.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33(1), 73–79.
- Blajenkova, O., & Kozhevnikov, M. (2009). The New Object-Spatial-Verbal Cognitive Style Model: Theory and Measurement. *Applied Cognitive Psychology*, 23(5), 638–665.
- Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology*, 20(2), 239–263.
- Bonner, M. F., Vesely, L., Price, C., Anderson, C., Richmond, L., Farag, C., ... Grossman, M. (2009).

- Reversal of the concreteness effect in semantic dementia. *Cognitive Neuropsychology*, 26(6), 568–579.
- Borenstein, M. (2009). Effect sizes for continuous data. In *The Handbook of Research Synthesis and Meta-Analysis* (pp. 221–235).
- Borghini, A. M., & Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. New York: Springer.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229–240.
- Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. (1958). Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychological Reports*, 4(g), 39–44.
- Bøyum, M. (2016). *A new word association test for Norwegian*. University of Oslo.
- Bozic, M., Tyler, L. K., Su, L., Wingfield, C., & Marslen-Wilson, W. D. (2013). Neurobiological Systems for Lexical Representation and Analysis in English. *Journal of Cognitive Neuroscience*, 25(10), 1678–1691.
- Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review*, 15(6), 1035–1053.
- Brown, R., & Berko, J. (1960). Word Association and the Acquisition of Grammar. *Child Development*, 31(1), 1–14.
- Brown, W. P. (1971). A retrospective study of stimulus variables in word association. *Journal of Verbal Learning and Verbal Behavior*, 10(4), 355–366.
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424.
- Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, 13(7–8), 992–1011.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990.
- 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.
- Brysbaert, M., Wijnendaele, I. Van, & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, 104(2), 215–226.
- Bybee, J. (2003). Mechanisms of change in grammaticalization: The role of frequency. *The Handbook of Historical Linguistics*.
- Bybee, J. (2006). From Usage to Grammar: The Mind’s Response to Repetition. *Language*, 82(4), 711–733.
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 216–236). New York, NY: Taylor & Francis.

- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, J., & Beckner, C. (2010). Usage-based theory. In B. Heine & H. Narrog (Eds.), *Oxford Handbook of Linguistic Analysis* (pp. 827–856). Oxford: Oxford University Press.
- Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 25(1), 85–95.
- Catling, J. C., & Johnston, R. A. (2009). The varying effects of age of acquisition. *Quarterly Journal of Experimental Psychology (2006)*, 62(1), 50–62.
- Charles, W. G., & Miller, G. A. (1989). Contexts of antonymous adjectives. *Applied Psycholinguistics*, 10(03), 357–375.
- Chomsky, N. (1957). *Syntactic Structures*. *Syntax*. The Hague: Mouton de Gruyter.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. *Aspects of the Theory of Syntax*. Cambridge, M.A.: The MIT Press.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Comput. Linguist.*, 16(1), 22–29.
- Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8), 370–374.
- Clark, H. H. (1970). Word Associations and linguistic theory. In *New Horizons in Linguistics* (pp. 271–286). Harmondsworth: Penguin.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Statistical Power Analysis for the Behavioral Sciences*.
- Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 82(6), 407–428.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33(A), 497–505.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125(3), 452–465.
- Connell, L., & Lynott, D. (2014). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General*.
- Consortium. (2007). *The British National Corpus, Version 3 (BNC XML Edition)*.
- Cools, E., & Van den Broeck, H. (2007). Development and validation of the cognitive style indicator. *The Journal of Psychology*, 141(4), 359–387.
- Cordier, F., Croizet, J. C., & Rigalleau, F. (2013). Comparing Nouns and Verbs in a Lexical Task. *Journal of Psycholinguistic Research*, 42(1), 21–35.
- Cousins, K. A. Q., York, C., Bauer, L., & Grossman, M. (2016). Cognitive and anatomic double dissociation in the representation of concrete and abstract words in semantic variant and behavioral variant frontotemporal degeneration. *Neuropsychologia*, 84, 244–251.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.

- Cramer, P. (1968). *Word Association*. New York, NY: Academic Press.
- Croft, W. (1990). A conceptual framework for grammatical categories (or: A taxonomy of propositional acts). *Journal of Semantics*, 7(3), 245–279.
- Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Cronin, V. S. (2002). The syntagmatic-paradigmatic shift and reading development. *Journal of Child Language*, 29(1), 189–204.
- Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding New Information to Semantic Memory: How Much Learning Is Enough to Produce Automatic Priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 581–591.
- Davis, C. J. (2005). N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37(1), 65–70.
- De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. J. (2016). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Big Data in Cognitive Science* (pp. 174–202). Hove: Psychology Press.
- De Deyne, S., Leopold, S., Navarro, D. J., Perfors, A., & Storms, G. (2012). Strong structure in weak semantic similarity: a graph based account. In *Building Bridges Across Cognitive Sciences Around the World: Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 1464–1469).
- De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2016). Measuring the associative structure of English: The “Small World of Words” norms for word association, 1–26.
- De Deyne, S., Navarro, D. J., & Storms, G. (2012). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, 45(2), 480–498.
- De Deyne, S., & Storms, G. (2008). Word associations: network and semantic properties. *Behavior Research Methods*, 40(1), 213–231.
- De Deyne, S., & Storms, G. (2015). Word Associations. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 465–480). Oxford: Oxford University Press.
- de Groot, A. M. B. (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.
- DeDe, G. (2010). Utilization of prosodic information in syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, 39(4), 345–374.
- DeDe, G. (2014). Verb Transitivity Bias Affects On-line Sentence Reading in People with Aphasia. *Aphasiology*, 27(3), 326–343.
- Deese, J. (1962a). Form class and the determinants of association. *Journal of Verbal Learning and Verbal Behavior*, 1, 79–84.

- Deese, J. (1962b). On the structure of associative meaning. *Psychological Review*, 69(3), 161–175.
- Deese, J. (1964). The associative structure of some common english adjectives. *Journal of Verbal Learning and Verbal Behavior*, 3(5), 347–357.
- Deese, J. (1966). The structure of associations in language and thought.
- Diependael, K., Grainger, J., & Sandra, D. (2012). Derivational Morphology and Skilled Reading: An Empirical Overview. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 311–332). Cambridge: Cambridge University Press.
- Dronjic, V., & Helms-Park, R. (2014). Fixed-choice word-association tasks as second-language lexical tests: What native-speaker performance reveals about their potential weaknesses. *Applied Psycholinguistics*, 35(1), 193–221.
- Druks, J. (2002). Verbs and nouns - A review of the literature. *Journal of Neurolinguistics*, 15(3–5), 289–315.
- Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, 53(8), 1560–1570.
- Duñabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin & Review*, 15(6), 1072–1077.
- Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2), 125–155.
- Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 26(5), 1103–1123.
- Ellis, N. C. (2002). Frequency effects in Language Processing. *Studies in Second Language Acquisition*, 24(02).
- Ellis, N. C., & Larsen-Freeman, D. (Eds.). (2009). *Language as a Complex Adaptive System*. Chichester: John Wiley & Sons.
- Ellis, N. C., Romer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of Construction Grammar*. Chichester: John Wiley & Sons.
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 8(7), 301–306.
- Entwisle, D. R. (1966a). Form Class and Children's Word Associations. *Journal of Verbal Learning and Verbal Behaviour*, 5, 558–565.
- Entwisle, D. R. (1966b). *Word associations of young children*. Baltimore: John Hopkins Press.
- Entwisle, D. R., Forsyth, D. F., & Muuss, R. (1964). The syntactic-paradigmatic shift in children's word associations. *Journal of Verbal Learning and Verbal Behavior*, 3, 19–29.
- Epstein, M. L., Szymanski, C. A., & Daggett, S. Z. (1977). Asch and Ebenholtz Revisited: A Test of the Doctrine of Associative Symmetry. *Journal of General Psychology*, 96(1), 111–116.
- Ervin, S. M. (1961). Changes with Age in the Verbal Determinants of Word-Association. *The American*

Journal of Psychology, 74(3), 361–372.

- Ervin, S. M., & Miller, W. R. (1963). Language Development. In *Child Psychology* (pp. 108–143). Chicago: University of Chicago Press.
- Eustache, F., Cox, C., Brandt, J., Lechevalier, B., & Pons, L. (1990). Word association responses and severity of dementia in Alzheimer disease. *Psychological Reports*, 66, 1315–1322.
- Evans, V. (2006a). Lexical concepts, cognitive models and meaning-construction. *Cognitive Linguistics*, 17(4), 491–534.
- Evans, V. (2006b). Towards a cognitive compositional semantics : An overview of LCCM theory. *New Insights in Semantics and Lexicography University of MarieCurie Press Poland*, 1–47.
- Evans, V. (2007). *A Glossary of Cognitive Linguistics*. Edinburgh: Edinburgh University Press.
- Evans, V., Bergen, B. K., & Zinken, J. (2007). *The Cognitive Linguistics Reader*.
- Evert, S. (2004). *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. University of Stuttgart.
- Evert, S. (2008). Corpora and collocations. *Corpus Linguistics. An International Handbook*, 1–53.
- Eysenck, H. J. (1994). Creativity and personality: Word association, origence, and psychoticism. *Creativity Research Journal*, 7(2), 209–216.
- Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition, (2).
- Ferrand, L., & Grainger, J. (1994). Effects of Orthography Are Independent of Phonology in Masked Form Priming, (2).
- Ferrand, L., & New, B. (2003). *Semantic and Associative Priming in the Mental Lexicon*. (P. Bonin, Ed.), *The Mental Lexicon: "Some Words to Talk About Words."* New York: Nova Publishers.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15.
- Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good Enough Language Processing: A Satisficing Approach. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, (1), 413–418.
- Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, 1(1–2), 71–83.
- Fillmore, C. J. (1988). The mechanisms of "construction grammar." *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 14, 35–55.
- 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. 121–145).
- Fitzpatrick, T. (2007). Word association patterns: unpacking the assumptions. *International Journal of Applied Linguistics*, 17(3), 319–331.
- 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).

- Fitzpatrick, T. (2012). Tracking the changes : vocabulary acquisition in the study abroad context, *40*(1), 81–98.
- 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.
- 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.
- Gahl, S. (2002). Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology*, *16*(12), 1173–1198.
- Geeraerts, D. (2006). Prospects and problems of prototype theory. In D. Geeraerts, R. Dirven, J. R. Taylor, & R. W. Langacker (Eds.), *Cognitive Linguistics: Basic Readings* (pp. 141–166). Den Haag: Mouton de Gruyter.
- Geeraerts, D. (2009). *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA.: MIT Press.
- Gentner, D. (2006). Why Verbs Are Hard to Learn. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). Oxford: Oxford University Press.
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281.
- Gerola, A. (2016). Review: Michael Tomasello, A Natural History of Human Morality. *Rivista Internazionale Di Filosofia E Psicologia*, *7*(2), 286–288.
- Gewirth, L. R., Shindler, A. G., & Hier, D. B. (1984). Altered patterns of word associations in dementia and aphasia. *Brain and Language*, *21*(2), 307–317.
- Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, *71*(1), 105–110.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.
- Goldberg, A. E. (2006). The inherent semantics of argument structure: The case of the English ditransitive construction. In D. Geeraerts, R. Dirven, J. R. Taylor, & R. W. Langacker (Eds.), *Cognitive Linguistics: Basic Readings* (pp. 401–438). Den Haag: Mouton de Gruyter.
- Gollan, T. H., Salmon, D. P., & Paxton, J. L. (2006). Word association in early Alzheimer's disease. *Brain and Language*, *99*(3), 289–303.
- Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology*, *61*(3), 348–353.
- Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The Time Course of and Phonological Orthographic

- Code Activation. *Psychological Science*, 17(12), 1021–1026.
- Gries, S. T. (2010). Useful statistics for corpus linguistics. In Aquilino Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: selected approaches* (pp. 269–291). Frankfurt am Main: Peter Lang.
- Gries, S. T., & Ellis, N. C. (2015). Statistical Measures for Usage-Based Linguistics. *Language Learning*, 65(S1), 228–255.
- Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language*, 38(3), 313–338.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211–244.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), 3–15.
- Grossman, L., & Eagle, M. (1970). Synonymity, antonymity, and association in false recognition responses. *Journal of Experimental Psychology*, 83, 244–248.
- 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.
- Guerrero, L., Claret, A., Verbeke, W., Enderli, G., Zakowska-Biemans, S., Vanhonacker, F., ... Hersleth, M. (2010). Perception of traditional food products in six European regions using free word association. *Food Quality and Preference*, 21(2), 225–233.
- Guida, A., & Lenci, A. (2007). Semantic properties of word associations to Italian verbs. *Italian Journal of Linguistics*, 19(2007), 293–326.
- Gulacar, O., Sinan, O., Bowman, C. R., & Yildirim, Y. (2014). Exploring the Changes in Students' Understanding of the Scientific Method Using Word Associations. *Research in Science Education*.
- Hahn, L. W., & Sivley, R. M. (2011). Entropy, semantic relatedness and proximity. *Behavior Research Methods*, 43(3), 746–760.
- Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge, MA.: The MIT Press.
- Hanks, P., & Pustejovsky, J. (2005). A Pattern Dictionary for Natural Language Processing. *Revue Française de Linguistique Appliquée*, 10(2), 63–82.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1), 10–20.
- Heyman, T., Hutchison, K. A., & Storms, G. (2015). Uncovering underlying processes of semantic priming by correlating item-level effects. *Psychonomic Bulletin & Review*, 540–547.
- Higginbotham, G. M. (2010). Individual learner profiles from word association tests : The effect of word frequency. *System*, 38(3), 379–390.
- Higginbotham, G. M. (2014). *Individual profiling of second language learners through word association*. Swansea University.
- Hirsh, K. W., & Tree, J. J. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, 14(1), 1–44.

- Hoaglin, D. C., & Iglewicz, B. (1987). Fine-Tuning Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*, 82(400), 1147.
- Hodgson, J. M. (1991). Informational constraints on pre-lexical priming. *Language and Cognitive Processes*, 6(3), 169–205.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3), 718–730.
- Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology. Human Perception and Performance*, 41(2), 385–402.
- Hopper, P. J., & Thompson, S. A. (1980). Transitivity in Grammar and Discourse. *Language*.
- Hunston, S. (2002). Corpora in Applied Linguistics. *ELT Journal*, 57, 416–420.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins Publishing Company.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10(4), 785–813.
- Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *The Quarterly Journal of Experimental Psychology*, 61(7), 1036–1066.
- Ibbotson, P., & Tomasello, M. (2009). Prototype constructions in early language acquisition. *Language and Cognition*, 1(1), 59–85.
- Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language*, 50(2), 165–181.
- Izura, C., & Playfoot, D. (2012). A normative study of acronyms and acronym naming. *Behavior Research Methods*, 44(3), 862–889.
- Jackson, R. L., Hoffman, P., Pobric, G., & Lambon Ralph, M. A. (2015). The Nature and Neural Correlates of Semantic Association versus Conceptual Similarity. *Cerebral Cortex*, 25(11), 4319–4333.
- Jakubíček, M., Kilgariff, A., & Kovář, V. (2013). The tenten corpus family. *Proceedings of the 7th International Corpus Linguistics Conference*.
- James, W. (1890). *The Principles of Psychology* (Vol. 2). Read Books Ltd.
- Jenkins, J. J. (1954). Transitional Organization: Association Techniques. In C. Osgood & T. Sebeok (Eds.), *Psycholinguistics: a survey of theory and research problems (Supplement to The Journal of Abnormal and Social Psychology)* (Vol. 49, pp. 114–116). Baltimore: Waverly Press.
- Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and Language*, 74(1), 103–112.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*,

66(2), 115–124.

- Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, 55(4), 534–552.
- Jorm, A. F. (1991). The validity of word age-of-acquisition ratings: A longitudinal study of a child's word knowledge. *British Journal of Developmental Psychology*, 9, 437–444.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2000). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Hopkins Press.
- Justeson, J. S., & Katz, S. M. (1991). Co-occurrences of Antonymous Adjectives and Their Contexts. *Computational Linguistics*, 17(1), 1–19.
- Kang, B. (2018). Collocation and word association: Comparing collocation measuring methods. *International Journal of Corpus Linguistics*, 23(1), 85–113.
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013–1040.
- Kemmerer, D. (2014). Word classes in the brain: Implications of linguistic typology for cognitive neuroscience. *Cortex*, 58, 27–51.
- Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, 22(4), 1068–1075.
- Kemmerer, D., & Eggleston, A. (2010). Nouns and verbs in the brain: Implications of linguistic typology for cognitive neuroscience. *Lingua*, 120(12), 2686–2690.
- Kemmerer, D., Miller, L., Macpherson, M. K., Huber, J., & Tranel, D. (2013). An investigation of semantic similarity judgments about action and non-action verbs in Parkinson's disease: implications for the Embodied Cognition Framework. *Frontiers in Human Neuroscience*, 7(April), 146.
- Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Insanity*.
- Keuleers, E., & Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology*, 68(8), 1457–1468.
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44(1), 287–304.
- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7–36.
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitkin, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153–165). Edinburgh: Edinburgh University Press.
- Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, 131, 20–244

- Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. *Psychological Bulletin*, 133(3), 464–481.
- Kruse, H., Pankhurst, J., & Smith, M. S. (1987). A Multiple Word Association Probe in Second Language Acquisition Research. *Studies in Second Language Acquisition*, 9(02), 141–154.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
- Kuperman, V. (2015). Virtual experiments in megastudies: A case study of language and emotion. *Quarterly Journal of Experimental Psychology*, 68(8), 1693–1710.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978–990.
- 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.
- Langacker, R. W. (1987). Nouns and Verbs. *Language*, 63(1), 53–94.
- Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63). Stanford: CSLI.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of english. *Applied Linguistics*, 27(4).
- Leech, G., & Rayson, P. (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Routledge.
- Lemaire, B., & Denhière, G. (2006). Effects of High-Order Co-occurrences on Word Semantic Similarity. *Current Psychology Letters - Behaviour, Brain, and Cognition*, 18(1).
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA.: MIT Press.
- Lewis, M. B. (1999a). Age of acquisition in face categorisation: Is there an instance-based account? *Cognition*, 71(1), 23–39.
- Lewis, M. B. (1999b). Are age-of-acquisition effects cumulative-frequency effects in disguise? A reply to Moore, Valentine and Turner (1999). *Cognition*, 72(3), 311–316.
- Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition*, 78(2), 189–205.
- Lieven, E., & Tomasello, M. (2008). Children's first-language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 168–196). New York, NY: Taylor & Francis.
- Louwerse, M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology*, 3(OCT), 1–11.
- Louwerse, M. M., & Connell, L. (2011). A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words. *Cognitive Science*, 35, 381–398.
- Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing.

Cognition, 114(1), 96–104.

- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*, 7(4), 618–630.
- Lynott, D., & Connell, L. (2009). Embodied conceptual combination. *Embodied and Grounded Cognition*.
- Ma, W., Michnick Golinkoff, R., Hirsh-Pasek, K., McDonough, C., Tardif, T., Hirsh-Pasek, K., ... Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language*, 36(2), 405–423.
- MacDonald, M. C. (1994). Probabilistic Constraints and Syntactic Ambiguity Resolution. *Language and Cognitive Processes*, 9(2), 157–201.
- Macgregor, L. J., Pulvermüller, F., Casteren, M. Van, & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, 3(711).
- Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101(1), 3–33.
- Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development*, 20(1), 121–136.
- McDonald, S., & Lowe, W. (1998). Modelling Functional Priming and the Associative Boost. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 675–680). Hillsdale, NJ: Erlbaum.
- McDonough, C., Song, L., Hirsh-Pasek, K., Michnick Golinkoff, R., Lannon, R., Hirsh-Pasek, K., ... Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science*, 14(2), 181–189.
- McGregor, K. K., Marian, V., & Sheng, L. (2006). Lexical – Semantic Organization in Bilingual Children : Evidence From a Repeated Word Association Task. *Journal of Speech, Language, and Hearing Research*, 49(June 2006), 572–587.
- McKoon, G., & Ratcliff, R. (1992). Spreading Activation Versus Compound Cue Accounts of Priming : Mediated Priming Revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6), 1155–1172.
- 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.
- McNeill, D. (1964). *Is child language semantically consistent?* Harvard University.
- McNeill, D. (1966). A study of word association. *Journal of Verbal Learning and Verbal Behavior*, 5, 548–557.
- McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and Associative Relations in Adolescents and Young Adults: 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). Washington, DC: American Psychological Association.
- Meara, P. (1978). Learners' Word Associations in French. *Interlanguage Studies Bulletin-Utrecht*, 3(2), 192–211.

- Meara, P. (1983). Word associations in a foreign language. *NOTTINGHAM LINGUISTIC CIRCULAR*.
- Meara, P. (2004). Modelling Vocabulary Loss. *Applied Linguistics*, 25(2), 137–155.
- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: Company, John Benjamins Publishing.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19–30.
- Merten, T. (2010). Factors Influencing Word-Association Responses: A Reanalysis. *Creativity Research Journal*.
- Merten, T., & Fischer, I. (1999). Creativity, personality and word association responses: Associative behaviour in forty supposedly creative persons. *Personality and Individual Differences*, 27(5), 933–942.
- Mervis, C., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, 32(1), 89–115.
- Miller, G. A., Beckwith, R., Fellbaum, C., & August, R. (1993). Introduction to WordNet : An On-line Lexical Database. *International Journal of Lexicography*, 3(August), 235–244.
- Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition*, 41(1–3), 197–229.
- Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 34(1), 65–79.
- Mitchell, D. (1987). Lexical guidance in human parsing: Locus and processing characteristics.
- Mohr, C., Graves, R. E., Gianotti, L. R., Pizzagalli, D., & Brugger, P. (2001). Loose but normal: a semantic association study. *Journal of Psycholinguistic Research*, 30(5), 475–483.
- Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory*.
- Morais, A. S., Olsson, H., & Schooler, L. J. (2013). Mapping the Structure of Semantic Memory. *Cognitive Science*, 37(1), 125–145.
- 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 ...*, 21(4), 863–883.
- Munby, I. (2011). *Development of a multiple response word association test for learners of English as an L2*. Swansea University.
- Murphy, K., & Hunt, H. (2013). The time course of semantic and associative priming effects is different in an attentional blink task. *Cognitive Processing*, 14(3), 283–292.
- Namei, S. (2004). Bilingual lexical development: A Persian–Swedish word association study. *International Journal of Applied Linguistics*, 14(3), 363–388.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition*, 28(6), 887–899.

- 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, ...*
- Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological Bulletin, 84*(1), 93–116.
- Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics, 16*(3), 389–408.
- Noble, C. E. (1952). An analysis of meaning. *Psychological Review, 59*(6), 421–430.
- Norvig, P. (2012). Colorless green ideas learn furiously. *Significance, 30*–33.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 786–803.
- Pace-Sigge, M. (2013). *Lexical priming in spoken English usage*. London: Palgrave Macmillan.
- Pace-Sigge, M. (2018). *Spreading Activation, Lexical Priming and the Semantic Web: Early Psycholinguistic Theories, Corpus Linguistics and AI Applications*. London: Palgrave Macmillan.
- Paivio, A. (1991). Dual Coding Theory: Retrospect and Current Status. *Canadian Journal of Psychology, 45*(3), 255–287.
- Paivio, A., Yuille, J., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness: Values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement, 76*(1/2), 1–25.
- Pecher, D., & Raaijmakers, J. G. W. (1999). Automatic Priming Effects for New Associations in Lexical Decision and Perceptual Indentification. *The Quarterly Journal of Experimental Psychology, 52A*(3), 593–614.
- Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does Pizza Prime Coin? Perceptual Priming in Lexical Decision and Pronunciation. *Journal of Memory and Language, 38*(4), 401–418.
- Perea, M., Gotor, A., & Nacher, M. J. (1997). Efectos de facilitación asociativa vs. semántica con una breve asincronía estimular señal-test [Associative and semantic priming effects at a very brief stimulus-onset asynchrony]. *Psicothema, 9*(3), 509–517.
- Peterson, E. R., Deary, I. J., & Austin, E. J. (2005). A new measure of Verbal-Imagery Cognitive Style: VICS. *Personality and Individual Differences, 38*(6), 1269–1281.
- Petrey, S. (1977). Word associations and the development of lexical memory. *Cognition, 5*(1), 57–71.
- Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual Word Recognition Volume 2: Meaning and Context, Individuals and Development* (pp. 24–43). Hove: Psychology Press.
- Playfoot, D., & Izura, C. (2013). Imageability, age of acquisition, and frequency factors in acronym comprehension. *Quarterly Journal of Experimental Psychology, 66*(6), 1131–1145.
- Politzer, R. L. (1978). Paradigmatic and syntagmatic associations of first year French students. In V. Honsa & M. Hardman-de-Bautista (Eds.), *Papers in Linguistics and Child Language: Ruth Hirsch Weir memorial Volume* (pp. 203–210). Den Haag: Mouton de Gruyter.
- Polzella, D. J., & Rohrman, N. L. (1970). Psychological aspects of transitive verbs. *Journal of Verbal*

Learning and Verbal Behavior, 9(5), 537–540.

- Porte, G. (Ed.). (2012). *Replication Research in Applied Linguistics*. Cambridge: Cambridge University Press.
- Postman, L. J. (1970). The California norms: Association as a function of word frequency. In L. J. Postman & G. Keppel (Eds.), *Norms of Word Association* (pp. 241–320). New York: Academic Press.
- Postman, L. J., & Keppel, G. (1970). *Norms of Word Association*. New York NY: Academic Press.
- Prior, A., & Bentin, S. (2003). Incidental formation of episodic associations: The importance of sentential context. *Memory & Cognition*, 31(2), 306–316.
- Prior, A., & Bentin, S. (2008). Word associations are formed incidentally during sentential semantic integration. *Acta Psychologica*, 127(1), 57–71.
- Prior, A., & Geffet, M. (2003). Mutual information and semantic similarity as predictors of word association strength: Modulation by association type and semantic relation. In F. Schmalhofer, R. M. Young, & G. Katz (Eds.), *Proceedings of the European Cognitive Science Conference 2003* (pp. 265–271). Mahwah, NJ: Lawrence Erlbaum.
- Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*.
- Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking? Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, 78(2), 143–168.
- Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain Signatures of Meaning Access in Action Word Recognition. *Journal of Cognitive*, 17(6), 884–893.
- Pustejovsky, J. (1991). The Generative Lexicon. *Language*, 17(4), 409–441.
- Qian, D. D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, 52(3), 513–536.
- Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21(1), 28–52.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355–371.
- Read, J. (1995). Refining the word associates format as a measure of depth of vocabulary knowledge. In *17th Language Testing Research Colloquium, Long Beach, CA*.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Riding, R., & Cheema, I. (1991). Cognitive Styles - an overview and integration. *Educational Psychology*, 11(3/4), 193–215.
- Riegel, K. F., & Zivian, I. W. M. (1972). A study of inter- and intralingual associations in English and German. *Language Learning*, 22(1), 51–63.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111(1), 205–235.

- Roininen, K., Arvola, A., & Lähteenmäki, L. (2006). Exploring consumers' perceptions of local food with two different qualitative techniques: Laddering and word association. *Food Quality and Preference*, 17(1–2), 20–30.
- Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, 4, 328–350.
- Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, 104(3), 192–233.
- Rosch, E. (1998). Principles of Categorization. In G. Mather, F. Verstraten, & S. Anstis (Eds.), *The Motion Aftereffect* (pp. 251–270). Cambridge, M.A., M.A.: MIT Press.
- Rosch, E., & Mervis, C. (1975). Family resemblances. *Cognitive Psychology*, 7, 573–605.
- 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.
- Saporta, S. (1955). Linguistic structure as a factor and as a measure in word association. *Associative Processes in Verbal Behavior: A Report of ...*
- Saussure, F. de. (1916). *A Course in General Linguistics* (Roy Harris, Trans.). London: Duckworth.
- Schmid, M. S., Kopke, B., & de Bot, K. (2012). Language attrition as a complex, non-linear development. *International Journal of Bilingualism*.
- Schmitt, N. (1998). Quantifying word association responses: what is nativelike? *System*, (26), 389–401.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N., Ng, J. W. C., & Garras, J. (2011). The Word Associates Format: Validation evidence. *Language Testing*, 28(1), 105–126.
- Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods*, 44(4), 971–977.
- Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, 25(2), 211–236.
- Schrijnemakers, J. M. C., & Raaijmakers, J. G. W. (1997). Adding new word associations to semantic memory: Evidence for two interactive learning components. *Acta Psychologica*, 96(1–2), 103–132.
- Schulte im Walde, S., & Melinger, A. (2005). Identifying semantic relations and functional properties of human verb associations. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, (October), 612–619.
- 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, 205–238.
- Shibotani, M. (1985). Passives and Related Constructions: A Prototype Analysis. *Language*, 61(4), 821–848.
- Sidhu, D. M., Heard, A., & Pexman, P. M. (2016). Is More Always Better for Verbs? Semantic Richness

Effects and Verb Meaning. *Frontiers in Psychology*, 7(May).

- Sidhu, D. M., Kwan, R., Pexman, P. M., & Siakaluk, P. D. (2014). Effects of relative embodiment in lexical and semantic processing of verbs. *Acta Psychologica*, 149, 32–39.
- Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: an empirically-based computational model. *Cognitive Science*, 31, 645–671.
- 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 Paris*, 102(1–3), 106–119.
- Söderman, T. (1993). Word associations of foreign language learners and native speakers—different response types and their relevance to lexical development. In B. Hammarberg (Ed.), *Problem, Process, Product in Language Learning: Papers from the Stockholm-Abo Conference, 21-22 October 1992* (pp. 157–169). Stockholm: Stockholm University.
- Sökmen, A. J. (1993). Word Association Results : A Window to the Lexicons of ESL Students. *JALT Journal*, 15(2), 135–150.
- Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, 32(2), 244–259.
- Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19(5), 317–330.
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, 38(4), 598–605.
- Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 550–569.
- Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks : Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29, 41–78.
- Stolz, S., & Tiffany, J. (1972). The production of “child-like” word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Behaviour*, 11(1), 38–46.
- Taikh, A., Hargreaves, I. S., Yap, M. J., & Pexman, P. M. (2015). Semantic classification of pictures and words. *Quarterly Journal of Experimental Psychology*, 68(8), 1502–1518.
- Tanaka-Matsumi, J., & Marsella, A. J. (1976). Cross-Cultural Variations in the Phenomenological Experience of Depression. *Journal of Cross-Cultural Psychology*, 7(4), 379–396.
- Taylor, J. R. (1998). Syntactic Constructions as Prototype Categories. In M. Tomasello (Ed.), *The New Psychology of Language, Volume 1* (pp. 177–202). New York: Psychology Press.
- Taylor, J. R. (2003). *Linguistic Categorization* (2nd ed.). Oxford: Oxford University Press.
- Taylor, J. R. (2008). Prototypes in Cognitive Linguistics. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 39–65). New York, NY: Taylor & Francis.
- Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.

- Theakston, A., Lieven, E., Pine, J., & Rowland, C. (2001). The role of performance limitations in the acquisition of verb argument structure. *Journal of Child Language, 28*, 127–152.
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College, Columbia University.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209–253.
- Tomasello, M. (2003). *Constructing a language: A usage-based approach to child language*. Cambridge, M.A.: Harvard University Press.
- Tomasello, M., Akhtar, N., Dodson, K., & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language, 24*(2), 373–387.
- van Hell, J. G., & de Groot, A. M. B. (1998). Disentangling Context Availability and Concreteness in Lexical Decision and Word Translation. *The Quarterly Journal of Experimental Psychology Section A, 51*(1), 41–63.
- 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.
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology (2006), 67*(January 2015), 1176–1190.
- Van Loon-Vervoorn, W. A. (1989). *Eigenschappen van Basiswoorden (Properties of Basic Words)*. Lisse, The Netherlands: Swets & Zeitlinger.
- Van Rensbergen, B., De Deyne, S., & Storms, G. (2016). Estimating affective word covariates using word association data. *Behavior Research Methods, 48*(October 2015), 1644–1652.
- Van Rensbergen, B., Storms, G., & De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic Bulletin & Review, (Advance online publication)*, 1–8.
- Vigliocco, G., Vinson, D. P., Druks, J., Barber, H. A., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and Biobehavioral Reviews, 35*(3), 407–426.
- Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology, 68*, 1–32.
- Vivas, L., Manoilloff, L., Garcia, A. M., Lizarralde, F., & Vivas, J. (2018). Core Semantic Links or Lexical Associations: Assessing the Nature of Responses in Word Association Tasks. *Journal of Psycholinguistic Research*.
- Warren, H. C. (1916). Mental association from Plato to Hume. *Psychological Review, 23*(3), 208–230.
- Warren, H. C. (1921). *A history of the association psychology*. New York: Charles Scribner's Sons.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45* VN-r(4), 1191–1207.
- Waxman, S. R., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H. (2013). Are Nouns Learned Before Verbs? Infants Provide Insight Into a Long-Standing Debate. *Child Development*

Perspectives, 7(3), 155–159.

- Wettler, M., & Rapp, R. (1993). Computation of word associations based on the co-occurrences of words in large corpora. *Proceedings of the First Conference on Very Large Corpora: Academic and Industrial Perspectives*.
- 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.
- Wheeler, K., & Bergen, B. K. (2010). Meaning in the Palm of Your Hand. *Empirical and Experimental Methods in Cognitive/Functional Research*, 1–14.
- Wilks, C. (2009). Tangled webs...: Complications in the exploration of L2 lexical networks. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical Processing in Second Language Learners: Studies in Honour of Paul Meara* (pp. 25–37). Bristol.
- 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.
- Williams, J. N. (1996). Is Automatic Priming Semantic? *European Journal of Cognitive Psychology*, 8(2), 113–162.
- Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, 49(5), 1105–1127.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell Publishing Ltd.
- 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.
- Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, 30, 315–329.
- Wolter, B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, 27(4), 741–747.
- Woodrow, H., & Lowell, F. (1916). Children's association frequency tables. *The Psychological Monographs*, 2(97).
- Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, 132(2), 173–189.
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60(4), 502–529.
- Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic Richness Effects in Lexical Decision : The Role of Feedback. *Memory & Cognition*, 43(8), 1148–1167.
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, 6(April), 72.
- Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification.

Psychonomic Bulletin & Review, 18(4), 742–750.

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.

Zareva, A. (2011). Effects of lexical class and word frequency on the L1 and L2 English-based lexical connections. *The Journal of Language Teaching and Learning*.

Zareva, A. (2012). Partial Word Knowledge: Frontier Words in the L2 Mental Lexicon. *IRAL: International Review of Applied Linguistics in Language Teaching*, 50(4), 277–301.

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.

Zwaan, R. A. (2003). *The Immersed Experiencer: Toward an embodied theory of language comprehension*. (B. H. Ross, Ed.), *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 44). New York: Academic Press.

Appendix 1 Test materials for replication of Fitzpatrick 2007, Chapter 2.

Note that while the words are the same as those used in the original study (and are presented in the same order), the format of the paper is not identical.

Task 1.

Please write down the first word you think of when you read each of the words listed. There are no right or wrong answers.

authority
context
distribution
export
income
labour
percent
required
significant
variables
aspects
complex
consumer
equation

assigned
cited
edition
federal
incentive
input
minimum
preceding
subsidiary
utility
chemical
converted
disposal
file

injury
normal
previous
relevant
select
transfer
consent
coordination
document
framework
instance
maximum
physical
removed
sufficient
volume
attitudes
concentration
dimensions

hierarchical
intervention
priority
simulation
thesis
voluntary
appreciation
commodity
currency
eventually
implicit
manipulation
practitioners
restore
thereby
widespread
behalf
incompatible
devoted

granted
imposed
mechanism
parallel
professional
series
summary
capacity
contact
enforcement
external
liberal
modified
orientation
ratio
sustainable
welfare

format
manual
mutual
protocol
rigid
suspended
vision
assembly
compiled
depression
forthcoming
intrinsic
likewise
odd
persistent
so-called
undergo

Task 2.

Please write down the first word you think of when you read each of the words listed. There are no right or wrong answers.

assessment
concept.....
create
definition
function
interpretation
involved
issues
legal
method
appropriate
categories
construction
distinction
final
impact

acknowledged
attached
domain
explicit
gender
inhibition
migration
scope
transformation
underlying
classical
confirmed
differentiation
equipment
global
innovation

positive
primary
purchase
survey
convention
deduction
ensure
immigration
initial
interaction
location
negative
published
validity
attributed
communication
emerged
goals
implications

phenomenon
publication
quotation
visible
appendix
automatically
clarity
contradiction
dramatic
exhibit
guidelines
intensity
predominantly
visual
analogous
assurance
confined
erosion.....
intermediate

label
occupational
principal
resolution
sum
amendment
consultation
energy
equivalent
fundamental
generated
notion
precise
psychology
version

minimal
portion
qualitative
relaxed
violation
adjacent
collapse
conceived
enormous
inclination
integrity
nonetheless
ongoing
reluctant
straightforward

Appendix 2: Cue word list for Nouns and Verbs study, Chapter 4

Please write the first word that comes to mind for each of the following words. There are no right or wrong answers and you shouldn't think too much about how you respond. Please don't change your responses after you have written them down. You don't need to give your name.

autopilot _____

horoscope _____

stifle _____

overshadow* _____

contaminate _____

outcast* _____

decode* _____

longitude _____

layman _____

twiddle _____

facet _____

shudder _____

detach _____

dissect _____

backfire _____

electrify _____

nightlife _____

deepen* _____

habitation* _____

assassinate* _____

migraine	_____
buffoon	_____
excursion	_____
confiscate	_____
lawsuit	_____
hypnosis	_____
devour	_____
reside	_____
circulate	_____
detonate	_____
infect	_____
repel	_____
priesthood	_____
discard*	_____
intrude	_____
grapple*	_____
recharge	_____
wallow	_____
scuttle*	_____
immerse	_____
forgery*	_____
expel	_____
crockery	_____
notify	_____
recapture*	_____
reconstruct*	_____

stardust _____

fallout* _____

infancy _____

overturn* _____

propel _____

redesign* _____

handiwork _____

smuggle _____

prescribe _____

vigil _____

commotion _____

symptom _____

loophole _____

witchcraft _____

reunite _____

borderline* _____

ancestry _____

entirety* _____

exterminate _____

rookie _____

migrate _____

voodoo _____

maniac _____

lordship _____

underworld _____

uphold* _____

medley _____

culprit _____

mimic* _____

hover _____

compose _____

conceal _____

piracy _____

sharpness* _____

cater _____

knighthood _____

depart _____

cosmos _____

humidity* _____

curfew _____

gourmet* _____

exclude _____

dissolve _____

biography _____

moron _____

overlap* _____

entity _____

reset* _____

sequel _____

getaway* _____

underdog _____

descend _____

geology

nutrition

Please circle one of the following:

1. I am a native speaker of English.
2. English is not my first language, but I don't have any problems using it on my daily life.
3. I have some difficulties using English.

*[Note: cues marked * were removed from analysis because they were found to break the morphological criteria for cue selection: see section 4.3.6]*

Appendix 3 Cue word list for Transitivity study, Chapter 5

Thank you for taking this test. Please follow the instructions below. You do not need to write your name.

All of the words which follow are **verbs**.

Please **write the first word that comes to mind** for each one. You can respond with any word (it doesn't have to be a verb); there are no right or wrong answers.

Please don't change your responses after you have written them down.

- irrigate _____
- thrive _____
- deduct _____
- blush _____
- migrate _____
- nourish _____
- purr _____
- glow _____
- abduct _____
- oscillate _____
- roar _____
- eject _____
- dawdle _____
- whack _____
- afford _____
- attach _____
- dwindle _____

limp

fall

sprinkle

extinguish

relieve

meditate

surge

grumble

triumph

flinch

hibernate

injure

bicker

exist

avoid

acquire

terrify

defy

swoop

vanquish

waddle

propel

inquire

expire

swoon

raise

illuminate _____
bark _____
remain _____
carry _____
appear _____
eliminate _____
seethe _____
hesitate _____
enhance _____
recede _____
devour _____
dilute _____
depart _____
lick _____
glisten _____
underline _____
rise _____
grovel _____
occur _____
emerge _____
mope _____
expedite _____
stitch _____
furnish _____
sparkle _____
dissect _____

snooze _____

chuck _____

flow _____

scold _____

rescue _____

baffle _____

shiver _____

pollinate _____

happen _____

subside _____

gratify _____

pacify _____

crush _____

cough _____

detect _____

tremble _____

strangle _____

stow _____

promote _____

bury _____

erupt _____

afford _____

fidget _____

trounce _____

knead _____

retrieve _____

inject

ache

kneel

wince

expel

Please circle one of the following:

1. I am a native speaker of English.
2. English is not my first language, but I don't have any problems using it on my daily life.
3. English is not my first language. I have some difficulties using English.

Appendix 4: Categorization of responses in Chapter 5.

Scheme of categorization for responses – specific to verb cues. Adapted from Fitzpatrick et al., 2015.

Basic category	Detailed category	Definition	Examples
Meaning-based	Synonym (1)	Cue and response are synonymous in some situations and represent a similar level of specificity.	<i>detach-separate, wallow-bathe</i>
	Troponym (2)	The cue is a way of doing the response (or vice-versa); one is more specific than the other; OR cue and response share a troponym at a higher or lower level of specificity.	<i>Walk-amble, eat-munch, fall-plummet</i> <i>Amble-stroll, crunch-chew</i>
	Contrast (3)	Cue and response are opposed in meaning; does not involve the use of an affix.	<i>Walk-run, sit-stand, Watch-ignore</i>
	Other conceptual – same part of speech (4)	Cue and response are related in meaning, but are not related in any of the above ways; same part of speech	<i>Afford – buy, nourish - eat</i>
	Other conceptual – different part of speech (5)	Cue and response are related in meaning, but are not related in any of the above ways: different grammatical class items which are unlikely to be position-based.	<i>Walk – legs, exterminate-daleks</i>
Position-based	Cue-response collocation (6)	Cue is followed by the response in common usage; includes compound nouns	<i>rookie-cop, migrate-overseas, voodoo-priest</i>
	Response-cue collocation (7)	Cue is preceded by the response in common usage; includes compound nouns	<i>medley-pop, culprit-caught, knighthood-bestow</i>
	Bi-directional (8)	Cue could precede or follow the response in a common phrase(s)	<i>circulate-paper, detonate-bombs, nightlife-good</i>
Form-based	Affix manipulation (9)	Cue is the response with the addition, deletion or changing of an affix	<i>propel-propeller, nutrition-malnourish, migrate-emigrate</i>

	Similar in form only (10)	Cue and response are similar in orthography and/or phonology but do not share meaning	<i>entity-entry, buffoon-baboon, wallow-willow</i>
	Two-step (11)	Cue and response appear linked only through another word	<i>layman-bed (via lay), hover-vacuum (via hoover), wallow-tree (via willow)</i>
Erratic	Erratic (12)	The link between cue and response seems illogical. Includes blank responses and repetition of the cue.	<i>moron-carrot, stifle-quite, facet-button</i>
Dual-category	Synonym <i>and</i> cue-response collocation (13)		
	Synonym <i>and</i> response-cue collocation (14)		
	Troponym <i>and</i> cue-response collocation (15)		
	Troponym <i>and</i> response-cue collocation (16)		
	Contrast <i>and</i> response-cue collocation (17)		
	Contrast <i>and</i> response-cue collocation (18)		
	Meaning and form-based (19)	Cue and response are similar in both form and meaning: this probably just means phonaesthemes; does not include affix manipulation. Defined as being meaningfully related in any of the ways 1-5, above, and also sharing a cluster of at least 2 letters in word-initial or word-final positions.	<i>Glisten-glimmer</i>