Second Language Word Association:
Processes, Methodologies, and Models

John Paul Racine

A thesis submitted to Cardiff University in fulfilment of
the degree of Doctor of Philosophy

2018

For Moka

May you dream big and always believe in yourself

Acknowledgements

I wish to express my deepest thanks to my supervisors, Professors Alison Wray and Tess Fitzpatrick, and Dr. Michelle Aldridge-Waddon. Without their wisdom and guidance, this dissertation would not have been possible. My research has also benefitted greatly from discussions with many other distinguished academics over the years. Among others, these include Irina Elgort, Todd Bailey, Ian Munby, George Higginbotham, Dale Brown, Brent Wolter, Henrik Gyllstad, John Read, and my friends and colleagues in VARG and the JALT Vocabulary SIG. All of you have helped to shape my approach to lexical research. Special thanks to the staff and students at Cardiff, Swansea, and Dokkyo Universities for your support, generosity, and warmth. I also wish to acknowledge four remarkable contributors to applied linguistics: Paul Meara, Norbert Schmitt, Paul Nation, and Batia Laufer. Their unparalleled contributions to our understanding of second language vocabulary have inspired me and will continue to resonate with generations of researchers to come. Finally, I wish to thank Yuko and Moka. Thank you always for your love and patience.

Abstract

This thesis examines processing in the bilingual mental lexicon by way of word association (WA) studies. The research reported here was designed to address three sets of key issues. Firstly, these studies attempt to establish whether or not WA methodology is a viable means of exploring the bilingual mental lexicon. Research questions to be addressed here concern the validity of current categorization schemes (i.e., whether they comprehensively account for all WA response data) and whether post-task interviews are necessary or useful in disambiguating responses. The second set of issues to be addressed here arise from the focus on cognitive processes in the lexicon. Research questions here are designed to address gaps in the literature concerning whether researchers are justified in conceptualizing subjects' responses as evidence of underlying cognitive styles. These questions will be addressed by implementing underused and never before utilized methods. Methods employed include restricted association tasks and an unconventional priming manipulation intended to alter response *types* (as opposed to altering *speed* of response, the conventional measure for inferring that priming has occurred). The final set of issues to be addressed here concerns the identification of the determinants of WA behaviour. Based on the results of the current studies, I will present a "dynamic" model of the WA process. The model attempts to account for the interplay among respondents' cognitive styles, features of the presented cues, and the influences of the experimental methodologies within which they meet.

In general, results of the studies showed that WA research is a viable means of investigating processes in the mental lexicon and that aspects of current WA methodologies (e.g. response categorization schemes) were up to the task. The findings also showed that, in some instances, subjects respond in accordance with what appears to be an underlying response preference, or cognitive style. In other cases, however, preferred response types can be altered by experimental manipulation. This suggested that WA responses are determined by interaction between types of cues, response preferences, and the methods employed. In the final chapter, a dynamic model of the word association process (DMWA) is introduced. Based on the conclusions described above, the DMWA describes and predicts WA behaviour in terms of the interaction of cue properties, characteristics of the respondents, and WA methodology.

Table of Contents

List of Tables and Figures

List of Abbreviations

The following abbreviations are used in this dissertation.

AWL	Academic Word List (Coxhead, 2000)
BIA	bilingual interactive activation model (Dijkstra & van Heuven, 1998)
BIA+	bilingual interactive activation plus model (Dijkstra & van Heuven, 2002)
BNC	British National Corpus
DMWA	the dynamic model of the word association process (presented here for the first time)
EAT	Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy, & Piper, 1973)
EFL	English as a foreign language
ERP	event related potential
ESL	English as a second language
EVST	Eurocentres Vocabulary Size Test (Meara & Jones, 1990)
IA	the interactive activation model (McClelland & Rumelhart, 1981)
L1	first language
L2	second language
LDT	lexical decision task
NNS	non-native speakers
NS	native speakers; native English-speaking respondents
RHM	Revised Hierarchical Model (Kroll & Stewart, 1994)
RT	reaction time
TOEIC	Test of English for International Communication
WA	word association
WRT	word recognition task

1. Introduction

1.1 Vocabulary research: From "neglected" to "vital"

> The fact is that while without grammar very little can be conveyed, without vocabulary *nothing* can be conveyed (Wilkins, 1972, p. 111).

Most people who have attempted to learn a second language will acknowledge the importance of vocabulary acquisition for achieving their learning goals. Language teachers and learners generally agree that aspects of lexical acquisition – such as spelling, pronunciation, form-meaning pairings, collocation, etc. – are essential for the successful acquisition and use of a second language. While the importance of vocabulary may be almost axiomatic for stakeholders in language education, the same has not always held true among language researchers. A century of applied linguistics studies has only recently begun to yield a substantial body of second language vocabulary research. Indeed, as of 1980 so little evidence of lexical research even existed that Meara dubbed vocabulary acquisition "a neglected aspect of language learning" (Meara, 1980, p. 221).

But times have changed and the transformation has been striking. Thirty years later Meara and his colleagues exclaim that "vocabulary is a vital and active area of language learning research" (Milton et al., 2008, p. 135). In 2013, Nation stated that "over 30 per cent of the research on vocabulary that has appeared in the last 110 years was published in the last 11 years" (Nation, 2013, p. 5). Meara also expressed this growth quantitatively, claiming that researchers were now producing approximately 120 vocabulary-related journal articles per year,

representing “a twenty-fold increase in outputs” (2012, p. 8) from the 1970s to 2012. In his modern history of applied linguistics, de Bot (2015) summarized this observation saying that vocabulary research has “grown exponentially” (p. 84). It would seem, then, that linguists now recognize what language learners had always known: vocabulary is an important field of research and worthy of rigorous investigation.

Among the many reasons for researchers’ current interest in vocabulary is the recent popularity of the view that lexis may be the central principle by which language is organized in the mind. Since its earliest conceptions as a “word-store” (Oldfield & Wingfield, 1965, p. 273), the *mental lexicon* – the notional repository of vocabulary knowledge in the human mind – has remained a key feature in psycholinguistic models of language. Levelt’s (1989) highly influential model of speech production, for example, specifies the lexicon as the driving force behind spoken language. According to this model, grammatical choices, morphology, and phonology are all predicated on lexical selection and the accompanying aspects of word knowledge within the lexicon. This “lexical hypothesis” (Levelt, 1989, p. 181) identifies the mental lexicon as essential to language abilities. In turn, we may wish to posit that the measurement of the development of the bilingual mental lexicon – i.e., of second language lexical acquisition – is key to investigating language development in general.

1.2 Investigating the mental lexicon via word association methodologies

This thesis aims to make an original contribution to research into second language vocabulary acquisition by exploring a series of questions about the developing structure of the mental lexicon, as revealed through word association (WA) tasks. Of all the methods researchers may choose to apply to the investigation of the mental lexicon, WA in particular may prove to be very fruitful due to its simplicity as a data collection method and its ease of administration for both researchers and subjects.

Simply put, data collection in WA research consists firstly of the presentation of cues (typically single words). Subjects are required to respond with the first thing that comes to mind, and these responses are recorded. Stimuli may be presented aloud, or in orthographic form on paper or via computer screen. Responses may be spoken or handwritten/typed and typically consist of a single word. The benefits of such a simple method of data collection are twofold. First of all, WA methods are easy and inexpensive to administer. Researchers require no specialized equipment to collect large amounts of data in a relatively short time. For subjects too, WA is both painless and unobtrusive. Indeed, many subjects enjoy the WA research experience, feeling as if they are playing a kind of word game.

As an aid to understanding WA research methods, Figure 1.1 depicts the WA research process as a six-stage model. While the model is designed to illustrate the WA research process, at the same time it is representative of much experimental research in a variety of fields. WA researchers – informed by their theoretical outlook concerning the nature of language, how it is acquired, the content and

organization of the mental lexicon, etc. – begin by forming hypotheses or research questions. These hypotheses inform their decisions concerning all aspects of research design and methodology in Stage 2. Stages 3 through 5 occur in sequence thereafter (indicated by the arrows between them in the figure). The specifics of these steps are determined by the methodological choices made prior (as indicated by the downward arrows). Finally, the results of the research are interpreted in Stage 6. Conclusions then feed back into researchers' beliefs, informing their theories about language in general and about WA specifically, potentially precipitating more hypotheses for further study.

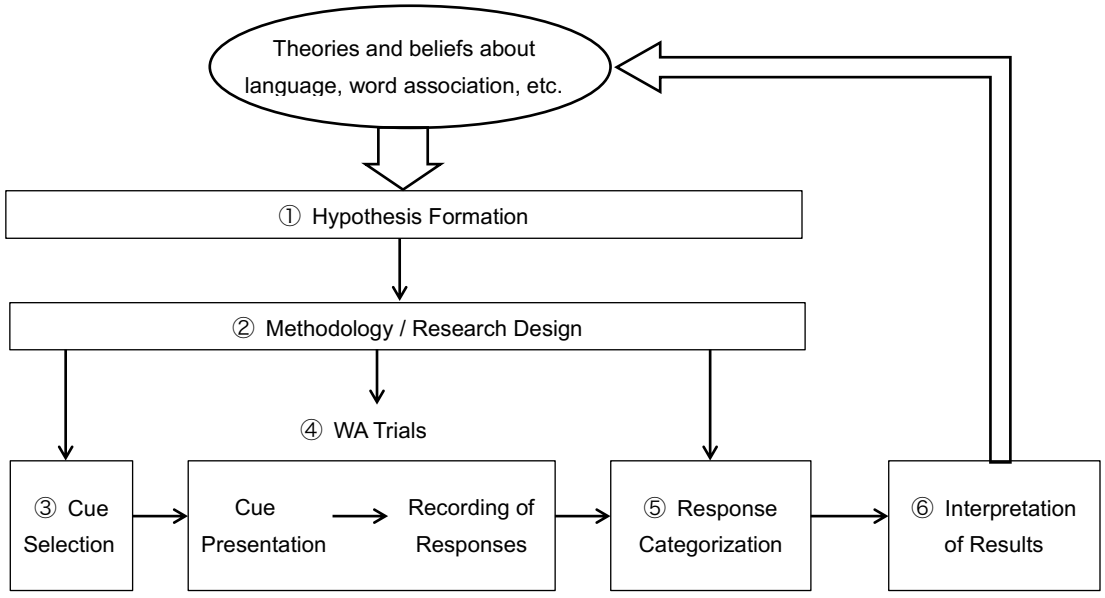


Figure 1.1. A six-stage model of the word association research process.

With this model as a guide, it is possible to position aspects of the research to be reviewed in the next chapter and the research to be presented for the first time in subsequent chapters within the WA research framework. Thus, in examining the variety of L2 research questions to which WA methods may be applied (Section

2.3), we may illuminate the early stages of the research process. These involve the beliefs held by the researchers about what can and cannot be investigated via WA methodology, and Stage 1 of the process in which decisions are made as to precisely which research questions they are to pursue. Section 2.4 – concerning the adoption and modification of WA methods – highlights the research process in Stages 2 through 5 with a specific focus on Stage 4. Critical scrutiny paid to the research designs in these studies will help to illuminate caveats that I will then attempt to address in my own research.

Stage 6, in which research results are interpreted and conclusions are drawn, is particularly significant. In my discussion of prior research, I will be very critical of the findings reported, possibly to an extent to which readers will wonder why I had selected these particular studies to review at all. Therefore, I should state now that the representative studies I have chosen to analyse in the next chapter and elsewhere are included not so much for their results, but for the methodologies employed. As I have stated above, one of the primary purposes of this chapter are to demonstrate that WA research – when certain caveats are heeded – is both a viable and a useful pursuit for language researchers.

Indeed, WA methodology has a well-established history both within linguistics and beyond. With its roots in 17th Century empiricist philosophy, some of the first attempts to systematically record data using WA were conducted by Galton (1879, 1883). His interest was not linguistic, but rather to uncover the "mental anatomy" (1883, p. 145) at the level of thoughts or concepts. In the early 20th Century, WA was utilized as a diagnostic tool in psychiatric and psychoanalytic studies (e.g.,

Jung, 1918/1969; Kent & Rosanoff, 1910) where the degree of typicality of response was held to be an indicator of mental health or illness. While these early studies may not have revealed compelling results in and of themselves – indeed, the use of WA as a diagnostic tool has been characterized as "disappointing" (Dunn, Bliss, & Siipola, 1958, p. 61) – they did precipitate a diagnostic research tradition that continues to this day (e.g., Vezzoli et al., 2007). Typicality of WA response data in comparison with normative responses is still utilized in a variety of psychological and neurological studies. WA methods have been applied to investigations of a variety of personality traits (e.g., Dunn et al., 1958; Merten & Fischer, 1999), mental illnesses (Kent & Rosanoff, 1910; Manschreck, Merrill, Jabbar, Chun, & DeLisi, 2012; Merten, 1993), and the study of aging and dementia, particularly Alzheimer's Disease (Gewirth, Shindler, & Hier, 1984; Gollan, Salmon, & Paxton, 2006; Goral, Spiro, Albert, Obler, & Connor, 2007; Hirsch & Tree, 2001) .

In linguistics, the WA research tradition is just as strong. One strand of WA research concerns first language (L1) development. This research focuses on tracking change as a function of maturation, with responses changing from primarily syntagmatic (i.e., responses belonging to word classes other than that of their cues) to primarily paradigmatic (i.e., belonging to the same word class). Evidence of this "syntagmatic-paradigmatic shift" (e.g., Entwisle, 1966a; Entwisle, Forsyth, & Muuss, 1964; S. M. Ervin, 1961; Francis, 1972; K. Nelson, 1977; Söderman, 1993; Woodrow & Lowell, 1916) is observed in the responses of young children and is explained in terms of a restructuring of lexical connections away from ones based primarily on frequency of co-exposure (i.e., "based on exposure to

spoken discourse"; Entwisle et al., 1964, p. 25) to ones based on deeper semantically-based connections.

Another strand of linguistic research in which WA methods are utilized mirrors earlier psychiatric applications of WA. These types of studies involve the comparison of individuals' responses to sets of normative data as a means of measuring deviation/stereotypy. Rather than being used as diagnostic tools, when WA norms are applied to linguistic research they are often employed as a measure by which language proficiency can be assessed. In other words, studies in this strand of research examine the response norms of native speakers in comparison with the types of responses made by L2 learners. This research is based on the assumption that second language (L2) learners will exhibit types of associations that become increasingly similar to those of native speakers as their L2 language proficiency improves. In fact, this assumption has been called into question in recent years by Fitzpatrick and her colleagues (Fitzpatrick, 2007, 2009; Fitzpatrick, Playfoot, Wray, & Wright, 2015) and will be scrutinized thoroughly in this dissertation as well.

Finally, there is one more strand of linguistic enquiry involving WA that is particularly germane to the studies reported in this dissertation. This research strand centres around the classification of response types and the distribution of subjects' responses across categories. While the classification of response types remains a key aspect of many WA studies to date, the use of particularly detailed categorization schemes and the conceptualization of these distributions as individual respondent "profiles" is relatively new (e.g., Fitzpatrick, 2007, 2009;

Fitzpatrick & Izura, 2011; Higginbotham, 2010). The assumptions underlying these new categorization schemes deserve as much scrutiny as the assumptions that have traditionally underpinned WA research. One of the purposes of this dissertation is to analyse these assumptions.

1.3 Scrutinizing and innovating WA methodology

As the examples above indicate, word association's status as a research method for investigating the mind in general, and the mental lexicon in particular, is long-established in both linguistics and other fields. The practicalities of its affordability and ease of use continue to promote WA as an appealing method for researchers interested in exploring the bilingual mental lexicon. Yet WA methods remain underutilized and the methodology not fully explored. One reason for this is that the results of L2 WA studies have proven inconsistent and that, at least in some cases, the research has been poorly conducted. Meara's (2009, Chapter 8) critical reviews of over 100 of the most influential L2 WA studies reveal a variety of contradictory findings. Likewise, Fitzpatrick et al. (2015) critically examined the most widely cited WA studies both from within linguistics and from other fields. This analysis revealed precisely how wide-ranging researchers' criteria can be in the selection of cues, adoption of norms, and the treatment of data. Indeed, their analysis reveals that many studies have not included adequate explanations of the manner in which data has been treated at all. The breadth of methodological and theoretical approaches made evident in this study has undoubtedly been a contributing factor to the inconsistent results noted in prior L2 WA research. Two more studies by these same researchers also took a critical approach to L2 WA

research and deserve mentioning here. Meara's (1982) summary paper may be regarded as the first attempt to outline methodological best practices for WA researchers, and Fitzpatrick (2007) evaluated many assumptions underlying WA research that had not been given careful consideration in prior research.

Partly as a result of this scrutiny paid to the methodologies of prior studies, a number of innovations have appeared in the L2 WA research published in the last decade. These innovations include new ways of classifying responses (e.g., Albrechtsen, Haastrup, & Henriksen, 2008; Fitzpatrick, 2006), new means of conceptualizing data (Fitzpatrick, 2009; Higginbotham, 2010), the application of WA research to an increasingly wider variety of first and second languages (e.g., Bøyum, 2016; De Deyne & Storms, 2008; Du & Gao, 2013; Higginbotham, Munby, & Racine, 2015; Li, 2011; Namei, 2004), the combination of response categories to identify "dual-link" associations, and the confluence of WA with methods traditionally considered to be within the purview of psycholinguistics (e.g., the recording of reaction times; Fitzpatrick & Izura, 2011).

This dissertation follows in the footsteps of these prior studies, critically assessing prior research while attempting to move the field forward through methodological innovation. To do so, reviews of representative papers in L2 WA will be presented in Chapter 2. Methodological issues raised in these reviews will receive further scrutiny in replications of recent innovative studies presented in Chapters 3 and 4. The original studies and discussion presented in subsequent chapters, will attempt to build on these findings and new approaches will be presented.

1.4 Focus on categorization

One of the recent methodological innovations came in 2006 when Fitzpatrick introduced an alternative classification of WA responses. While non-traditional categorization schemes have been introduced by a number of researchers over the years (e.g., Aitchison, 2012, Chapter 9; Albrechtsen et al., 2008, Chapter 2; Sökmen, 1993), Fitzpatrick's (2006) scheme differed from traditional categorizations in a number of regards. Until that time, most WA researchers (e.g., Entwisle et al., 1964; Fillenbaum & Jones, 1965; Gewirth et al., 1984; Greidanus & Nienhuis, 2001; Politzer, 1978; Wolter, 2001) had adopted classifications based on four broad categories: paradigmatic responses (belonging to the same word class as the stimulus), syntagmatic responses (having intra-sentential, syntactic relationships with the stimulus), clang responses (sharing phonological or orthographical characteristics with the stimulus), and an *other* category (for responses with no apparent link to their stimuli). Fitzpatrick's categorization was also based on four broad categories: meaning-based, position-based, form-based, and erratic responses (see Table 1.1). While these categories (to be explained in detail in Chapter 3) may yet prove almost indistinguishable from the traditional categories in actual practice, Fitzpatrick's new, user-friendly labels defied a century-old tradition: linguists had been using the *paradigmatic* and *syntagmatic* labels since at least the early 20th Century.[1]

[1] The introduction of these two terms to the field of linguistics is often attributed to Saussure (e.g., 1916/2006). While they appear in many English translations of his semiotic theory, some argue that Saussure himself never used the terms in the manner in which they tend to be portrayed. Indeed, Harris (2001, p. 90) states that "'paradigmatic' is not a Saussurean concept at all".

Perhaps more importantly, Fitzpatrick's categorization scheme also introduced a taxonomy of 17 subcategories (see the central column of Table 1.1). Based in part on the aspects of word knowledge outlined by Nation (2001), these subcategories were intended to create a sharper instrument with which to measure WA response behaviour. Indeed, an analysis of data categorized in accordance with this scheme yielded promising results in that study (2006) and in subsequent research (e.g., Fitzpatrick, 2007, 2009). Precisely what promise these subcategories may hold for researchers is an issue I will return to in Section 1.5 and in subsequent chapters.

Table 1.1. Fitzpatrick's (2006, p. 131) category/subcategory classification scheme for word association responses (*x* = stimulus word, *y* = response word).

Category	Subcategory	Definition
Meaning-based associations	Defining synonym	x has the same meaning as y
	Specific synonym	x can mean y in some specific contexts
	Hierarchical/lexical set relationship	x and y are in the same lexical set, are coordinates, or have a meronymous or super/subordinate relationship
	Quality association	y is a quality of x or x is a quality of y
	Context association	y provides a conceptual context for x
	Conceptual association	x and y have some other conceptual link
Position-based associations	Consecutive xy collocation or compound	y follows x directly (or with an article between them)
	Consecutive yx collocation or compound	x follows y directly (or with an article between them)
	Phrasal xy collocation	y follows x with a word (non-article) or words between them
	Phrasal yx collocation	x follows y with a word (non-article) or words between them
	Different word class collocation	y collocates with x + affix
Form-based	Derivational affix difference	y is x +/- a derivational affix

associations	Inflectional affix difference	y is x +/- an inflectional affix
	Similar form only	y looks or sounds similar to x, with no clear meaning link
	Similar form association	y is an associate of a word with a similar form to x
Erratic associations	False cognate	y is related to a false cognate of x in the L1
	No link	y has no decipherable link to x

The utility of Fitzpatrick's scheme for future research lies in its ability to provide a more nuanced means of analyzing WA responses through the subcategories contingent on the degree to which it reveals new insights into the processes underlying associations. While there is some evidence that the scheme is beginning to be adopted by other researchers (e.g., Higginbotham, 2010), its utility has yet to be fully determined. Beyond what it may reveal in terms of response patterns and processes, there are three factors in particular by which it should be assessed. First of all, a WA categorization scheme must be judged on its ability to comprehensively account for all types of responses. Indeed, we may consider *comprehensiveness* to be the defining feature of any successful categorization scheme. If we are to sort roses by colour and we do not have a pile for the red ones, there may be a fundamental flaw in our system. *Exclusivity* between categories is another, related, consideration. If our scheme includes a *red* category and a *white* one, where would we put a pink rose? We might consider including a *pink* category, or we may consider the shade to be close enough to red or to white to fit in one of our pre-existing categories.

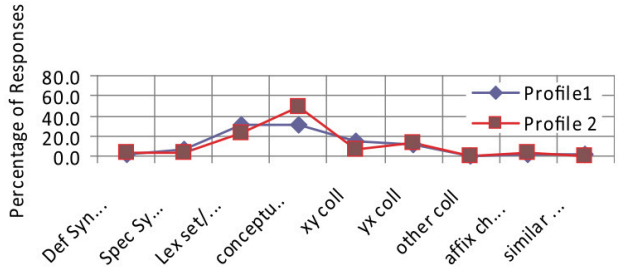
Another feature by which to assess the value of a taxonomy of WA responses is its *psychological validity*. This refers to the ability of a scheme to identify the association processes that underlie the responses. Thus, if a suggested category accurately identifies and labels a distinct type of response, it will have utility in identifying a distinct underlying process or processes. For example, suppose a WA researcher decided to employ a *furniture* category for all responses that refer to types of furniture. Insofar as the researcher is committed to psychological validity, doing so would imply that the series of processes mediating these responses is somehow different from those mediating other types of response. In other words, unless we believe that the lexical representations of furniture are stored and accessed differently from other items in the lexicon, then we should consider the category to be psychologically invalid[3] and we should not include it within the classification scheme. Importantly, if the categories do in fact accurately identify and label underlying processes, then any experimental manipulation intended to facilitate or inhibit these processes should be observable in variations in the numbers and types of responses observed. This is a crucial assumption in the research to be presented throughout this dissertation. Therefore, two of the goals of the next two chapters are 1) to examine how comprehensively and exclusively Fitzpatrick's (2006) categories account for the types of responses elicited in L2 WA studies, and 2) to assess the psychological validity of these categories.

[3] Note that I am in no way implying that the positing of the existence of a potentially meaningful *furniture* category – or any other category, for that matter – in the cognition of research participants is invalid *as such*. I am using psychological validity specifically as a criterion by which to assess the relations between categories of associations and the processes presumed to underlie them.

1.5 Beyond categorization: Response profiles and cognitive styles

In quick succession following the study in which the two-tiered categorization scheme described above was first employed (2006), Fitzpatrick published two WA studies in which she described responses in terms of subject *profiles* (2007, 2009). By employing the 17 subcategories to categorize responses, the data highlights – particularly when presented in graphic form – a detailed "profile" of a subject's response preferences. Such profiles were presented by Higginbotham (2010) and are duplicated in Figure 1.2. An explanation of the categories utilized here will appear in Chapter 3. For now, however, it is only important to note the broad differences between Graphs A and B. Graph A represents the responses of a single subject. Profiles 1 and 2 represent the subject's distribution of responses collected at two different times, elicited by two different sets of cues. At both times, she tended to produce an abundance of responses with conceptual links to their stimuli. Graph B illustrates the profiles of a different respondent – again, elicited at two different times by two different sets of cues. There are clear between-subjects differences (i.e., between Graphs A and B). This contrast illustrates what appear to be differing *cognitive styles* – referred to as *response sets* in some WA studies (e.g., Cramer, 1968, Chapter 7; Moran, 1966) – between the two subjects. The similarities observed between Profiles 1 and 2 within each respective graph show that these response preferences have remained stable over time, despite the use of different sets of cue words.

Graph A



Graph B

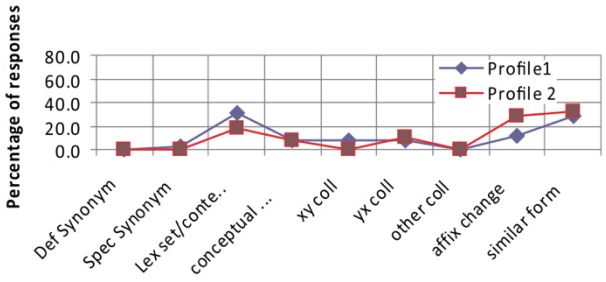


Figure 1.2. Examples of individual subject response profiles (from Higginbotham, 2010, p. 385, Figures 6 and 7).[4]

Two aspects of this research contribute to the reconceptualization of response distributions across categories as a function of respondents' cognitive styles. Firstly, the methodological change – to a more detailed classification of responses – has rendered profiles that highlight the details of response styles in a more nuanced manner than could conventional four-part categorizations. Secondly, the empirical finding that one's responses remain relatively stable across WA tasks (Fitzpatrick, 2007, 2009; Higginbotham, 2010) support the claim that association responses reflect or represent subjects' cognitive styles. A cognitive style is thus a trait-like tendency to respond with an abundance of certain types of associations. In fact, there is evidence (e.g., Fitzpatrick, 2009) that response preferences remain stable *even across a participant's languages* (if a certain degree of L2 proficiency

[4] It should be noted that Higginbotham's graphs are somewhat misleading. His use of line graphs enhances the "shape" of the profiles and aids in highlighting the contrasting styles between subjects. However, they also imply that there are midpoints between discrete categories that simply do not exist.

has been attained). These important findings provide support for the usefulness of reconceptualising individual respondents' associations as reflections of their cognitive styles.

With only a few exceptions (mostly in L1 psychological studies; e.g., Arthur & Freemantle, 1966; Carroll, Kjeldergaard, & Carton, 1962; Jenkins, 1960), however, WA researchers have not conceptualized their data in terms of cognitive styles. Fitzpatrick's (2007, 2009) *profiles*, for example, are "numerical" (personal communication), rather than conceptual. In her studies, a "profile" refers to a set of response data, typically collected from one person at one time. Higginbotham (2010), as seen in his use of the *profile* labels in Figure 1.2, uses *profile* in the same manner as Fitzpatrick (i.e., as the numerical distribution of a set of responses across a detailed categorization scheme). Both graphs highlight the similarities between individual's profiles as measured at different times with different types of cues. This parallels Fitzpatrick's findings (to be described in detail in subsequent chapters) that profile data appear to remain stable over time (2007) and even across languages (2009). Despite these results, neither researcher cites cognitive styles (or response sets) as the underlying reason for the similarities.

Indeed, throughout most of the WA literature, no assumption is made that an individual subject's response preferences might be a determining factor in which responses are elicited. Researchers' underlying assumption appears to be that, while association data are in fact representations of a respondent's productions, the types of responses elicited are primarily determined by specific cues, presented in a specific language, under a certain set of conditions. I would like to

suggest that interpreting data in terms of factors exclusively external to the research subject has led to a missed opportunity. One can imagine a large number of research hypotheses being generated if each set of data (i.e., each *profile*) were interpreted as an approximation – or as a single representative sampling – of a subject's proclivity to respond in a certain manner. While Fitzpatrick and Higginbotham do not take this step from response profiles to positing the existence of underlying cognitive styles, they have laid important groundwork for the studies described in this dissertation. Among the questions raised by the results of these studies, are the following: If cognitive style is an underlying determinant of an individual's responses, what conditions yield association data consistent with one's style? Conversely, what kind of experimental manipulation (e.g., to cue words or participant instructions) elicits a distribution of response types not in accordance with subjects' presumed cognitive preferences? The extent to which one's response preferences remain stable despite experimental manipulation is an empirical issue that will be addressed in Chapters 5 and 6.

1.6 Validation through replication

In order to approach the first goal of this thesis – to test the validity of these new response categories and the reconceptualization of responses as reflections of subjects' cognitive styles – two systematic replication studies will be presented. The first one is of Fitzpatrick's (2006) study in which she introduced the two-tiered categorization scheme duplicated here in Table 1.1. The replication has two purposes. These purposes are distinct, but together, they may be considered two aspects of a test of the *conceptual validity* of the categorization scheme. The first

purpose is to address whether or not the scheme comprehensively and exclusively categorizes all types of responses likely to be observed in WA studies, specifically in L2 WA studies. This may be considered a test of the scheme's *surface validity*. That is, if the scheme *appears* to take into account all possible association types, then we may consider it to be valid, at least on its surface. The second aim of the replication study is to determine the *psychological validity* of the categories by assessing whether or not they successfully identify and label distinct cognitive processes. These processes are presumed to mediate between the access of lexical representations (as instigated by the perception of WA cues) and the activation of responses within the mental lexicons of WA research subjects. This assessment of psychological validity is necessary if we are to make claims about the kinds of links assumed to exist in the lexicon based on types of response.

The second replication attempted to reproduce Fitzpatrick's (2009) L2 study in which subject responses were characterized as profiles. The aims of this replication are also twofold. One is to confirm or refute the apparent stability of response preferences across languages. In other words, the study is designed as a means of testing the relative strength of subjects' cognitive styles by examining responses in a first and second language. The original study compared the responses of native English speakers learning Welsh in Wales. The focus of the replication will be responses from Japanese learners of English in Japan. For reasons to be outlined in Chapter 4, we may consider learners of Welsh as a second language as an unusual population from which to draw subjects. Learners of English as a foreign language (EFL) in Japan are a more standard research population. Data obtained from these Japanese non-native speakers of English

(NNS) may thus be generalizable to a broader range of language learner communities. The second purpose of this replication is to re-examine one of the important findings of the original study. Fitzpatrick (2009) found that learners' L2 profiles more closely matched their own L1 profiles as L2 proficiency increased. The implication is that WA norms gathered from native-speakers as a potential yardstick by which to measure L2 proficiency are actually unsuitable for the task. In other words, Fitzpatrick's finding contradicts the traditional assumption that L2 proficiency facilitates nativelike association behaviour. Rather, these results suggest that investigations into L2 proficiency should really begin with the examination of individual subjects' own L1 associations.

The relationship between the two replication studies and the original studies upon which they are based can be seen in Figure 1.3. These replications will appear in Chapters 3 and 4.

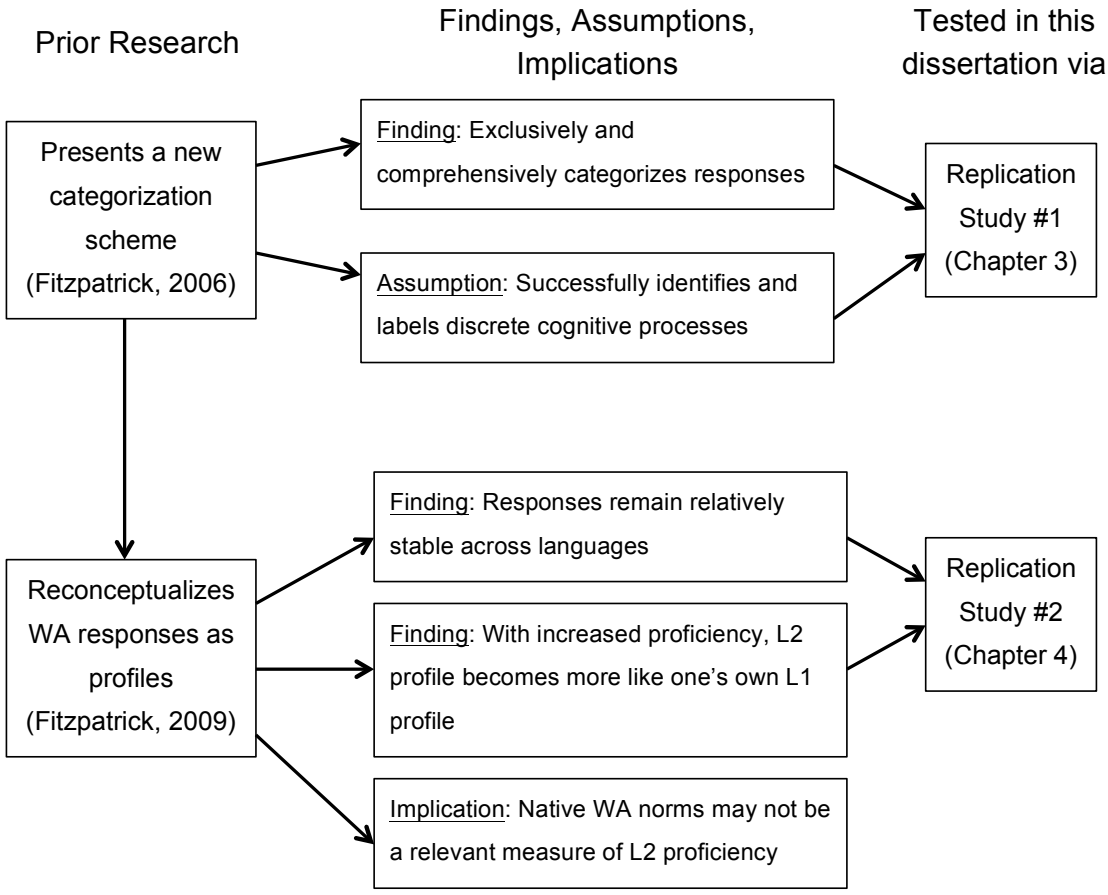


Figure 1.3. Two replication studies and their relation to prior research.

1.7 Revealing the processes: Adopting and adapting priming methodologies

Replicating the findings of these studies would add weight to the claim that Fitzpatrick's (2006) categorization scheme comprehensively categorizes response types. If that is the case, then the categorization can be further employed to assess the findings of Fitzpatrick's (2007) study that showed that 1) individuals' response types remain relatively stable across their first and second languages and 2) that one's L2 responses become more like one's L1 responses as greater L2 proficiency is attained. At the same time, these replication studies will include critical analyses of the two-tiered response categorization scheme and of the reconceptualization of response profiles as evidence of subjects' cognitive styles. In turn, the replication

studies will lay the groundwork for three original studies I will present in Chapters 5 and 6. In this research, alongside typical WA methodologies, I will adopt adapted versions of techniques used routinely in the fields of cognitive psychology and psycholinguistics. In essence, these are priming methodologies. However, for reasons that will become apparent in subsequent chapters, I will introduce modified versions of these common psycholinguistic research techniques. The application of these borrowed, adapted psycholinguistic methodologies to the L2 WA studies presented here will grant us greater confidence in any conclusions we may draw from the data concerning the underlying processes of WA.

Priming effects are robust phenomena that have been observed in numerous studies in experimental psychology. In the context of linguistics research, priming refers to "the phenomenon in which prior exposure to language somehow influences subsequent language processing" (McDonough & Trofimovich, 2009, p. 1). For the purposes of the current studies specifically, "subsequent language processing" is that which underlies the production of WA responses. Precisely to what extent this processing is affected will be reflected in changes in the distribution of types of responses. "Prior exposure" will take two forms here. In Chapter 5, I will introduce a methodology by which cue order is manipulated as a means of inducing priming effects. In the two studies presented in Chapter 6, I will adopt a restricted association task for the same purpose.

The details of these methodologies will be explained subsequently, but for now it is important to note that the type of priming measurement to be utilized here is not the same as priming techniques typically adopted throughout the psycholinguistic

literature. Data collection in priming studies from within both psychology and linguistics typically takes the form of reaction time (RT) measurement. In other words, priming manipulations may have the effect of facilitating or interfering with subsequent processing, as observed in shorter or longer RTs respectively. In the current studies, however, the effects of the priming manipulations will be measured by examining their impact upon response profiles. In other words, effects of the experimental measures should be observable in changes in the distribution of responses across the categorization scheme. It is the potential influence of these manipulations on response types – rather than response times – that makes the priming methodology a useful – indeed, a necessary – component of the current investigations.

1.8 Modelling the determinants of WA data

All told, then, the research presented in this dissertation consists of five WA studies. Two of these are replication studies designed to verify the findings of recent studies (Chapters 3 and 4). The studies upon which they are based pose challenges to traditional WA methodology that demand the kind of scrutiny that these replications are intended to provide. The three subsequent studies (in Chapters 5 and 6) will attempt to build upon the findings of the replications, providing further evidence by which to evaluate the implications of those prior studies. These three original studies will implement unique, modified versions of the kind of priming procedures more typically employed in psychology research. Situating these new studies alongside the replications, Figure 1.4 (as in Figure 1.3)

shows the relationship between the challenging results and assumptions of the prior studies discussed above and the five studies to be presented here.

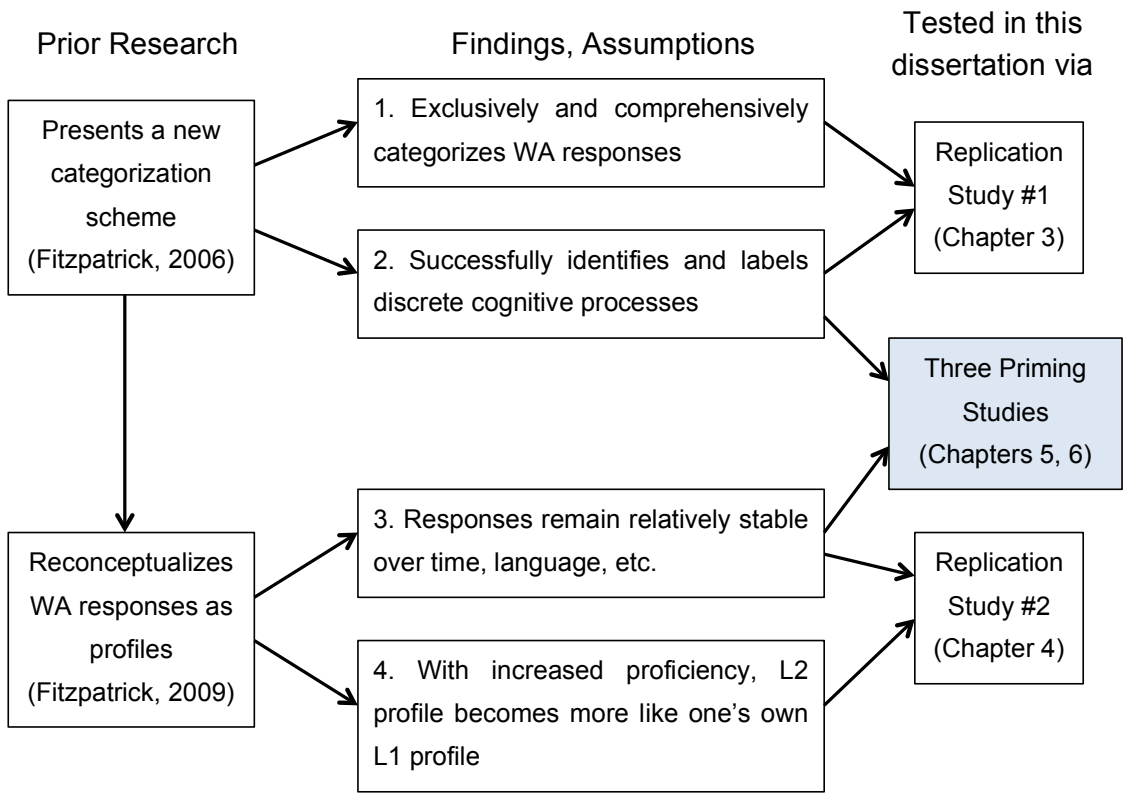


Figure 1.4. The current studies and their relation to prior research.

In Chapter 7, a thorough discussion of the research findings will precipitate the final element of this investigation. Namely, the presentation of a new model of the determinants of WA behaviour.

1.9 Organization of this dissertation

The next chapter will be devoted to establishing the extent to which previous research shows the concepts and methods of WA described above to be robust and defensible. Specifically, as my interest lies in the organization of vocabulary in the bilingual lexicon, I will be examining L2 WA studies. Literature germane to a

critical examination of these concepts and methods will be introduced and reviewed. Three broad strands of second language applied linguistics research will be covered, each exemplified by a small number of papers to be scrutinized in detail. The first of these strands examines ways in which researchers have modelled lexical processes and the mental lexicon. These models provide the theoretical framework upon which this entire dissertation is based. The second and third research strands provide methodological background for the current studies. Specifically, the second strand of research to be examined consists of a selection of papers exemplifying the breadth and variety of approaches to which L2 WA methods may be applied. Finally, the third strand consists of studies through which these methods and conceptual issues raised above (e.g., response categorization and response profiles) will be examined in detail.

Chapters 3 through 6 will consist of five empirical studies. Chapters 3 and 4 are replications of Fitzpatrick's 2006 and 2009 studies in which she first introduces the two-tiered classification of WA responses described above, and applies response profiles to L2 associations, respectively. Chapter 5 describes an original study in which cue order is manipulated as a means of eliciting priming effects in subsequent responses and Chapter 6 presents two more studies in which restricted association tasks are used as priming techniques.

Finally, the discussion in Chapter 7 will re-examine the concepts and issues introduced above in light of the results of the studies presented. These findings will be utilized to achieve the final goal of this dissertation, to present a model of the determinants of WA.

Chapter 2. Second Language Word Association and the Mental Lexicon: Applications and Methods

2.1 A framework for second language WA research

The series of original second language WA studies I will present in subsequent chapters are founded on two simple premises that I will attempt to establish in this chapter. The first is that WA is a viable method – that is, a robust and effective one – for gaining insight into second language lexical organization and processes. To establish this premise, I will present in Section 2.3 a number of prior studies that demonstrate the versatility and relevance of WA to L2 research in general, and to L2 lexical research in particular.

The second premise is that WA methods are adaptable in a variety of ways, and that this adaptability makes WA methodology particularly valuable for lexical research. As I will demonstrate in Section 2.4, the simple cue-response format can be modified to elicit multiple responses (e.g., Fitzpatrick & Munby, 2014; Kruse, Pankhurst, & Sharwood Smith, 1987), yielding a richer view of the lexical network around individual words/nodes. For the same purpose, specific types of associations (e.g. synonyms or antonyms) can be elicited via instruction, rather than elicited freely (e.g., Ramsey, 1981; Riegel, Ramsey, & Riegel, 1967; Riegel & Zivian, 1972). Also, response times (RTs) between the onset of cues and the elicitation of associations can be recorded (e.g., Fitzpatrick & Izura, 2011) and RTs between multiple responses can be measured in continuous word association tasks (e.g., Aldridge, Fontaine, Bowen, & Smith, 2018). These methods yield

another type of data from which researchers may infer processes in the lexicon, their speed, potential pathways of activation, and strength of links within the mental lexicon. Moreover, when adopted in conjunction with other research methods (see Albrechtsen et al., 2008 for a variety of examples), WA methodology can provide a unique perspective into second language lexical processes. The discussion here will investigate key components of the methodology and research design that are then reapplied in the original studies to follow in subsequent chapters. At the same time, it will reiterate the diversity of linguistic studies to which WA methods might be applied.

Before examining the relevance of prior WA studies to applied linguistics research, the many ways in which WA methods might be adapted, and the diversity of research to which WA methods might be employed, it is necessary to establish the theoretical background upon which this research is based. For this reason, I will begin (in Section 2.2) with a discussion of a variety of models that researchers have proposed to describe the processes and structure of the mental lexicon. While this does indeed provide the "theoretical" background upon which we can examine WA as a means of investigating the L2 mental lexicon, it is important to note that extant theories and models of the mental lexicon have developed on the back of more than a half-century of empirical research. That is, while these models are necessarily hypothetical (or *theoretical*) notions of a structure within the minds of language learners, the approach taken in prior studies – and continuing in the original research to be reported in subsequent chapters – is an *empirical* one. In other words, while the lexicon itself is the *notional* repository of our lexical knowledge, language behaviour as observed in linguistics research – particularly in

the WA studies and other lexical studies to be reported here – is an observable phenomenon. It is from the observations made in prior research, and in the current studies, that we infer the processes and structures of the mental lexicon.[1]

2.2 Modelling bilingual processes and the mental lexicon

In a sociolinguistic publication entitled *Languages in Contact*, Weinreich (1953/1968) was one of the first language researchers to attempt to classify types of bilingualism. The following year, Susan M. Ervin and Osgood (1954) published a psycholinguistic chapter defining bilingual language systems in a similar manner, labeling them as either "compound" (where the L2 is structured and dependent upon the L1) or "coordinate" (where the two languages exist as independent grammars). While neither of these classifications remain in use today, these studies helped to give rise to a new strand of linguistic research focusing on the nature of language as it is represented in the minds of bilinguals. Nearly 70 years later, as many as 20 major models/hypotheses have been proposed to account for the way bilinguals process their languages. While it is beyond the scope of this chapter to thoroughly describe and evaluate each of these contributions, I wish to briefly mention just a few of these landmark studies. Each of these have helped to shape the way we think about the bilingual mental lexicon and the way we process languages.

[1] It is interesting to note that there has not been complete unanimity among language researchers about this approach. Some researchers (e.g., Glucksberg, 1984), particularly in the early days of modelling the mental lexicon, have argued that issues central to the way we view the lexicon are ultimately conceptual and not empirical issues at all.

One of the most influential ideas in the early days of research into bilingualism was Kolers's (1963) *shared* vs. *separate lexicon hypotheses*. Potter, So, Von Eckardt, and Feldman (1984) also made significant contributions to the field in the *word association hypothesis* (that L1 and L2 lexical items may be linked, i.e., associated, in the lexicon)[2] and the *concept mediation hypothesis* (to be explained in the analysis of the Revised Hierarchical Model in the section immediately below). De Bot's (1992) model of bilingual production was an important step forward, itself based on Levelt's (1989) highly influential monolingual model of speech. In 1994, Poulisse and Bongaerts introduced the notion of spreading activation to their model of bilingual lexical access. Kroll and de Groot introduced both the *developmental hypothesis* (that L2 words would shift from L1 mappings to conceptual ones as L2 proficiency increased) and the distributed feature model (focusing on specific aspects of lexical and conceptual processing) in their 1997 chapter.

While I can only mention them in passing, the significance of each of the studies above, at least at one time during the development of research into bilingualism, was great. In some cases, these represent excellent first attempts to explain the complexities of the bilingual lexicon. Others remain state-of-the-art descriptors and predictors of bilingual processing, but have been relegated here to a mere mention as they are not necessarily the most germane models to the current purposes (i.e., analyzing L2 word association behavior and what it may reveal

[2] This is not the sense in which "word association" is used elsewhere throughout this thesis.

about the bilingual lexicon). In the sections below I will examine in detail a few of the most pertinent models for the studies at hand.[3]

2.2.1 The Revised Hierarchical Model (RHM)

By far, the most referenced model of the bilingual lexicon is Kroll and Stewart's (1994) Revised Hierarchical Model (RHM). This is due in part to its simple design (seen in Figure 2.1) and its ability to parsimoniously account for a great number of research findings concerning the manner in which bilinguals process language. The model makes specific predictions concerning processing during translation. For example, as L1 representations are likely to access meaning directly, translation to the L2 may be conceptually-mediated and hence slower. Translation from the L2 to L1, however, should be facilitated by strong L2-L1 ties. Kroll and Stewart's (1994) own experiments showed evidence of this asymmetrical processing. The results of these studies showed that translation from L2 to L1 was in fact slower than translation from L1 to L2. Also, only L1 to L2 translation was affected by the manipulation of semantic variables. The researchers concluded that transfer from their L1, as a mediator of access to their L2, was responsible for subjects' longer translation times.

Another prediction made by the RHM is that, as greater L2 proficiency is developed, L1 transfer is no longer necessary. Direct access can then be achieved from the L2 to the underlying concept. The model received a great deal of support

[3] For detailed histories of these and other models and hypotheses proposed to account for bilingual representation and processing, I direct you to Keatley (1992), Smith (1997), French and Jacquet (2004), and Jiang (2015). Each of these papers include thorough summaries of bilingualism research until that point in time.

from developmental studies in this regard (e.g., Kroll, Michael, Tokowicz, & Dufour, 2002; Sunderman & Kroll, 2006; Talamas, Kroll, & Dufour, 1999). These studies showed that less proficient learners appeared to access L1 representations during L2 translation. There was also evidence of a shift from using transfer from the use of L1-L2 transfer as a translation strategy to being able to conceptually process the L2 without L1 mediation. This shift from L1 lexical mediation to direct conceptual access is illustrated in Figure 2.1 by way of asymmetric connections between the components of the model. Thus, there is a stronger link (represented by a solid line) from the L2 to the L1 than there is in the opposite direction (the dotted line). There is also a stronger connection between one's L1 representation and its concept than the one between one's L2 representation and the concept.

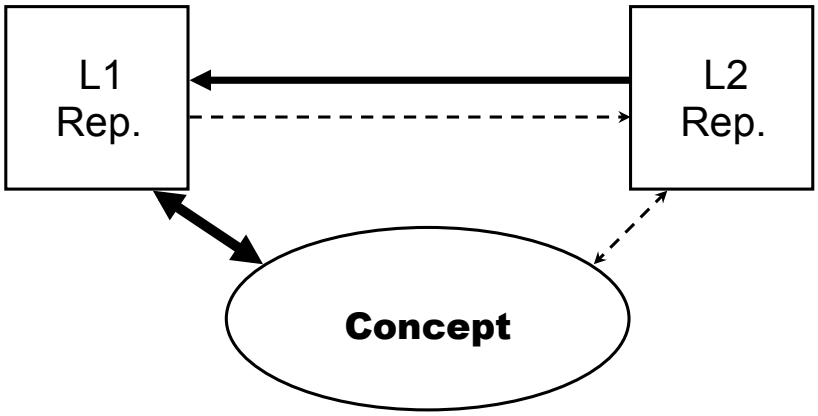


Figure 2.1. The Revised Hierarchical Model (Kroll & Stewart, 1994, p. 158).

While hundreds of researchers have cited the paper in which the RHM first appeared, not all of the subsequent research has supported the conclusions described above. Thierry and Wu (2007), for example, conducted a study using highly proficient Chinese-English bilinguals. By employing a measure of event related potentials (ERPs), the results showed that L1 representations were still activated during translation despite subjects' having achieved proficiency in their

second language. Guo, Misra, Tam, and Kroll (2012) also observed L1 activation in both ERP measures and in behavioral ones when proficient Chinese-English bilinguals were asked to determine whether a Chinese word was the correct translation of an English one. However, when the researchers manipulated the timing of the presentation of the L1 (Chinese) words, subjects appeared not to access L1 representations. The researchers concluded that proficient learners in these kinds of studies may be accessing L1 equivalents *after* performing translation tasks.

The studies described above are only two of many inspired by the RHM since its introduction. In their survey of the research, Brysbaert, Verreyt, and Duyck (2010) found more than 300 citations to the paper in which Kroll and Stewart (1994) introduced the model only 15 years earlier. Depending on which aspect of the model was being tested, however, some of the key implications of the RHM have not been borne out in the literature. One of the key hypotheses of the model is the asymmetry between translation from the L1 and from the L2. Brysbaert et al. (2010) reported that out of 54 studies designed to evaluate this aspect of the model, only 29 (54%) found evidence in support of this asymmetry.[4] Yet Brysbaert and his colleagues also found that 19 of 20 studies examining the developmental shift believed to occur with increased L2 proficiency (i.e., the move from L1 mediation to direct L2-concept access) showed support for this aspect of the model. Similarly, eight of nine studies testing the notion of a single, common

[4] See Kroll and De Groot (1997) and Kroll and Tokowicz (2001) for discussions of factors that may lead to these discrepant findings. See also Kroll, Van Hell, Tokowicz, and Green (2010) for their defence of the RHM and a review of studies both supporting and failing to support the model.

conceptual store found support for this aspect of the RHM. Based on these mixed findings, Brysbaert and Duyck (2010) have argued that it is time to "leave behind" the RHM for another model of bilingual processing. To date, however, only one other model has shown such promise: the bilingual interactive activation model (BIA).

2.2.2 The bilingual interactive activation model (BIA)

The bilingual interactive activation model (BIA; Dijkstra & van Heuven, 1998; Dijkstra, van Heuven, & Grainger, 1998; van Heuven, Dijkstra, & Grainger, 1998) was developed, in part, to account for the fact that recent research findings appeared not to support the notion of independent lexical representation between languages – one of the assumptions underlying the development of the RHM. Thus, its creators needed the model to simulate an integrated L1/L2 lexicon that allowed parallel access. At the same time, however, the BIA had to account for asymmetrical processing and the effects of language context (see Marian & Neisser, 2000).

The BIA borrows part of its name and some of its features from the interactive activation model (IA; McClelland & Rumelhart, 1981) introduced almost 20 years earlier. The IA model is essentially a three-tiered system for word recognition. The three levels of nodes are designed to recognize (from the bottom) the features of letters, the letters themselves, and finally words. Activation is competitive within nodes, such that activation of one letter, *P* for example, simultaneously inhibits the activation of other related possibilities, such as *R* or *B*. Across tiers, nodes may inhibit or excite each other, both in top-down and bottom-up directions. The most

obvious difference between this model and the BIA is that the BIA is bilingual. Thus, the BIA incorporates an integrated bilingual lexicon at the word level. The other important difference is that the BIA includes a language node at the highest level of the model (Dijkstra et al., 1998). This fourth tier performs an important role, recognizing and labeling lexical input as members of either the learner's L1 or L2 vocabulary. In so doing, it strongly influences language selection and inhibition, facilitating further language processing in one or the other of the learner's languages. In this way, the model well accounts for many of the finding described above.

While I cannot include a complete review of all the literature relevant to comparing and evaluating these models, I may say that my own view has shifted over the course of completing these studies. Originally, I considered the RHM to be an elegant and succinct (and intuitive) means of depicting the complexities of the mental lexicon in L2 learners. Having reviewed the literature, however, my view now is somewhat like that suggested by the title of Brysbaert and his colleagues' (2010) review paper. I attribute the "popularity" of the RHM to the way its simple principles generate clear, easily testable hypotheses. These result in data sets that either support or fail to support the hypotheses in what is often an unambiguous way. Such results are easily digestible by journals' editors and readerships and thus a rapid proliferation of articles and citations is precipitated.

The BIA on the other hand, has "interaction" at its core. Letter features, letters, words, and even languages are continually activated and inhibited in cycles that feed back and forth between its component nodes. Slow translation times could be

attributable to inhibition or interference at any of these stages, or perhaps to excitation/activation within the lexicons of subjects who had translated more quickly. The BIA is thus an inherently messy depiction of the workings of the bilingual lexicon, but I would say it is the most accurate model we have. In this sense, it is a "roadmap", not a "hypothesis-generator", to borrow Brysbaert's labels. The bilingual lexical pathway has many twists and turns, but the BIA may be the most accurate tool we have with which to navigate. The RHM now appears to me to be a somewhat simplistic view of the highly complex bilingual lexical process. Its value as a hypothesis-generator and stimulator of research are clear, but it may not provide the most accurate depiction of the lexical process.[5]

With these views of the mental lexicon as background, we may now turn our attention to the methodology we will use to investigate it in the current studies: word association.

2.3 Applications: What L2 word association can tell us

Before looking at the details of the WA research methodologies to be applied in the current studies, it is important to have an understanding of the breadth of second-language research to which WA methodologies may be applied. It is beyond the scope of this dissertation to provide a detailed account of the wide variety of L2 research studies in which WA methodologies have been utilized (see Meara, 2009,

[5] It should be mentioned here that in the years following the introduction of the BIA, Dijkstra and van Heuven (2002) expanded the scope of the model by developing the BIA+. This bilingual "architecture" of word recognition includes a semantic node, in which meaning is processed, and a task schema which incorporates specific steps for the task at hand (e.g., production or translation).

Chapter 8 for an annotated bibliography of more than 100 of these studies). However, reviews of a few example studies will provide a snapshot of this research and help to illustrate WA's potential as a research tool in linguistics generally and in second language studies in particular.

To launch the discussion, I will present one landmark study to represent one research area to which WA methods may be applied in each of the next subsections. Selected as a result of extensive reading in the field, the selection process was based on a number of priorities. Among these were the facts that the studies speak directly to the research areas of this thesis. Secondly, they had to be pivotal in their contribution to the field. That is, either these studies presented new and important findings or, in particular, they provided new methodological approaches. Finally, notwithstanding the criticisms I will raise here, they are well-executed studies. They are likely to be replicable and thus have longevity as points of reference in future research.

While the purpose here is really to illustrate how WA research is conducted – rather than the results of the studies, per se – I will scrutinize the methodology of each study such that problematic features might be avoided in my own research as presented in the chapters to follow. Specifically, the L2 association studies to be reviewed in this section involve, firstly, the investigation of L2 vocabulary depth and the acquisition of the many elements that contribute to it (e.g., spelling, knowledge of collocations, knowledge of word class). Another strand of research to be presented here involves the exploration of the structure of the mental lexicon via WA methods. Rather than the content of the mental lexicon, these studies

explore what we can know about the organization of the mental lexicon via WA. The third strand looks at the creation of WA tests. The aim of these tests is to use WA methods to assess L2 proficiency. These studies differ not only in terms of the research areas to which the WA method is applied, but also in their research approaches (e.g., longitudinal case studies vs. controlled laboratory experiments). At the same time, however, all three attempt to answer important questions about second-language lexis by way of WA.

2.3.1 WA as a measure of L2 vocabulary development

One of the beliefs that a number of linguists hold about WA is that associative knowledge represents a single aspect of word knowledge. One of the hypotheses stemming from this belief (i.e., Stage 1 in Figure 1.1) is that an exploration of the development of this knowledge will reveal important aspects of an individual's vocabulary development – either in terms of its breadth or depth. A variety of vocabulary studies in which WA was used as a measure of word knowledge have been founded on this assumption (including Henriksen, 2008; Read, 1993; 2004, etc.). Among these is a longitudinal case study of vocabulary acquisition by Schmitt (1998b). In this research, he traces the acquisition of 11 English words by three adult ESL learners over the course of a year. Two of Schmitt's stated purposes are to go beyond the simple "not acquired/acquired" dichotomy (p. 283) and to examine deeper word knowledge than merely the acquisition of meaning-form pairs. Schmitt's approach was thus to examine the development of vocabulary knowledge generally, as the incremental acquisition of a variety of specific types of word knowledge for a specific group of words. By "word knowledge", I am referring to both productive and receptive aspects of formal (spoken, written, etc.),

semantic (synonymous, conceptual, etc.), and usage-related (collocations, register, etc.) features of a word.[6] To uncover aspects of his participants' word knowledge, Schmitt tested four variables specifically: spelling, morphology, word senses, and – of particular relevance to the current studies – word associations.

In a bid to track the incremental growth of word knowledge, Schmitt selected words that the learners were likely to encounter in the course of their academic life (i.e., postgraduate learning in the United Kingdom). The words were polysemous, having at least three meaning senses. In order to trace early acquisition, he also included low frequency items likely to be unknown to the participants. The method consisted of three interviews held at approximately 6-month intervals. Each included the following questions:

1. *How do you spell _____?* – to measure knowledge of spelling.
2. *Please give the first 3 words you think of when you hear the word _____.* – to measure knowledge of associations. Data were compared to a list of native speaker norms.
3. *What word class (part-of-speech) is _____?* – to measure grammatical knowledge. Depending on their responses, participants were asked if they could provide other forms of the word (i.e., the remaining three forms from: noun, verb, adjective, adverb).
4. Finally, as a measure of productive knowledge, participants were asked to explain the senses of the words they knew using definitions, examples, drawings, gestures, or by using the word in a sentence. As a receptive measure, the words and their respective word classes were used as prompts.

Schmitt's results showed that participants were quite adept at spelling. However, even advanced learners such as these showed rather limited development in the

[6] For a detailed outline of what it means to know a word, see Nation (2013, Chapter 2). See also Richards (1976) and Bogaards (2000) for earlier classifications of word knowledge.

mastery of meanings. In fact, only two of the three subjects appeared to show steady improvement in this knowledge over time. Generally speaking, association knowledge (as measured by the number of responses to Question 2 that coincided with normative responses from NS respondents) and grammar knowledge increased with knowledge of word meaning. In fact, these types of knowledge could be demonstrated even in the absence of meaning knowledge.[7] For example, even if respondents were uncertain about which specific tool "wrench" referred to, they might still have been able to respond with nativelike associations (e.g., "I thought of a hammer.") and would understand that it referred to noun.

Schmitt's study provides a solid example of one of the ways in which WA methodology can be utilized in the investigation of L2 vocabulary acquisition. However, the paper also reveals a number of methodological issues that deserve further scrutiny. One of these is the manner in which knowledge of word meanings was scored. As an indicator of meaning knowledge, Schmitt required that subjects use the words in sentences. This appears to conflate knowledge of meaning with knowledge of use, or perhaps collocational knowledge. This point is perhaps minor and might easily be rectified by labelling the measure differently. However, in a study examining the acquisition of specific types of word knowledge, it would be best to keep meaning and use well-defined.

[7] Ultimately, Schmitt's goal was to uncover whether the acquisition of individual knowledge types followed a discernible pattern (e.g., mastery of form-meaning pairs followed by mastery of associational and collocational knowledge). While he "found no evidence of a developmental hierarchy for word knowledge types" in this study (N. Schmitt, 1998b, p. 309), he and other prominent vocabulary researchers continue to speculate on the incremental nature and the acquisition order of aspects of word knowledge (Nation, 2013; N. Schmitt, 2010, 2014; N. Schmitt & Meara, 1997).

I raise this issue as a reminder of the importance of how vocabulary knowledge is conceptualized. Meaning and usage may represent an important distinction warranting treatment as discrete variables, or they may represent two inextricably linked aspects of word knowledge that, by necessity, must be assessed together. Issues like these, and the methodological decisions they precipitate, require continuous scrutiny by vocabulary researchers. Indeed, this is only one of the many types of issue that must be considered when categorizing and conceptualizing WA responses. These considerations will be addressed in detail in Chapters 3 and 4.

Another aspect of Schmitt's study to consider is the comparison of word associations with native norms as a means of measuring lexical development. Native-likeness, in terms of the distribution of types of response, may no longer be the cut-and-dried measure of proficiency that it was once considered to be. Fitzpatrick (2007, 2009) appears to have demonstrated that L1 responses are not necessarily homogeneous, nor do L2 associations appear to become more native-like with increased L2 proficiency. Based in part on those findings, my colleagues and I have since argued strongly against their use in WA studies (e.g., Fitzpatrick & Racine, 2014; Higginbotham et al., 2015; Racine, Higginbotham, & Munby, 2014). One of these background studies in particular (Fitzpatrick, 2009) showed that L2 associations become more similar to respondents' own L1 responses (rather than to the associations of native speakers) as L2 proficiency increased. This is an important finding and it will receive due scrutiny in Chapter 4.

Thirdly, it would be interesting to know why Schmitt "went back and explained the different meaning senses to the student" (p. 294) after each interview. The stated reason was to maintain the learners' interest. However, as the goal of the research was to measure the natural incremental change in word knowledge over time, this action might defeat the purpose of the study. As we cannot know how much meaning knowledge the subjects would have acquired on their own outside of the research setting, it is difficult to assess how realistic the findings are in regard to the measure of sense knowledge in the second and third interviews.

The analysis of this paper benefits from a comparison with a more recent case study conducted by Churchill (2007). Churchill's study was a three-month case study of the acquisition of a single, two-character Japanese compound by a single subject – himself. The studies differ in that Churchill's approach is observational and introspective; a diary account of his growing lexical knowledge in "real time" over three months. He did not attempt to objectively test his knowledge at any point or to in any way quantify it. Nor did he attempt to draw conclusions that might potentially be generalizable to a broader population of language learners.

A major disadvantage of Churchill's approach is that the study, unlike Schmitt's, lacks objectivity and hence does not offer construct validity. Construct validity is the extent to which variables are operationally defined in a way that makes them objectively measurable. While Schmitt attempted to objectively define the varying types of word knowledge in terms of incremental, numerical scales, Churchill described aspects of vocabulary knowledge less methodically, more in the *acquired/not acquired* manner that Schmitt set out to avoid. In terms of external

validity (the extent to which the results can be generalized to a broader population), neither of these studies – nor most case studies for that matter – can make strong claims. Both involve extremely limited subject pools. In neither case are quantitative analyses applicable, so we cannot see clearly how generalizable these findings may be beyond these specific participants. As neither researcher writes about the generalizability of their findings, we can infer that they feel it would be inappropriate to generalize from them.

In some respects, these two studies are complementary. Churchill's findings are interesting, but they are a set of personal observations, the associations he had about his knowledge of a single Japanese compound – rather than a set of results from a formally operationalized experiment. They tell us about authentic, contextualized lexical acquisition. One of the benefits of this kind of case study is the detailed scrutiny of data that it affords. Indeed certain insights may be yielded that might not become apparent in less *in situ*, experimental studies. Schmidt and Frota (1986; see also Schmidt, 1990), for example, postulated the *noticing hypothesis* based on case study data.[8] Later, results from a number of experimental studies found support for this hypothesis (see Truscott, 1998 for a review) and *noticing* is now a well-established concept for many teacher-researchers. This is not to oversimplify the research process or imply a linearity that does not exist, but it is somewhat ironic that Schmitt's (1998b) study preceded Churchill's (2007) by almost a decade. Churchill makes many interesting and detailed observations

[8] The *noticing hypothesis* states that lexical and grammatical forms are more likely to be acquired, as evidenced by their appearance in subsequent output, if the form was consciously recognized during prior input.

that appear to provide the basis for the kinds of hypotheses that could be tested in subsequent, quasi-experimental studies like Schmitt's.[9]

2.3.2 L2 associations as an indicator of language proficiency

In studies like Schmitt's above, word associations are collected to potentially reveal development of a particular aspect of word knowledge (i.e., associative knowledge) over time. A related methodology stems from the assumption that differences in L2 proficiency will be reflected in differences in WA responses. Based on this belief, researchers have hypothesized (Stage 1 in Figure 1.1) that the associations of native speakers may be used as normative data according to which the language proficiency of L2 learners could be compared. Driven in part by this premise, a variety of WA norms lists have been compiled (e.g., Entwisle, 1966b; Kiss et al., 1973; H. Moss & Older, 1996; D. L. Nelson, McEvoy, & Schreiber, 1998; Postman & Keppel, 1970). A number of language proficiency measures that use word associations as data have also been developed (e.g., Kruse et al., 1987; Meara & Fitzpatrick, 2000; Read, 1998).

It may appear reasonable to assume that associations represent one type of word knowledge and that this knowledge develops like other types of word knowledge, becoming more nativelike as L2 proficiency increases. Indeed, this is the premise behind Schmitt's (1998b) use of the WA measure in his study analysed in Section 2.3.1. In fact, this is one of the essential premises for all native WA norms lists and

[9] Incidentally, Churchill's (2007) findings (while not completely relevant to the current focus on methodology) were that the development of word knowledge exhibits the qualities of a dynamic system in that it was non-linear and proceeded through a number of quasi-stable states.

tests of language proficiency that employ WA methods, like those cited in the paragraph immediately above. The problem with this reasoning, however, is that there is almost no empirical basis to suggest that this is actually the case. There have been very few longitudinal studies examining change in WA responses over time, fewer still have examined second language associations.

The idea of WA norms as a measure of stereotypy dates back to the use of free association as a psychometric measure in the early 20th Century. One study in particular – Kent and Rosanoff (1910) – has had long-lasting, undue influence on subsequent studies. Essentially, the researchers used a list of words to gather normative data. The notion was that if one's responses deviated far enough from those norms, one could be diagnosed as "insane". While very few researchers or clinicians would find validity in this method today, this approach – using norms to measure stereotypy – remains. Indeed, even the stimulus words themselves, now known as the *Kent-Rosanoff list*, were repeatedly adopted in subsequent studies for decades thereafter (e.g., Meara, 1978; Postman & Keppel, 1970; Singleton, 1999, Chapters 5 and 6). Besides mere convention, it is not always clear on what grounds these cues were recycled so extensively.

More recently, Fitzpatrick and Munby (2014), describe how they designed such a word association test, modelled on a test originally constructed by Kruse et al. (1987). Their first step was to carefully select cue words. To ensure that subjects (Japanese learners of English with proficiency levels ranging from "elementary to intermediate"; p. 102) were likely to understand the cues, they used the first 1000-word band of the British National Corpus (BNC). From these 1000 items, they

removed those that would elicit strong primary responses, proper nouns, those that would elicit proper nouns, and those that existed as loanword cognates in Japanese. With the aid of the Edinburgh Associative Thesaurus (EAT; Kiss et al., 1973), they further removed cues where the primary response was the same as another word on the list, as well as those where relatively popular responses (comprising at least 6% of total responses) were the same as those of other cues. The remaining 50 words were then used as cues in a preliminary study in which responses from 82 (L1 Japanese) participants were compared to a list of native-speaker response norms. The results for each cue were converted into scores and examined against subjects' TOEIC scores. The 10 cues with the greatest potential to discriminate between learners of varying proficiency levels were then selected to be cues in the final WA test.

Seventy-one Japanese learners of English took the WA test. The test included 20 cues: 10 selected via the criteria described above, and 10 from the original study (Kruse et al., 1987). Subjects were instructed to enter up to 12 associations for each of the 20 cues. They were also given a cloze test, an L1-to-L2 translation test, and the TOEIC test as measures of L2 proficiency. Associations were measured in two ways: terms of number of responses and stereotypy. For the latter, responses to the original cues were compared to the Minnesota norms (Jenkins, 1970).[10] Responses to the new cues were compared to a list of native-speaker norms created from the responses of 114 native speakers who were asked to provide five associations to each cue.

[10] Actually, the Jenkins (1970) study was mistakenly cited by Fitzpatrick and Munby as "Postman and Keppel (1970)", the edited volume in which the Jenkins paper appears.

The authors draw our attention to three sets of findings and conclusions in particular. First, correlation scores between the cloze test results and both the number of responses and stereotypy scores are very similar to the findings of Kruse et al. (1987). Secondly, while there were positive, statistically significant correlations between the three proficiency measures and both of the WA measures (number of responses and stereotypy), stereotypy correlations were consistently stronger. This contradicts Kruse et al.'s finding that the number of responses was the best predictor of proficiency and suggests that the quality of WA responses, rather than the quantity, becomes more nativelike as proficiency increases. Given the criticisms of the Fitzpatrick and Munby paper I offer below, it is difficult to determine whether the quantity or quality of WA responses is the better measure of proficiency, and whether there may simply be differences between the subject pools that account for the discrepant findings across studies. Finally, in the authors' view, the strong correlations between the three rather varied measures of proficiency and the stereotypy scores indicate that, with careful selection of stimuli, Kruse's original test can be improved upon and can be utilized to discriminate levels of L2 proficiency.

The study is well considered in that Fitzpatrick and Munby have systematically justified the need for the creation of this test through their examination of the problematic issues in the original study (Kruse et al., 1987). The research has also been conducted in a principled manner, seen particularly in the methodical process by which the association cues were selected. At the same time, however, aspects of the account raise questions. One set of questions surrounds the

positioning of the research in light of the authors' prior studies. Aspects of the methodology employed here also require a closer examination.

Firstly, there is a discrepancy between the authors' prior work and one of the methodological choices made in this study. Both Fitzpatrick and Munby have presented arguments against the use of native-speaker norms as a measure of proficiency in L2 WA research (Fitzpatrick et al., 2015; Fitzpatrick & Racine, 2014; Higginbotham et al., 2015; Racine et al., 2014). Indeed, Fitzpatrick (2009) has presented evidence that L2 word associations actually become more like respondents' own L1 association profiles with increased proficiency – that is, *not* more like native-speaking respondents' profiles. In the present study, however, stereotypy measures are employed to compare learners' responses to native norms, both in the preliminary study and after administering the WAT itself. If Fitzpatrick's prior assertions are correct, comparisons to respondents' own L1 associations might have been a more valuable means of discriminating between proficiency levels. While I think it is not unreasonable to continue to compare second language word associations with native speaker norms, an acknowledgment of their prior studies and an explanation of the rationale behind why they chose to do so would have been a welcome addition to this study.

A second methodological issue to consider here is that the researchers could not apply the same norms to the two groups of responses they wished to examine. This makes it very difficult to draw any strong conclusions about the efficacy of either group of test items. Further, the authors argued against the use of published native norms lists (as used in Kruse et al., 1987) on the grounds that they were dated and

were compiled from participants' single (rather than multiple) responses. One wonders, then, why they continued to employ them here, and only to compare the responses collected from Kruse's original cues. Also, it is not clear why they chose to compile their own list of native norms from only five responses per cue, while non-native respondents were asked to provide up to 12 responses. While these are not fatal flaws for the research, I have once again drawn attention to the types of methodological issues to be considered in WA research. This study reveals another aspect of language research – second language proficiency testing – to which WA methods might fruitfully be applied.

2.3.3 Revealing connections and processes in the bilingual mental lexicon

The assumption underlying this third strand of enquiry is that development of the lexicon – as reflected in changes in the organization, processes, and relative strengths of connections between lexical entries – can be revealed through changes in WA responses and, in particular, through changes in types of WA response. Researchers basing their studies on this premise have investigated, for example, the syntagmatic-paradigmatic shift. This term refers to a change in WA response types over time, from predominantly collocational (syntagmatic) relationships with their cues to ones belonging to the same word class as their cues (i.e., in paradigm with it). This apparent shift has been observed in the L1 development of children and a parallel process has been hypothesized for the development of L2 responses (e.g., Entwisle et al., 1964; Francis, 1972; K. Nelson, 1977; Söderman, 1993).

Zareva (2007) presents a study based on these assumptions. Unlike Schmitt's, Churchill's, and Fitzpatrick and Munby's studies described in the sections above, the research is neither a case study nor a contribution to the language testing literature. Instead, the study attempts to use WA methods as a means of comparing the organization and processes in native speakers' mental lexicons with those in second language learners' mental lexicons. It is interesting to note that despite Zareva's title promising a study of the "structure of the mental lexicon", she neither grounds her work in models of the lexicon (see Section 2.2) nor cites the relevant literature (e.g., Kroll & Stewart, 1994; van Heuven et al., 1998). How this study's results relate to the models is an issue I will return to below.

In the paper, Zareva (2007) describes a word association study in which the responses of 29 native speakers of English (NS) are compared to those of 29 advanced- and 29 intermediate-level non-native speakers (NNS). This research is a somewhat unusual example of WA research in that Zareva did not adopt the traditional WA methodology in which stimulus words are presented aurally or in written form and subjects respond with the first word that comes to mind (orally or in writing). Instead, stimuli were presented along with information about their respective word classes, and also a familiarity measure based in part on Paribakht and Wesche's (1993). In total, 73 stimulus words were chosen via fixed random selection from a learner's dictionary and subjects were required to respond to each word with one of the following responses (Zareva, 2007, p. 133):

1. I have not seen this word before;
2. I have seen this word before but I don't remember what it means;
3. I think this word means _____ [provide a synonym or brief explanation];
4. I know that this word means _____ [provide a synonym or brief explanation].

Subjects who responded with either (3) or (4) were asked to complete a fifth option (I associate this word with _____, _____, _____) in which they were to answer with up to three words they associated with the stimulus.

Quantitative analyses of the results were conducted to examine differences between the numbers of associations between groups, as well as differences in response commonality and response heterogeneity within groups. Qualitative analyses examined proportions of response types based on the traditional classification of paradigmatic (same word class as the stimulus), syntagmatic (different word class), and clang (phonologically related) responses. Quantitative and qualitative results were then compared to examine possible interactions. The results of an ANOVA showed significant main effects for all three of the quantitative measures: total number of associations (NS > advanced NNS > intermediate NNS), response commonality within groups (NS > advanced NNS > intermediate NNS), and response heterogeneity (advanced NNS > NS > intermediate NNS). Post hoc comparisons showed that these differences were attributable mostly to significant differences on all three measures between the intermediate NNS and the other two groups. Advanced NNS and NS did not differ significantly on any of these measures. In terms of the qualitative analysis, notably, there were no clang responses by any of the three groups. Also, all groups had responded with a majority of paradigmatic responses. There were no significant differences between the proportions of paradigmatic vs. syntagmatic responses. A correlation measure showed that the quantitative and qualitative measures were not strongly related.

Taking these results as a whole, Zareva concluded that 1) meaning connections in the lexicons of advanced learners and NS closely resemble each other, 2) "qualitative characteristics of lexical organization" (p. 141) do not differ between NS and L2 learners who have achieved at least an intermediate level of proficiency, and 3) that quantitative and qualitative features of associative patterns develop fairly independently of each other.[11]

Zareva's study utilized an interesting combination of word knowledge and word association methodologies. Notwithstanding the critical analysis I present below, they provide a valuable opportunity for exploring WA methodology more deeply. I will focus on the researcher's choice of stimuli and her adoption of a measure of word familiarity in place of the traditional WA methodology below.

Although Zareva stresses the systematic approach by which she selected the stimulus words, some of these should have been omitted from this study for a variety of reasons. For example, loanwords from foreign languages (e.g., *abattoir, cassava*) may exist as atypical representations in the lexicons of both learners and NS participants. As yet, there is very little evidence from WA studies to indicate

[11] Zareva does not situate her findings in the context of models of the mental lexicon (as discussed earlier in this chapter), but an argument can be made that the results generally provide support for Kroll and Stewart's (1994) Revised Hierarchical Model. Since similar results were elicited from both highly proficient learners and from the NS respondent groups, there is at least tangential support for Kroll's developmental hypothesis (Kroll & De Groot, 1997). Both qualitatively and quantitatively, the similarities between proficient learners' associations and those of native speakers suggest the possibility that the reliance on L2-L1 links in the lexicon has been replaced by stronger L2-concept links as L2 proficiency has developed.

how loanwords are stored in the lexicon or accessed during WA tasks. It is certainly possible, therefore, that loanwords at least in some cases are tagged and processed as a kind of foreign word even within one's L1 lexicon. This argument finds support in the fact that, for many, one salient aspect of word knowledge for these words consists in knowing that they are in fact foreign words. Therefore, it is difficult to say precisely what association data elicited by these kinds of cues tell us about the lexicons of either population of respondents. Other stimuli were extremely difficult or infrequent cues (e.g., *bursar, gambol*) or were derivational forms far less frequently occurring than their more typical base forms (*amoral, rigidity*). Still others perhaps should have been omitted on account of their phonetic or morphemic complexity (*contravention, lackadaisical*). The processes brought to bear in deciphering these difficult words may be different from those applied to more typical cue words utilized in studies like these (e.g., drawn from the most frequent bands of a corpus or word list). While all of these words were drawn from a learner's dictionary, it is not clear how relevant they would be in revealing elements of the emerging lexicon except in the case of quite advanced learners. This criticism is not to say that Zareva's cue selection process was in any way unprincipled. Once again, I am attempting to focus on the types of issues that shape all WA research methodologies including my own below. Zareva's subjects may have been substantially more proficient L2 learners than the respondents who took part in the research I will present in subsequent chapters. Thus, the relative difficulty of infrequent or morphologically/phonetically complex cue words may have been less of an issue in her study.

Also, Zareva might argue that the adoption of the familiarity measure eliminates the problem of difficult, infrequent, or unknown stimuli. Indeed, her final analyses only included associations for stimuli with which individual respondents reported to be familiar (Options 3 and 4 of the familiarity measure) and had correctly identified one of its meanings (as judged by the researcher). One can imagine, however, how NS were able to respond correctly to more stimuli than NNS (especially in the intermediate group), who may only have been able to respond to a small fraction of them. In the final analyses then, it is likely that responses were elicited from potentially different cues, and certainly different-sized sets of cues between groups. For my own research (Chapters 3 through 6), this comparison of responses from different sets of cues would be a methodological confound that would prevent me from drawing conclusions from the comparison of NS and NNS responses.

Perhaps the most serious problem with the familiarity measure is that, by nature, it requires respondents to attend to word meaning. In considering whether or not they know the cue, subjects necessarily give the stimulus deliberate, conscious attention and activate semantic knowledge related to the word. Semantic knowledge is thus made salient in the minds of subjects and impacts upon any subsequent WA tasks. Associations elicited under these conditions are not representative of the type of spontaneous, "free" responses believed to reveal links in subjects' mental lexicons. This effect, known as *semantic priming*, is a robust and well-documented phenomenon in the psycholinguistic literature. The predominance of paradigmatic responses in Zareva's study is indicative of semantic priming. This priming effect, while somewhat problematic for Zareva's

study, is serendipitous for the current dissertation. Indeed, Zareva has inadvertently introduced yet another methodological detail that will be (intentionally) applied within the studies introduced in Chapters 5 and 6. Prior research in which priming methods were implemented will be introduced in the following section.

2.4 Methodology: Adopting and adapting WA methods

The descriptions above – of research areas to which WA methods have been applied – are intended as a brief sampling of the range of L2 phenomena that may be fruitfully researched using WA. The focus of the current section will be on WA methodologies themselves, illustrating aspects of Stages 2 through 5 of the WA research process (as depicted in Figure 1.1). Besides illuminating aspects of the WA research process in general, the studies examined here present methodologies that will be adopted in the research in subsequent chapters. The exploration of the studies presented here will demonstrate how WA methods can be adopted – either alone, or in conjunction with other methods – and adapted for a variety of research purposes. In so doing, the nuances of WA research methodology and the complex decision-making in which WA researchers must engage will also be revealed. As illustrated in Figure 2.1, these decisions impact upon each step of the WA research process. The research methods to be introduced here involve the collection of reaction time data and the implementation of lexical decision tasks and restricted word association tasks. The processes of adopting and modifying these methods will be brought to bear in the research I will present subsequently and are essential to furthering the goals of this dissertation as outlined in Chapter 1. As in

the previous section, the research examined here will again contribute to the bigger picture of the breadth of L2 lexical studies to which WA may be applied.

2.4.1 Associations, reaction times, and lexical decision tasks

Stage 4 of the WA research process (see Figure 1.1) comprises the actual WA trials. Often limited to simply the presentation of cues and the recording of responses, this stage of WA research is ripe with potential variations to advance opportunities for addressing new research questions. As in Section 2.3.2 above, one simple variation on the conventional WA trial is to collect multiple responses from each respondent for each cue. A number of researchers have elected to conduct WA trials in this manner (e.g., Fitzpatrick & Munby, 2014; Kruse et al., 1987; Randall, 1980). One of the underlying assumptions of this method is that the elicitation of multiple responses will provide a richer view of respondents' lexicons and particularly of the representations of individual words therein. Another seldom explored, but potentially fruitful variation on the conventional WA trial involves the collection of reaction time (RT) data. In WA research, an RT score refers to the time elapsed between the presentation of a cue and the production of a response. Researchers who choose to record RTs do so under the assumption that differences in RTs allow them to infer differences in the processes intervening perception of stimuli and the production of responses.

RT methods remain a staple methodology in a variety of wide-ranging cognitive psychological research strands (e.g., Meiran, Chorev, & Sapir, 2000; M. J. Nissen & Bullemer, 1987; Rosch & Mervis, 1975; L. M. Schmitt, Ankeny, Sweeney, & Mosconi, 2016; Sexton & Cooper, 2017; Shiffrin & Schneider, 1977). Among these strands

are a wide variety of priming studies (Arai, van Gompel, & Scheepers, 2007; Branigan, Pickering, Stewart, & McLean, 2000; Dagenbach, Carr, & Barnhardt, 1990; Ellis, 1982; Feldman, 2003; Kinoshita & Lupker, 2003; McRae & Boisvert, 1998; H. E. Moss, Ostrin, Tyler, & Marslen-Wilson, 1995). Priming effects will be explained in detail in Chapters 5 and 6 in relation to the studies presented there. For now, it suffices to say that RT, as it is employed in priming studies, measures the degree to which language processing is facilitated or hindered by prior language input.

One can see how RT data can aid in the assessment of models of the mental lexicon as presented above (Section 2.2). The RHM, for example, makes very specific claims about the speed of lexical processing during translation tasks. RTs are dependent on whether processing is mediated by access to L1 representations and on the proficiency of the L2 user. In fact, priming studies have the potential to illuminate a wide variety of cognitive functions and processes both within and outside of linguistic research. Differences in latency have been used to reveal the existence of lexical competition while listening (Chambers & Cooke, 2009), facilitation in word recognition tasks (Meyer & Schvaneveldt, 1971), and in recognition tasks for collocations (Wolter & Gyllstad, 2011), to give just a few examples. Indeed, it appears that RT research is attracting growing interest from second language researchers (e.g., Finkbeiner, Forster, Nicol, & Nakamura, 2004; Jiang, 2012; McDonough, 2006; McDonough & Trofimovich, 2009; Pellicer-Sánchez & Schmitt, 2012; Racine, 2014; Scheffler, 2015; Trofimovich & McDonough, 2011).

One experiment in which an RT measure was integrated into the WA framework was conducted by Fitzpatrick and Izura (2011). This study illustrates the versatility of WA methodology as well as the utility of RT measures. An examination of this study is also important to this dissertation because the manner in which responses were categorized – a topic to be dealt with in depth in the next chapter – is quite innovative. Here, perhaps for the first time in WA research history, responses that were potentially classifiable into more than one category were placed in "dual-link" categories. An example of this would be the response of *hairdryer* to the cue *hairdresser*. Obviously, the words are related in form, but there is a clear meaning-based link between them as well. This type of response has proven difficult for WA researchers in the past who were forced into an either/or decision. In this study, however, the response was placed into a "form and meaning" category. Finally, the inclusion of a lexical decision task (LDT; a common psycholinguistic method in which subjects must determine whether a sequence of letters constitutes a word) also points to the versatility of WA methods when combined with other approaches.

The aims of the study included attempting to address three research questions (Fitzpatrick & Izura, 2011, pp. 379-380):

1. Are certain response types more frequent or produced more quickly than others?
2. Do subjects' L2 RT profiles mirror their L1 RT profiles?
3. Are L2 responses mediated by the L1, and if so, is the mediation dependent on language proficiency?

To address these questions, 24 native speakers of Spanish were first given a 95-word L1 WA task and then another 95-word task in their L2 (English). Great care was taken in the selection of cue words. Inter-lingual homographs (i.e., words

similar in form across languages) and cognates (words similar in both meaning and form across languages) were not included. Members of each list were balanced in terms of certain features (e.g., number of letters, imageability, frequency, word class, and age of acquisition in the respective L1). Based on the lexico-semantic relations between cues and their responses, each response was placed into one of six categories including two dual-link categories: form and meaning (e.g., *newsagent–newspaper*), meaning and collocation (*peacock–feather*), collocation (*bat–man*), form (*mustard–mustang*), equivalent meaning (essentially synonymous; *sofa–couch*), and nonequivalent meaning (related in meaning, but not synonymous; *party–celebrate*). RTs were recorded for each response. After that, participants received a 72-word lexical decision task (LDT) in their first language, consisting of 36 real Spanish words and 36 invented words. Eighteen of the real words were translation equivalents of words already encountered in the English WA task. The remaining real words were unrelated to items in the previous tasks. Once again, RT was recorded.

The results showed that responses with a dual-link (i.e., form and meaning or meaning and collocation) were produced more quickly than other responses in either language. This is an interesting finding in that it suggests that the number of links between pairs of words in the lexicon determines the speed with which they can be processed. In other words, when two or more links exist between a pair of nodes in the lexicon, they appear to exist as a single, fast pathway, rather than as separate slower paths. I will unpack this notion of dual links and their possible contributions to processing in my discussion below. Another finding was that nonequivalent meaning associations (i.e., conceptual ones) were the slowest but

most frequent responses in both languages. L2 associations took longer to produce than associations in the subjects' L1, but the time difference lessened as L2 proficiency increased. The results of the LDT indicated that a priming effect existed for the L1 translations of words that had appeared as cues in the L2 WA task, at least in the case of lower-proficiency participants.

Fitzpatrick and Izura's examination of RT in this study is at the cutting edge of WA research. Elsewhere (e.g., Racine, 2008, 2011b, 2011c, 2013), I have called for WA researchers to implement RT measures as a means of more precisely examining the cognitive processes underlying WA. However, very little work has been done so far. Another aspect of this research that puts it at the forefront of WA studies is the attempt to classify certain responses as *dual-links* (i.e., simultaneously belonging to more than one response category). With few exceptions (e.g., Fitzpatrick et al., 2015; Meara, 1982), researchers have tended to ignore the fact that many WA responses can easily be classified into more than one response category. Issues involved in creating an effective categorization scheme will be addressed in detail in the next chapter. For the time being, though, this study appears to be an encouraging step forward in terms of both the classification of responses and the acknowledgment of RT measurement as a potentially useful tool for WA researchers.

Fitzpatrick and Izura's study also draws attention to other key features of WA research relevant to the studies I will present in the chapters to follow. One regards what can be inferred about the underlying lexical processes involved in WA trials. For example, their use of RT methodology is underpinned by the

assumption that if L1 equivalents are recognized more quickly in subsequent LDTs (i.e., they had been primed in previous tasks) participants must have translated L2 (English) cues into their L1 (Spanish) before responding. Leaving aside the possibility that a common semantic representation could have been accessed, this is a logical assumption. However, there is a substantial body of evidence – from eye movement and LDT research – showing that L1 representations are activated during L2 tasks *even when the L1 is not utilized in the performance of the task* (e.g., Marian & Spivey, 2003; Tannenhaus, Magnuson, Dahan, & Chambers, 2000). So, while it is reasonable to assume that L1 activation has occurred if a priming effect is observed in subsequent tests, it may not be completely attributable to the direct (i.e., intentional) translation that the researchers seem to infer has taken place in their study.

Another interesting feature of this study is how it contrasts with Fitzpatrick's own prior studies (2006, 2007, 2009). As noted earlier and laid out in Table 1.1, Fitzpatrick began, in 2006, to utilize a logical and user-friendly system of WA response classification (to be examined in detail in the next chapter). Based in part on Nation's (2001) categories of word knowledge, WA links were classified as meaning-based, position-based (i.e., collocational), or form-based (i.e., orthographically- or phonologically-related). Responses were further divided into a series of subcategories within the three main groups. Part of the intuitive appeal of this classification scheme is that meaning-based responses correspond to conceptual or semantic links in the minds of respondents while the other two categories refer to different types of lexical links. However, in the current study (i.e., Fitzpatrick & Izura, 2011), meaning-based responses are subdivided into

equivalent meaning (synonyms and lexical set members) and nonequivalent meaning (conceptual associations) subgroups. The researchers do not address the implications for further research of this departure from recent approaches. Readers are left to decide for themselves if the categorization scheme employed here has greater utility in explaining the L2 WA process and whether these categories are more psychologically valid to respondents than previous categorizations. The concept of psychological validity will be central to the arguments I make concerning the categorization of WA responses in the next chapter and in the discussion in Chapter 7.

Another feature to examine here is the manner in which the researchers interpret their results. Specifically, in discussing the preponderance of form-based responses by L2 learners they offer two possible reasons. The first – that it is common for L2 learners to focus on the form of new words, whereas L1 acquisition involves the holistic labelling of concepts – seems to make sense. The second reason, however, appears to be based on a misinterpretation or an over-extension of Kroll and Stewart's (1994) Revised Hierarchical Model (RHM) of bilingual lexical representation. The RHM, as described in Section 2.2, states that semantic links to L2 representations are characteristically weak in non-proficient learners, while connections between L2 and L1 representations are strong. What this means is that (for a native speaker of Spanish learning English as an L2) *truck* (the L2 representation) is more strongly connected to *camión* (the L1 representation) than it is to the concept of a truck. But this kind of L1 form to L2 form connection is entirely different from form-based associations where, for example, *truck* elicits

the response *duck*. The RHM cannot be used in this way to interpret form-based connections between strictly L2 representations.

Fitzpatrick has argued elsewhere (2006) that post-task interviews could be used to disambiguate associations appearing to belong to more than one response category. Using *brother-sister* as an example, when asked, respondents might say, "we always say 'brothers and sisters'" in which case this is a collocational response (i.e., position-based). If a participant says, "they're both family members", then this is a meaning-based response. Without such interviews to help disambiguate responses, the assumption in the current study is that both possible connections exist in the mind of the respondent and that both have contributed to the faster RTs found in these dual-link categories. Personally, I don't think there is enough data here to draw strong conclusions either way, but there are certainly alternative interpretations of this result that do not suggest that two links contributed to the faster RTs. If we call the explanation above the *2 connections-2 contributions* hypothesis, then I would posit the following models as potential explanations of the findings as well:

> *2 connections-1 contribution*: In this model, both connections exist, but only one (the most easily accessible, i.e., faster) is made. The fact that two (or more) connections exist implies that this item has been accessed a lot and has slightly stronger connections than single-link responses, thus leading to quicker RTs than responses determined by only a single link.
>
> *1 connection*: There is no reason to believe that L2 respondents (especially non-proficient ones) are aware of all the aspects of word knowledge implied in the connections between their responses and the cues that elicited them. While they appear to be "dually-determined", these

responses may actually have been elicited by only a single aspect of word knowledge.

One possible means of further exploring the relative contributions of multiple links between representations in the bilingual lexicon may be to combine an RT measure with restricted association tasks. That is, rather than responding freely to the cues, subjects could be given specific tasks such as the following:

- "Respond with a word that has a similar meaning": to intentionally elicit meaning-based responses;
- "Respond with a word that would follow this word in a sentence or phrase": to elicit position-based responses; or
- "Respond with a word that sounds like this word": for form-based responses.

Whether this type of research design proves useful in determining which *connection-contribution* models best explain connections between RTs and possible dual-link associations is an empirical question beyond the scope of this dissertation. On the other hand, restricted association tasks are germane to the current research, as I will employ them in two studies to be presented in Chapter 6. Therefore, in the next section I will present prior research that utilized restricted WA methods.

2.4.2 Restricted word associations

Restricted associations, as I have suggested above, are a potentially fruitful – albeit underused – source of data for certain types of second language lexical research. Restricted association tasks (also referred to as *directed* or *controlled* association) are WA trials in which subjects are required to give specific types of responses – as opposed to "free" associations. The reasons this strand of research is potentially

valuable to WA researchers include the fact that it can yield connections within the lexicon that are not revealed through conventional WA methods. The majority of conventional WA studies to date were intended to elicit the "first thing that comes to mind." These primary responses are considered to represent the strongest, most accessible links within the lexicon. Thus, we have very few clues about individuals' possible responses beyond these primary, "free" associations. Also, the fact remains that the majority of recorded responses in practically all WA studies to date reveal semantic (meaning-based) links to their cues. We thus have very little data about potential non-semantic links in the lexicon. For these reasons, restricted association data may yield another perspective on the mental lexicon.

One of the few examples of an L2 WA study in which a restricted association methodology was employed was published by Riegel et al. (1967). This research involved 24 native English speakers learning Spanish as an L2, and 24 native Spanish speakers learning English as an L2. Participants completed two sets of WA tasks – one in their first language, and one in their L2 – three weeks apart. Each set of tasks required participants to respond to 35 stimuli in accordance with seven sets of instructions each designed to elicit a different aspect of word knowledge. The English version of the instructions and examples of each type of association appear in Table 2.1.

Table 2.1. Instructions and examples for seven restricted association tasks in Riegel et al. (1967, pp. 537-538).

Instructions	Examples
Find a class-name for the stimulus.	*fork → silverware, utensil* *limousine → car, vehicle*
Name another member of the class to which the stimulus belongs.	*fork → spoon, knife* *car → train, bike*

Find a word that means essentially the same as the stimulus.	*fork → rake, branch* *car → auto, automobile*
Find a word that means essentially the opposite of the stimulus.	*fork → knife, spoon* *wisdom → foolishness, stupidity*
Find a word that denotes the usage of the stimulus.	*fork → eat, take-up* *car → travel, drive*
Find a word that denotes a quality of the stimulus.	*fork → pointed, heavy* *car → fast, shiny*
Name an essential part or attribute of the stimulus.	*fork → handle, metal* *wisdom → experience, maturity*

Results of the study showed, unsurprisingly, that subjects left more responses blank (i.e., null responses) when responding in a second language than in their native tongues. Of the seven types of restricted associations, providing a word that was synonymous to the cue proved to be the most difficult for subjects. In terms of the number of different responses, subjects provided approximately the same number regardless of the language in which they were responding (10.4 in Spanish; 9.9 in English). The greatest number of types of response was elicited by the "quality" instructions. This finding was attributed to the number of such words available in the target languages. Finally, as a measure of language proficiency, the researchers also examined *group overlap.* Put simply, this refers to a comparison of native speaker responses to those of L2 learners. By examining the number of L2 responses that were identical to those of the NS groups, the researchers concluded that the English speakers learning Spanish were less proficient than the Spanish-speaking subjects learning English as an L2.

This study is important to the current dissertation as it was one of very few studies (including Miron & Wolfe, 1964; Ramsey, 1981; Riegel & Zivian, 1972) to employ a

restricted association task. This underused method is central to the research I will present in Chapter 6. A review here of a number of the features of the Riegel et al. study will help to avoid potential pitfalls in the research design I will present later.

One of these features is the manner in which the overlap scores were interpreted. Essentially, the researchers' measures of overlap are numerical representations of nativelikeness based on the number of times NNS subjects provided answers mapping on to those of native speakers. I have discussed some of the issues surrounding the use of native norms in the context of other studies above (see Sections 2.3.1 and 2.3.2). Such use of normative data is by no means uncommon in WA research. Indeed, the canon of WA research includes a plethora of studies (e.g., Greidanus & Nienhuis, 2001; Namei, 2004; Racine, 2008; N. Schmitt, 1998b) in which such comparisons are made between learner responses and native associations for the purposes of measuring L2 proficiency or making claims about the organization of the developing lexicon . A key distinction, however, between most of the other studies in which native norms have been used as comparative measures and the current study (i.e., Riegel et al., 1967) is that, in the current study, associations were elicited by – and restricted by – researcher instruction. What I mean by this is that the underlying premise of virtually all WA research concerns responses elicited by cues *without deliberate consideration on the part of subjects*. I suggest, however, that the Riegel et al. (1967) study required subjects to adopt deliberate strategies for each of the tasks involved. For example, an NNS participant's inability to respond with a synonym when presented the cue *fork* may be due to a lack of knowledge of words like *sword* or *pointer.* Failing to generate that kind of response, the subject may adopt a default strategy of either not

responding, or perhaps responding with a word like *spoon*, despite the fact that the instructions say explicitly that this is an "incorrect" response – i.e., *spoon* is the opposite (sic) of *fork* (see Table 2.1). In other words, the instructed association method adds an extra dimension of uncertainty to researchers' assumptions about the processes underlying WA responses. L2 WA researchers have always assumed that – despite subjects potentially adopting a response strategy when a response didn't come immediately to mind – associations reveal connections within their mental lexicons. Differences in responses between NS and NNS respondents are thus said to indicate differences in language proficiency brought about by differences in word knowledge and lexical processing. In the case of the restricted association method adopted by Riegel et al., it is difficult to say what the data reveal about the mental lexicons of language learners. Response differences between NS and NNS may reflect differences in response to the tasks rather than anything about lexical processing at all.

My criticism here is that strong conclusions about the L2 lexicon based on restricted association data – one of the purposes of Riegel's study – may not be justified. This criticism aside, however, if not for its results and their interpretation, but for its methodology, this study was ahead of its time. As I will demonstrate in Chapter 6, used more carefully, restricted association methods do offer researchers a useful twist on the conventional WA methodology. Data gathered via this technique – as long as the implications thereof are not misinterpreted or overextended as they have here – are another useful WA method for language researchers. It should also be noted that while L2 WA studies are commonplace now, most WA research taking place at the time that Riegel and his colleagues

published these findings concerned the first language. This study is – along with the work of Lambert and his colleagues (Lambert, 1956; Lambert & Moore, 1966; Lambert & Rawlings, 1969) and only a few others (e.g., Davis & Wertheimer, 1967; Fishman & Cooper, 1969) – one of very few L2 WA studies published prior to 1970.

2.5 The research process: To be continued

In the attempt to illustrate aspects of the WA research model outlined in Figure 1.1, three broad areas of L2 lexical research have been covered in this chapter. Before turning to WA research, it was necessary to outline the models that have been proposed to explain the bilingual mental lexicon and its processes (Section 2.2). Illustrating the first stages of the WA research process, Section 2.3 focused upon the types of research to which WA methodology might be applied and the theories and assumptions that guide researchers' hypothesis formation and experimental design. Section 2.4 was concerned with the later stages of the WA research process, focusing primarily on the methodological aspects of WA research. While I have offered criticism of these prior studies, it is by no means intended as a condemnation of the specific works described above or more generally of the methods employed. On the contrary, my assessment of these studies is meant in the spirit of fine-tuning research methods that have much to offer. Indeed, the scrutiny paid to aspects of the studies described above helps to guide the rationale for the research I will present below.

Finally, it is hoped that this examination of WA methodology's role in L2 research has demonstrated the value and utility WA methodology has to offer, and that the

focus on specific stages of the research process has piqued my readers' interests. It may appear that this chapter has been overly brief, at the expense of further illuminating some of the issues inherent to some of the under-represented stages of the research process (e.g., cue selection and the categorization of responses; Stages 3 and 5 of the research process model presented in Figure 1.1 respectively). This is not to say that these other aspects of WA research are somehow less crucial to the process, or that this chapter is presumed to adequately depict the research process despite their absence. In fact, research illustrating stages of the process not covered in detail here will be scrutinized piecemeal in the chapters to follow. Their introduction later in the dissertation is intended to present them at points when their relevance to the research is at a zenith and the issues raised will contribute most to the discussion at hand.

Chapter 3. Replicating rabbits: Toward a comprehensive categorization of word association responses

3.1 Introduction

The dominant purpose of adopting word association methods in lexical research is to examine presumed links within the mental lexicon. As explained in the previous chapters, however, the WA methodology itself is deceptively simple. A typical WA trial (see Stage 4 of the WA research process in Figure 1.1) consists of subjects reading or hearing a word and responding (orally, or in writing) with the first word that comes to mind. I have described a number of ways in which other aspects of the WA research design may be adapted and combined with other methods to further our understanding of lexical processes. However, the basic cue-response format as initially applied to research in second language processing during the 1950s and 1960s (e.g., Kolers, 1963; Lambert, 1956) has undergone very little change since that time. Furthermore, as noted already, the same may be said about the manner in which WA responses are categorized (Stage 5 of the WA research process model). In the vast majority of WA studies to date, responses were classified by researchers into the following categories: paradigmatic responses, if a response belonged to the same word class as its stimulus; syntagmatic responses, if stimulus and response were from different word classes but might appear in a sequential relation in a sentence; and clang responses, if stimulus and response shared orthographic or phonological features in the absence of some kind of semantic or sequential relation. Responses that appeared not to fit into any of these were typically placed in an *other* category.

This response classification scheme has survived more than a half-century despite being rather problematic in actual practice. One issue that arises repeatedly in WA research is what to do with the many responses that are potentially classifiable into more than one category. *Pepper* in response to *salt* appears to be a paradigmatic response in that they are both nouns and both reside in the same lexical set. Yet at the same time they maintain a collocational relationship in the phrase "salt and pepper", and can thus be classified as a syntagmatic cue-response pair. As already discussed in Section 2.4.1, Fitzpatrick and Izura (2011) assign both links for this kind of response. It remains unclear, however, whether dual-links reflect the motivation for these responses.

Not only does the traditional paradigmatic/syntagmatic/clang classification present difficulties in assigning responses unambiguously to categories, the data yielded from this type of classification may not offer a particularly rich view of the connections between representations in the mental lexicon. Recent vocabulary research takes a more nuanced view of word knowledge. In Nation's (2001) model, for example, vocabulary knowledge is conceptualized as the product of as many as 18 ways in which we can know a word.[1] This contrasts sharply with the broad grammatical and sequential categories utilized in the traditional classification of

[1] While beyond the scope of this dissertation to go into detail about all of these types of word knowledge, it is important to note that Nation's taxonomy includes various types of formal knowledge (of a word's spoken and written forms, and its parts), semantic knowledge (e.g., the link between a word's meaning and form, and its concepts, referents, and associations), and knowledge of word use (its grammatical functions, occurrence in collocations, knowledge of register constraints, etc.) and that each of these may be represented as both receptive and productive knowledge. See Chapter 2 of Nation (2013) for a detailed examination of types of word knowledge.

word associations. A study by Fitzpatrick (2006) has attempted to modernize the way in which WA responses are categorized, introducing a detailed categorization scheme – to be described in Section 3.2 and analyzed thoroughly throughout this chapter – that is more in accordance with current notions of word knowledge.

In the context of the multifaceted nature of word knowledge depicted in vocabulary acquisition research more generally, Fitzpatrick (2006) observes that WA studies – in particular, L2 WA studies – have yielded unclear and inconsistent results. In particular, she notes differences between L1 and L2 response behaviours. She attributes this, in part, to the utilization of the traditional categorization scheme which places "artificial constraints on both association behaviour and the exploration of response types" (2006, p. 121), among other factors. In an attempt, then, to more precisely reveal the differences between L1 and L2 responses, Fitzpatrick designed a categorization scheme that is certainly among the most inventive to be implemented in a WA study. Based in part on Nation's (2001) taxonomy of word knowledge, she designed a 4-category classification of response types, further analysable into 17 subcategories. She also incorporated post-test interviews to help disambiguate responses based on participants' retrospective accounts of the association process.

Given the innovative nature of Fitzpatrick's response categorization scheme it is important to examine precisely how well it accounts for WA response types. It is equally important to examine how psychologically valid these categories are in the minds of subjects. As explained in Section 1.4, if a given categorization scheme is not psychologically valid, then its utility in identifying underlying lexical processes

will be compromised. One of the roles of the categorization scheme is thus to accurately identify the underlying processes presumed to give rise to the various types of WA responses. If the scheme does not function adequately in this regard, then it will be difficult to make strong assertions about the effects of other experimental manipulations (e.g., priming methods; to be introduced in Chapters 5 and 6) on WA findings. Likewise, the utility of post-test interviews to assist in the categorization of responses – a key element in Fitzpatrick's study – should be carefully considered. With these issues in question, Fitzpatrick's (2006) study is worthy of the type of analysis that a systematic replication[2] may provide.

A systematic replication can also begin to establish the extent to which her findings are generalizable to different populations of second language learners of English. As we know from Schmitt's (1998b) research reviewed above (see Section 2.2.1) the various aspects of word knowledge are not acquired simultaneously. In fact, Schmitt (2010, 2014) speculates that knowledge of associations may be among the last to be acquired. The implication, then, is that lower-level learners will not have acquired the same kind of word knowledge as have more advanced learners and thus the conclusions in Fitzpatrick's study may not hold true for less proficient

[2] In applied linguistics, a "systematic" replication may be defined as a study in which a single key variable from the original research (such as participants' proficiency levels or L1 backgrounds) is manipulated. In fact, the current study manipulated two variables rather than one. The subjects are from a different language background (i.e., Japanese students learning English in an EFL context, as opposed to English learners studying in the UK, in the original study) and are of a different level of proficiency (i.e., lower than those who participated in the original study). Despite this, *systematic* replication remains the most accurate label for the current study (see *Language Teaching* Review Panel, 2008 for definitions of other types of replication research).

learners. In this way, the replication presented here tests the generalizability of the original study's findings to a population of lower-level L2 learners.

Besides the issue of how generalizable these results may be to other populations of learners, there are also specific questions concerning Fitzpatrick's methodology that need to be examined. These include the extent to which post-test interviews are useful to WA researchers trying to disambiguate responses. This is an issue of practical concern as most WA research can be conducted with multiple respondents simultaneously (either with handwritten responses, or with series of computer terminals for typed/online responses). If there is substantial benefit to be gained through post-task interviews, they will have to be conducted one-on-one. This would add greatly to the time needed to collect and analyse data. For these reasons, post-test interviews will also be scrutinized here.

3.2 Reviewing rabbits: The original study (Fitzpatrick, 2006)

Fitzpatrick's (2006) study, entitled, "Habits and rabbits: Word associations and the L2 lexicon", involved 80 participants in total: 40 native speakers of English (NS), and 40 English language learners (non-native speakers; NNS) from a variety of first language (L1) backgrounds. The association stimuli consisted of 60 English words (see Table 3.1) selected from the Academic Word List (Coxhead, 2000). These cues were selected to represent a variety of word classes and frequencies.[3] The words

[3] Fitzpatrick's stated intention for selecting these particular stimulus words was to avoid the influence of specific frequencies and word classes on the results. However, she does not cite studies revealing what specific effect these might have. Similarly, very high-frequency items were eliminated from the study, but no rationale is offered for why it was important to eliminate them from this particular

were presented in written format and subjects were asked to respond by writing the first word that came to mind. Each subject participated individually such that, after all responses were collected, an interview could be conducted to more precisely classify ambiguous responses (i.e., those potentially belonging to more than one category and those seemingly belonging to none).

Table 3.1. Stimulus items from Fitzpatrick (2006).

1	consistent	21	integration	41	somewhat
2	environment	22	overall	42	voluntary
3	income	23	regime	43	chart
4	method	24	undertaken	44	detected
5	response	25	conflict	45	implicit
6	variables	26	equivalent	46	paragraph
7	commission	27	liberal	47	schedule
8	cultural	28	objective	48	visual
9	injury	29	stability	49	coincide
10	positive	30	whereas	50	distorted
11	resources	31	brief	51	manual
12	transfer	32	estate	52	portion
13	contribution	33	incentive	53	scenario
14	dominant	34	lecture	54	vision
15	instance	35	rational	55	colleagues
16	partnership	36	utility	56	encountered
17	sequence	37	confirmed	57	intrinsic
18	volume	38	eliminate	58	notwithstanding
19	commitment	39	hierarchical	59	posed
20	emerged	40	paradigm	60	whereby

As already mentioned, the responses were not categorized according to the traditional paradigmatic/syntagmatic/clang response categorization used in the majority of word association studies to date. Instead, Fitzpatrick developed a set of

study. Finally, concrete nouns were eliminated apparently on the grounds that they "tend to produce predictable responses and … are more likely to share a conceptual representation in the L1 and the L2 (Kroll and de Groot 1997)" (p. 128). Again, it is not made clear why these would be problematic for the current study.

categories and subcategories, and defined and labelled them in a more user-friendly manner than were the traditional categories. They were developed so they would be inclusive of the response types elicited in prior research (i.e., Meara & Fitzpatrick, 2000) and would account for the various types of word knowledge outlined by Nation (2001, p. 27). Responses were thus placed into three main categories: meaning-based (semantically related to their cues; similar to the traditional paradigmatic response category and accounting for most of Nation's *meaning* category), position-based (related to their cues via collocation; similar to syntagmatic responses and Nation's *use* category), and form-based associations (based on orthographical and/or phonological characteristics of the associates; similar to clang responses and Nation's *form* category). These categories were divided into 15 subcategories and an erratic associations category was added to account for responses based on false cognates of the stimulus or having no discernible connection to the stimulus. This classification, including definitions of the subcategories, as outlined in Table 1.1, is duplicated here as Table 3.2 for ease of reference.

Table 3.2. Fitzpatrick's (2006, p. 131) category/subcategory classification scheme for word association responses (*x* = stimulus word, *y* = response word).

Category	Subcategory	Definition
Meaning-based associations	Defining synonym	x has the same meaning as y
	Specific synonym	x can mean y in some specific contexts
	Hierarchical/lexical set relationship	x and y are in the same lexical set, are coordinates, or have a meronymous or super/subordinate relationship
	Quality association	y is a quality of x or x is a quality of y
	Context association	y provides a conceptual context for x

	Conceptual association	x and y have some other conceptual link
Position-based associations	Consecutive xy collocation or compound	y follows x directly (or with an article between them)
	Consecutive yx collocation or compound	x follows y directly (or with an article between them)
	Phrasal xy collocation	y follows x with a word (non-article) or words between them
	Phrasal yx collocation	x follows y with a word (non-article) or words between them
	Different word class collocation	y collocates with x + affix
Form-based associations	Derivational affix difference	y is x +/- a derivational affix
	Inflectional affix difference	y is x +/- an inflectional affix
	Similar form only	y looks or sounds similar to x, with no clear meaning link
	Similar form association	y is an associate of a word with a similar form to x
Erratic associations	False cognate	y is related to a false cognate of x in the L1
	No link	y has no decipherable link to x

Utilizing the new categorization, Fitzpatrick (2006) found that NS produced significantly more position-based associations than NNS, while NNS produced significantly more form-based responses than the NS group. In terms of the subcategories, NS produced significantly more defining synonyms and consecutive collocations than NNS. NNS produced more associations similar only in form to the stimuli, or which were associated through context or a loose conceptual link.[4]

[4] In trying to account for differences in response patterns between groups, Fitzpatrick raises an interesting question concerning ability vs. motivation (i.e., "whether non-native speaker subjects are unable to produce collocational responses to stimulus words or whether they are simply unwilling to", p. 138). She states that the results of the study do not yield answers to this question and does not speculate further. Indeed, respondent motivation has rarely been addressed in any discussion of WA research results to date. This is in spite of the fact that assumptions concerning this aspect of the associative process underpin decisions about how responses are classified. This is a topic I will return to again in the discussion below.

In accounting for the differences between subject groups described above, Fitzpatrick offers a number of important inferences. First of all, she describes differences between the lexicons of native speakers and those of learners not in terms of absences of lexical items, but in terms of weakness of connections between items in learners' lexicons. This is important for the current dissertation. The research throughout this thesis is concerned primarily with *processes* within the lexicon. Unlike much of lexical research where word knowledge is a dependent variable (e.g., tests of vocabulary size and depth) the currency of much contemporary WA research lies not in what may be revealed about word knowledge, but in what we can infer about lexical processes from these findings. Indeed, the inferences I will make from the data in this chapter and throughout the rest of this dissertation will primarily address presumed processes in the lexicon by which responses are produced.

Fitzpatrick goes on to account for the significantly large number of position-based (collocational) responses produced by native English-speaking subjects in terms of Wray's (2002, Chapter 11) model which contrasts the manner in which L1 and L2 phraseological units are acquired. Wray asserts that L1 phrases are holistically incorporated into the lexicon as chunks, while classroom learning of an L2 may first involve the breakdown of collocations depending on learners' expectations of how language needs to be internalized and on their beliefs about what constitutes manageable and reusable L2 input. This model, Fitzpatrick claims, accounts for the lack of correlation between proficiency and the number of collocational responses elicited from NNS informants. That is, more proficient L2 learners did not produce proportionately greater numbers of position-based responses than did lower-level

learners. Fitzpatrick's conclusion appears to be based on the assumption that an individual stimulus is more likely to elicit a collocate from NS than from NNS subjects. However, Wray's model does not necessarily support this view. As collocations are integrated wholly into the L1 lexicon as unanalysed chunks, individual words may remain invisible within such strings. For example, *once* is less likely to elicit *upon* (as in *once upon a time*) from NS respondents precisely because the string may have been stored unanalysed. If this is the case, for NS in particular, *upon* and *time* would remain relatively invisible within the wordstring and would not likely be elicited during WA trials. While Wray does not make specific claims about the activation of representations during word association, we can infer from this argument that the holistic manner in which collocations are stored in the L1 lexicon make individual elements less likely to elicit each other. Fitzpatrick assumes that the strength of these wordstring representations is the cause of elicitation of collocational associations between individual words in the strings. However, Wray's model actually states that it is the holistic nature of those representations and the invisibility of their individual elements that makes elicitation of collocational associations less likely in the case of NS subjects.[5]

Returning to Fitzpatrick's methodology: as one of the purposes of the research reported in this chapter is to examine the exclusivity and comprehensiveness of Fitzpatrick's response categories, it is important to critically examine the manner in which she classified responses. Let us begin with the *defining synonym*

[5] In fact, Wray (personal communication) states that Fitzpatrick's reasoning may have more in common with models of ballistic processing (see Favreau & Segalowitz, 1983; Segalowitz, 2010, Chapter 4) than it does with Wray's own model.

responses, part of the *meaning-based* supercategory. She provides a table of *defining synonym* responses to the stimulus *contribution* that includes, among obvious synonyms like *donation*, the following responses: *money*, *input*, *help*, *part*, *gift*, *payment*, *sharing*, and *add*. None of these items is actually fully synonymous with *contribution* and they are certainly not synonymous with each other. There is, in fact, a broad argument that no two words are truly synonymous (otherwise the words in question would need not exist simultaneously). If we are to consider the words listed above as synonymous with *contribution*, our judgment is entirely dependent upon the specific contexts in which the words are used. According to Fitzpatrick's own categorization scheme (Figure 3.2) – that makes these *specific* synonyms, rather than *defining* ones. However, Fitzpatrick had additional information from the interviews. It is possible that, when asked why they had responded to *contribution* with, for example, *money*, subjects replied that "a contribution is money". Accepted as is, this sounds like a defining synonym, but in the context of a longer interview, the subject may go on to say that they had imagined a church collection plate or a charity drive, both very specific contexts. Therefore, *money* once again would be a specific synonym. In this way, use of Fitzpatrick's detailed categorization scheme may present a number of difficulties in actual practice. One can imagine how interrater reliability could become a concern given the kinds of interpretation necessary to categorize certain responses. Such issues concerning response categorization may best be examined in light of the use of post-test interviews as a means of clarifying response types.

The implementation of post-study interviews to aid in the classification of associations and the introduction of the two-tiered categorization scheme are both

innovative aspects of Fitzpatrick's (2006) study. While WA researchers know well the difficulties in trying to determine which category a given response should be assigned to (e.g., as noted above, *salt-pepper* appears to be both a collocational and a meaning-based/paradigmatic response), very little has been done to rectify this issue. The utilization of "dual-link" categories (see Section 2.3.1) as in Fitzpatrick and Izura's (2011) study has already been mentioned. In Fitzpatrick's (2006) study, however, she took a different approach. Respondents were interviewed with the aim of determining precisely what they were thinking when they made their responses. This, at least on its surface, appears to be a reasonable means of distinguishing between competing categorizations, though I will argue later that it may not be effective.

The participants in Fitzpatrick's (2006) study were highly proficient learners in an ESL context. As explained in the discussion of models of the bilingual mental lexicon in the previous chapter, there may be reason to believe that the mental lexicons of ESL and EFL learners differ. These differences may consist of differences in strength between lexical and conceptual nodes (as in the RHM) or as differences between activation levels for each language in the lexicon (as in the BIA and BIA+). Such differences may manifest themselves in WA data as a variety of between-group differences. These may include differences in response types to specific cues or different response distributions across categories. Within-subjects differences may also be observed in stability of response types over time or over task. Taking these issues into account, along with those raised in the review above, it was deemed appropriate to conduct a systematic replication of the 2006 study.

3.2.1 New questions

In regards to the issue raised immediately above – concerning potential variance in response data due to differences in respondent proficiency – the replication was conducted to examine one substantive research question in particular:

1. Will Fitzpatrick's main findings be replicated with a population of less proficient learners of English in an EFL context?

In addition, a number of methodological questions are also raised by the original study. For example, as Fitzpatrick's is one of very few WA studies to incorporate post-test interviews, I am motivated to answer the following:

2. To what extent are post-test interviews useful in word association research? That is, do they aid researchers in disambiguating WA responses and if so, can they be implemented practically within the context of WA studies?

Also, as one of the key features of Fitzpatrick's study is the unveiling of a new categorization scheme of WA response types, one of the purposes of the replication is necessarily to address the question:

3. How comprehensively and inclusively do Fitzpatrick's response categories account for participants' responses?

3.3 The replication

3.3.1 Method

This study, like the original, involved two sets of participants: 40 native speakers (NS) of English and 40 non-native speakers (NNS). The NS group consisted of 1st-year students at a university in the United Kingdom. Unlike the original study, in which the L1 backgrounds of the non-native speakers consisted of as many as 10 different languages, the NNS group here were all native speakers of Japanese. The

group consisted of 2nd-year students and members of staff at a private university in Japan (none of whom were majoring in linguistics, nor were the staff already linguists). Each had achieved a score of at least 600 on the TOEIC (Test of English for International Communication). As a further measure of language proficiency, the Eurocentres Vocabulary Size Test (EVST; Meara & Jones, 1990) was administered. These scores appear in Table 3.3. For comparison, the scores from the original study (Fitzpatrick, 2006, p. 128) are also included.

Table 3.3. EVST scores of NNS subjects.

	n	Mean	Minimum	Maximum	SD
Current study	40	5520	2600	9800	1631
Fitzpatrick (2006)	40	6614	3550	9900	1660

As the table shows, the NNS group in the present study scored somewhat lower on average than did those in the original study. Fitzpatrick found that her NNS respondents had an average receptive vocabulary of 6,614 words (and averaged 6.6 on IELTS). The current group of NNS knew an average of 5,520 words as estimated by the EVST, and had achieved a mean score of 760 on the TOEIC. These differences are at least partly attributable to the increased vocabulary acquisition presumed to stem from an increase in exposure to the target language in the English as a second language (ESL) context – the UK – in which the original study was conducted. NNS in the current study were all students of English as a foreign language (EFL) in Japan.

The stimulus words in this study were identical to those chosen by Fitzpatrick in the original study (see Table 3.1): 60 words from the Academic Word List (Coxhead, 2000). These stimuli included 30 nouns, 13 adjectives, nine verbs, two

prepositions, one conjunction, one adverb, plus three words that could have been considered adjectives or nouns (e.g., manual), and one word that could have been interpreted as a verb or a noun (transfer). The use of a variety of word classes was intended to avoid the influence of a specific class or classes on response types (as observed by Deese (1962) and Sökmen (1993), among others). These particular cues were also selected to represent a broad range of frequencies, with the exception of the most frequent 2000 words in English (see Coxhead, 2000) which Fitzpatrick assumed would not elicit differences in responses between NS and NNS subjects (see Meara, 1982).

In line with the original experiment, participants took part individually. They each received a printed list of the stimulus items and were asked to respond to each by writing the first English word that came to mind in the space provided to the right of each word. There was no time limit for completing the task. As in the original study, most participants were able to complete this part of the procedure within 15 minutes. Immediately thereafter, respondents were interviewed in order to disambiguate responses for which categorization was unclear. The time required for this part of the procedure varied greatly depending on the number of ambiguous responses elicited from individual subjects and the ability of respondents – in particular, NNS participants – to articulate why they had responded as such.[6] Finally, the NNS respondents were asked to report their TOEIC scores and were asked to complete the EVST. Most NNS subjects took

[6] Note that the assessment of ambiguous responses occurred "in real time" during the interviews with each participant. An experienced WA researcher is able to scan the list and recognize immediately which responses do not readily fit into a single response category.

approximately 40 minutes to complete the entire procedure. Responses were categorized according to the 17-subcategory classification developed by Fitzpatrick (2006; see Table 3.2).

3.3.2 Results and discussion

3.3.2.1 NS vs. NNS and the generalizability of the original findings

To help examine the first research question (whether Fitzpatrick's findings generalize to NNS subjects at lower proficiency levels), response scores in the three main categories are represented in Figure 3.1. A cursory examination reveals a response pattern similar to that found by Fitzpatrick: the majority of responses are meaning-based for both subject groups while relatively few responses fall into the position- and form-based categories. It should be noted that the data presented in Figure 3.1, and throughout these analyses, are based on mean percentages, not actual means as Fitzpatrick reported. Although the same stimuli were used in both studies, the relative difficulty and infrequency of these academic English words for learners in Japan left many NNS respondents unable to respond to a variety of stimuli. In fact, 22 stimulus words failed to elicit a response from even 10 (of 40) NNS participants.

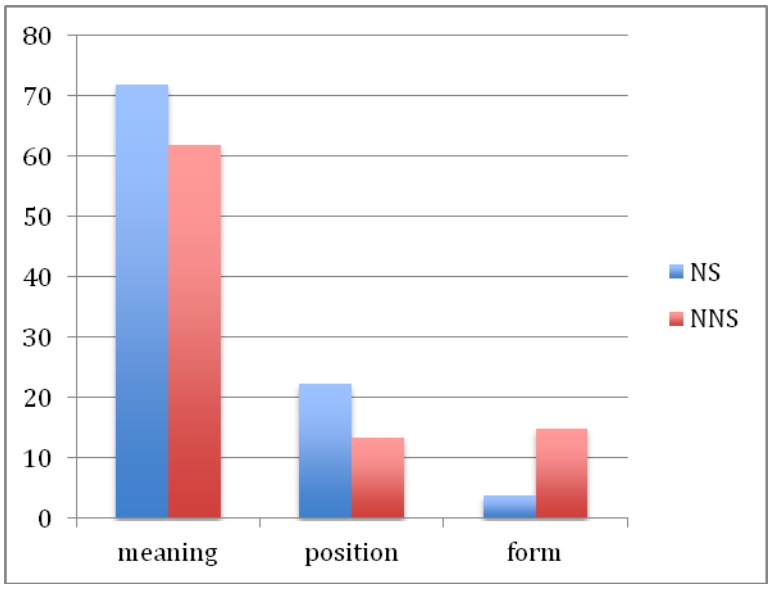


Figure 3.1. Mean response percentages per subject group.

It should also be noted that the majority of erratic responses was produced by the NNS group. In total, 9.8% of NNS responses were erratic, while only 1.7% of NS responses fell into this category. Fitzpatrick did not report her findings for the erratic category at all, possibly due to the small number of these responses produced by the subjects in her study (i.e., native speakers and relatively proficient NNS respondents). The highly elevated erratic response counts from NNS in this replication (almost six times those of NS) may reflect weak connections within the lexicons of learners, or a lack of word knowledge with regard to the academic words used as stimuli here, as explained above. Indeed this is a common problem in word association research involving lower-level learners (Higginbotham, 2010). When encountering stimuli they don't know, pressure to respond with "the first thing that comes to mind" may lead NNS respondents to reply without attempting to make even a formal connection to the stimuli in front of them.

Table 3.4 allows a more detailed examination of the results and reveals a number of noteworthy findings. For one, the differences in responses between groups (NS and NNS) in all three of these categories are significant: NS produced significantly more meaning- and position-based associations than did NNS. NNS produced significantly more form-based responses than did NS. The table also reveals that these findings are not in complete agreement with those of Fitzpatrick (2006): in this study, NS responded with significantly more meaning-based responses than did NNS, while Fitzpatrick found no significant difference (in fact, NS had produced fewer than NNS). However, the table also shows that position- and form-based responses followed the same pattern in both studies. NS produced significantly more position-based responses than NNS, while NNS produced significantly more form-based responses than NS. Generally speaking, then, the answer to the first research question is *yes*: Fitzpatrick's basic findings are generalizable to Japanese respondents. Her classification, at least as far as the main categories are concerned, does reveal differences between the response behaviours of NS and Japanese learners of English. An analysis of the differences in results within the subcategories appears below.

Table 3.4. Comparison of means between groups.

	Category	t	Group with higher mean
Current study	Meaning-based associations	2.531*	NS
	Position-based associations	2.443*	NS
	Form-based associations	4.676***	NNS
Fitzpatrick (2006)	Meaning-based associations	1.254	NNS
	Position-based associations	10.581***	NS
	Form-based associations	2.940**	NNS

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

The three broad categories described here (meaning-, position-, and form-based associations) coincide to a large extent with traditional response categories as defined in most word association research to date (paradigmatic, syntagmatic, and clang/orthographic responses, respectively). The fact, then, that NS produced significantly more meaning- and position-based responses than did NNS, provides support for the widely accepted finding that native speakers produce an abundance of paradigmatic responses while most clang responses are elicited from NNS subjects (e.g., Kudo & Thagard, 1999; H. B. Nissen & Henriksen, 2006; Orita, 2002; Söderman, 1993). At the same time, however, this renders the new scheme – at least at the supercategory level – as merely a relabelling of responses, rather than a complete reconceptualization.

It is important to note again, that the NNS subject group in this study were not as proficient as those in Fitzpatrick's study, having attained a receptive vocabulary of approximately 1100 fewer words, according to the results of the EVST (Table 3.3). Thus, certain effects that appear to be inconsistent between the two studies (e.g., meaning-based responses were elicited from NS significantly more often than from NNS in this study, while Fitzpatrick did not find this difference between groups) may be attributable to the difference in proficiency levels between the NNS groups. As a point of discussion, it is interesting to speculate about the meaning of these proficiency-based differences in response behaviour. In keeping with the WA research thread that examines responses in terms of lexical development (e.g., investigations into the syntagmatic-paradigmatic shift cited in the previous chapter), it is tempting to explain these differences in terms of a shift in response patterns brought about by reorganization within the expanded and deepened

bilingual lexicons of L2 learners. However, while the present findings provide clear support for the notion that NS response patterns differ from those of NNS subjects, it would be mere conjecture to infer that NNS response patterns undergo a shift towards nativelike responses as learners attain sufficient levels of language proficiency. Further research may determine whether a shift is actually occurring in this case, and if so, at what stage of lexical development (i.e., level of proficiency) it occurs.[7] I will re-examine the possible connection between nativelike responses and L2 proficiency in Chapter 4.

3.3.2.2 Subcategories

Table 3.5 yields a more precise view of the differences in responses between groups. It becomes clear here that only certain subcategories contribute to the main effects exhibited in Figure 3.1 and Table 3.4. For example, NS participants produced significantly more meaning-based responses overall. This is primarily due to an abundance of defining synonyms and specific synonyms. This was also in spite of the fact that NNS produced significantly more conceptual associations than the NS group. The high percentages of consecutive *xy* and *yx* collocative associations contributed to NS participants' production of significantly more position-based associations, despite the fact that NNS produced significantly more different word class associations (i.e., responses that collocate with a different grammatical form of the stimulus). NNS produced significantly more responses in

[7]In fact, research published within the last decade (including Fitzpatrick, 2007, 2009; Higginbotham, 2010; Racine et al., 2014; Zareva, 2010) is beginning to question the utility of examining second language learners' response data in terms of "nativelike-ness" at all. Indeed, there is a growing argument against making a distinction between native and non-native speakers and evidence of the potential harmful effects of these labels (e.g., Mauranen, 2011; Swan, Aboshiha, & Holliday, 2015).

all of the form-based subcategories except "inflectional affix difference" which were not produced by any of the participants in either group. The mean response percentages for these subcategories are represented graphically in Figures 3.2, 3.3, and 3.4.

Table 3.5. Comparison of means between groups (subcategories).

Category	Subcategory	*t*	Group with higher mean
Meaning-based associations	Defining synonym	2.681**	NS
	Specific synonym	4.409***	NS
	Hierarchical/lexical set relationship	1.517	NS
	Quality association	1.778	NS
	Context association	0.022	NNS
	Conceptual association	5.356***	NNS
Position-based associations	Consecutive *xy* collocation	2.450*	NS
	Consecutive *yx* collocation	2.276*	NS
	Phrasal *xy* collocation	0.561	NNS
	Phrasal *yx* collocation	0.248	NS
	Different word class collocation	2.530*	NNS
Form-based associations	Derivational affix difference	3.498***	NNS
	Inflectional affix difference	-	-
	Similar form only	2.012*	NNS
	Similar form association	4.384***	NNS

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

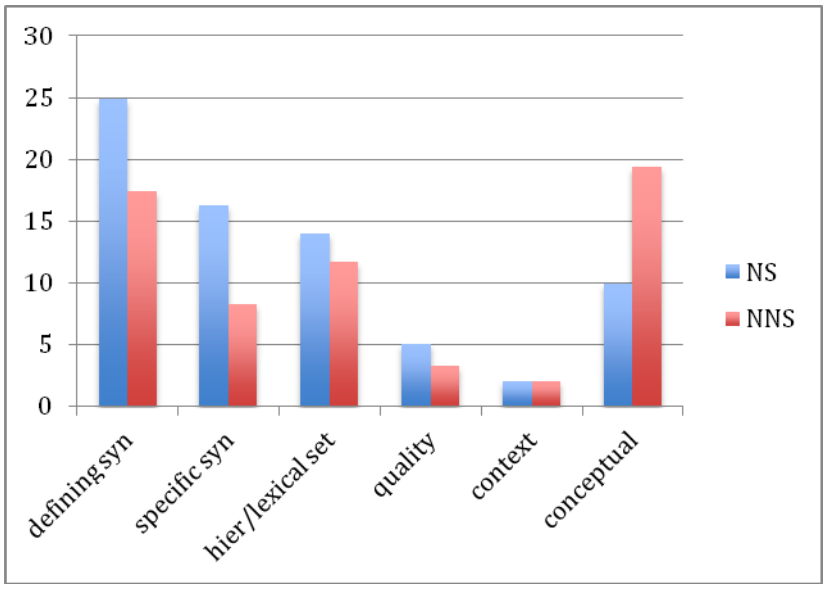


Figure 3.2. Mean response percentages for meaning-based association subcategories.

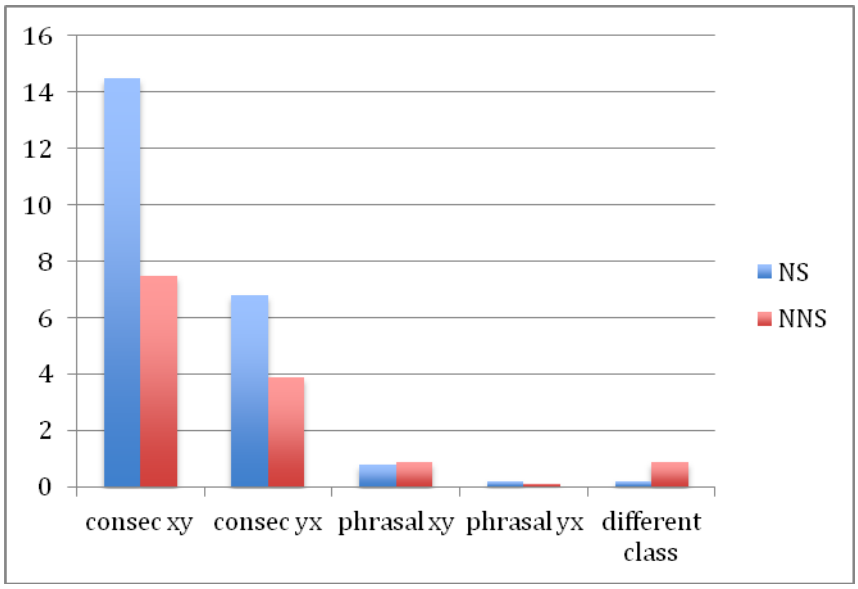


Figure 3.3. Mean response percentages for position-based association subcategories.

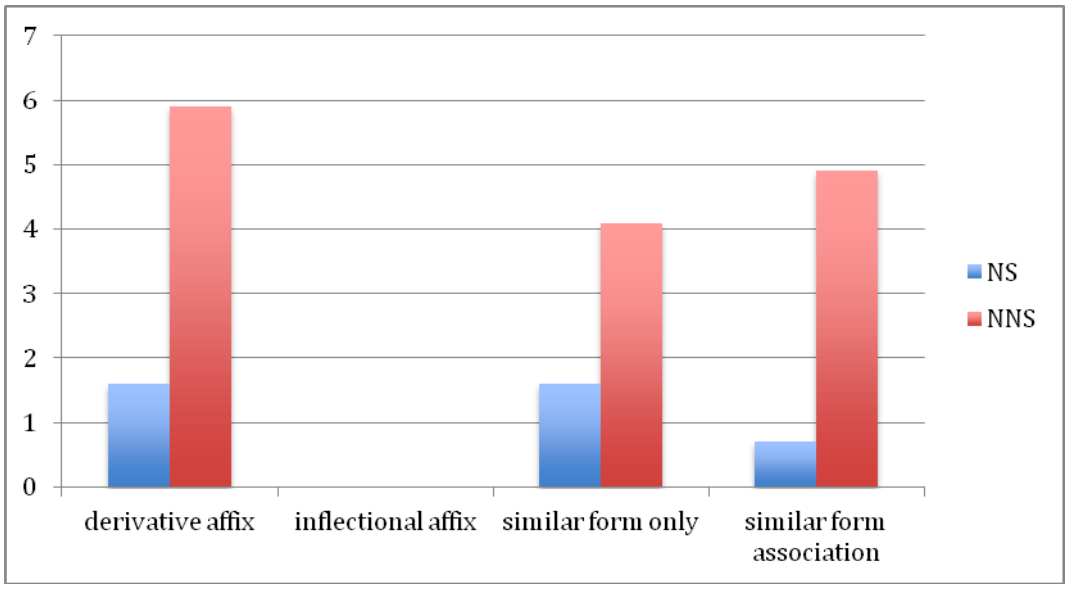


Figure 3.4. Mean response percentages for form-based association subcategories.

Table 3.6 displays a comparison of Fitzpatrick's findings with those of the current study. In 10 of the 15 subcategories there were no differences between the two studies in terms of the significance or direction of effect. These are listed in the centre column of Table 3.6 in the non-shaded cells. Specifically, in both studies, NS produce significantly more defining synonyms and consecutive collocations (both *xy* and *yx*) than did NNS. NNS produced more conceptual associations and similar form only associations. Five subcategories showed no significant differences (represented by the equals sign) between NS and NNS in either study (i.e., hierarchical/lexical set relationships, quality associations, both types of phrasal collocations, and inflectional affix differences). The five subcategories that revealed differences across studies are listed in the columns on the right side of the table in the non-shaded cells.

Table 3.6. Comparison of statistically significant effects across studies.

Category	Subcategory	Same effect in both studies	Current study	Fitzpatrick (2006)

Meaning-based associations	Defining synonym	NS > NNS		
	Specific synonym		NS > NNS	=
	Hierarchical/lexical set relationship	=		
	Quality association	=		
	Context association		=	NNS > NS
	Conceptual association	NNS > NS		
Position-based associations	Consecutive xy collocation	NS > NNS		
	Consecutive yx collocation	NS > NNS		
	Phrasal xy collocation	=		
	Phrasal yx collocation	=		
	Different word class collocation		NNS > NS	=
Form-based associations	Derivational affix difference		NNS > NS	=
	Inflectional affix difference	=		
	Similar form only	NNS > NS		
	Similar form association		NNS > NS	=

Note. The equals sign (=) denotes no significant differences between groups.

Overall, we can interpret Table 3.6 as indicating that the results of the replication were quite consistent with those of the original study. The results of the examination of subcategory response patterns answer the first research question in the affirmative: largely speaking, Fitzpatrick's findings do generalize to these Japanese learners of English, despite the fact that they had not attained the same levels of proficiency as the NNS subjects in the original study. None of the subcategories showed contradictory significant results across studies. Indeed, only five subcategories showed inconsistent results at all (i.e., where a significant effect was found in one study but not in the other). In the current study for example,

relatively few NNS responded with specific synonyms (a meaning-based subcategory) resulting in a significant difference in mean responses between subject groups. Similarly, NNS produced relatively large numbers of derivational affix differences and similar form associations (form-based) as well as different word class associations (position-based). Fitzpatrick (2006) found no significant differences between groups in any of these subcategories, as can be seen in the right-hand columns of Table 3.6.

There are a number of possible reasons for these few differences in the results across studies. First of all, they may be attributable to issues of inter-rater reliability. As noted in the examination of synonym categorization above, there are clear differences in the way Fitzpatrick coded responses and the way they have been dealt with here. A second means of accounting for differences in findings between these two studies is their statistical calculation. Many of these subcategories had very few responses in either study. A single response may have made the difference between significant and non-significant differences in some cases. Third, and perhaps most interestingly, these differences could reflect differences in learner proficiency. As noted above (Table 3.3), EVST scores showed a noticeable difference in proficiency. NNS subjects in the current study scored almost 1100 points lower than Fitzpatrick's NNS subjects. While this third possibility cannot be confirmed simply by comparing the results of these two studies alone, the observed pattern is consistent with previous research findings of a proficiency-driven shift from form-based to meaning-based responses.

A more direct means of measuring the possible link between language proficiency and response data may be found in correlation coefficients. However, Table 3.7 indicates that, with the exception of position-based responses, correlations between response types and proficiency scores were weak or nonexistent. Position-based associations showed a weak, but statistically significant, correlation with levels of receptive vocabulary knowledge (as assessed via the EVST). They also shared a moderate and highly significant correlation with TOEIC scores. The Fitzpatrick study also found a weak correlation between position-based responses and EVST scores (r = .30), though it did not quite reach significance (p < .055).

Table 3.7. Correlations (r) between response types and proficiency scores.

	Current study		Fitzpatrick (2006)
	EVST	TOEIC	EVST
Meaning-based association	.16	-.01	.24
Position-based association	.34*	.46**	.30
Form-based association	-.26	-.18	-.09

Note. * p < .05; ** p < .01

Table 3.8 shows the correlation coefficients between individual response subcategories and proficiency scores on the EVST and TOEIC tests. It is clear that in the replication consecutive collocations in particular have contributed to the significant correlation between position-based scores and levels of proficiency. Fitzpatrick's finding that a significant correlation may exist between proficiency and phrasal *xy* collocations was not replicated here. As with other differences in the results between these two studies, this finding may be accounted for by the differences in proficiency levels between the two NNS groups. It should also be noted again, however, that the small number of responses in this subcategory may

render these findings unreliable. While Fitzpatrick found a significant correlation between these phrasal collocation responses and language proficiency for NNS subjects, her findings were based on mean scores of only approximately one (1.0) phrasal *xy* collocation per subject.

Table 3.8. Correlations (*r*) between response types (subcategories) and proficiency scores.

		Current study		Fitzpatrick (2006)
		EVST	TOEIC	EVST
Meaning-based associations	Defining synonym	.25	-.01	.23
	Specific synonym	-.20	-.19	.20
	Hierarchical/lexical set	.12	.06	.19
	Quality association	.16	-.13	-.19
	Context association	-.18	-.03	.13
	Conceptual association	.01	.16	.22
Position-based associations	Consecutive *xy* collocation	.35*	.48**	.24
	Consecutive *yx* collocation	.24	.37*	.25
	Phrasal *xy* collocation	-.16	-.03	.36*
	Phrasal *yx* collocation	.24	-.06	.24
	Different word class assoc.	.04	-.18	-.01
Form-based associations	Deriv. affix difference	-.13	-.03	.13
	Inflectional affix difference	-	-	.26
	Similar form only	-.11	-.06	-.02
	Similar form association	-.28	-.26	-.46**

Note. * $p < .05$; ** $p < .01$

3.3.2.3 Post-test interviews

We turn next to the second research question, concerning post-test interviews and their utility in WA research. The interviews are intended to be a means of differentiating between categorizations of ambiguous responses, but I will argue below that the utility of such interviews is limited and that researchers may choose not to implement them for reasons of both inaccuracy and impracticality. Based on

the argument below then, the answer to the second research question (Are post-test interviews useful in word association research?) is *no*.

Fitzpatrick implemented post-test interviews as a means of disambiguating responses. Part of the reasoning behind utilizing the interview technique is that word association studies traditionally relied upon the intuitions of researchers to disambiguate associations that could potentially be classified into more than one response category. For example, a researcher may have considered a typical stimulus-response pair like *salt-pepper* to be semantic coordinates, and thus classified the response as paradigmatic. Another researcher might just as easily have regarded these as elements of the collocation "salt and pepper". Thus, the response would be considered a syntagmatic one. Without asking the respondents themselves, such responses could only be classified according to the intuitions of individual researchers and were typically based on relations between the grammatical classes of stimuli and responses. Post-test interviews, Fitzpatrick explains, yield a classification that "refers to the subject's own perspective of the link between words rather than to any external referent" (2006, p. 132).

My argument here (see also Racine, 2011b), however, is that "the subject's own perspective" may yield data that is no more accurate than the traditional intuitions of word association researchers. For introspection reports to accurately reveal connections between words in their mental lexicons, subjects must have conscious knowledge of the processes initiated during word association trials. By definition, however, associations are responses consisting of the "first thing that comes to mind". In most cases, respondents do not engage in a conscious and effortful

process to generate responses. Responses are retrieved automatically and via an unconscious process from memory. For that reason, introspection protocols inevitably reflect the subjective post hoc inferences of respondents, not unlike those traditionally made by WA researchers.

Not only are post-test interviews potentially inaccurate sources of information from which to disambiguate response types, they are particularly problematic in the case of NNS subjects. While conducting the replication study, it quickly became apparent that it is extremely difficult for language learners who have achieved only the proficiency levels reported here to articulate why they had responded the way they had. Even when only enquiring about responses that were difficult to categorize, NNS experimental sessions averaged 40 minutes in length. Given that each interview had to be conducted one-on-one, this became a particularly labour-intensive method. For practical purposes, then, many researchers may find it necessary to continue to rely on their own intuitions while coding response data.

A third issue is that post-test interviews may yield information that is just as difficult to interpret as the word association responses that they are intended to clarify. For example, one NNS respondent in the replication study gave the response *consistent* to the cue *coincide*. It was not clear if it was a form-based association (based on an orthographic or phonetic connection) or a conceptual association. During the interview (and after a long pause), she commented, "If something coincides, it is consistent with something." Are we to take her at her word that these two are synonymous in her lexicon, or has she simply been unable to articulate the type of conceptual connection she has in mind? Similarly, an NNS

respondent responded to *dominant* with *chance*, and later said that they have the "same meaning". To the respondent, the words were synonymous, but should the response be given that categorization. Without the interview information, they would have been counted as *erratic*, as no dictionary would suggest that *dominant* and *chance* are synonymous. The participant provides an explanation, suggesting the link is not erratic, but the formal criteria for allocating the categories would have to be overridden in order to adhere to the participant's unlikely proposal. This introduces a rather ad hoc approach to the defined criteria that is not scientifically desirable. Examples like these cast doubt on the utility of interview data as a tool for more accurately categorizing association responses.

3.3.2.4 Response categorization

The creation of any classification scheme – whether it be for word association responses or some other phenomena – requires due consideration of the issues of inclusiveness (Does the scheme adequately cover all types of the tokens to be classified?), exclusivity among categories (Based on the category definitions, might a given token be placed in more than one category?), and the relative specificity of each category (Should certain types be given a separate category of their own, or would it be more pertinent to group them together in a single, more encompassing category?). These questions are all aspects of the third research question concerning whether or not Fitzpatrick's categorization comprehensively and inclusively accounts for all possible WA responses. If we are to use WA as a means of examining processes within the lexicon, then we must have valid categories that cover all types of response. Without accurate categories to label the responses, we may be mistaken in any inferences we make about the processes that underlie

them. While Fitzpatrick's (2006) approach to WA response categorization is innovative, only a clear answer to these questions can determine the true utility of her response taxonomy for further research. Her categorization scheme does appear promising as a tool that will allow future researchers more precision in mapping the subtleties of word association responses. Indeed, the subcategory-level detail of the scheme will provide the structure for the conceptualization of responses as profiles (to be discussed in the next chapter). However, the findings from this replication study suggest that the scheme might be further improved by a number of refinements that I will suggest below.

By definition, an association is the "first thing that comes to mind". That is, for a response to be considered indicative of a connection between entries in the mental lexicon, responses should be automatic, not brought about by conscious deliberation. Therefore, any response for which a respondent can give an accurate explanation ("I was thinking about …") is not really an "association" at all. Further, in the case of truly automatic responses (in which respondents do not engage in conscious, effortful thinking about the cues and their responses to them), any introspective report of the process is simply a conscious effort to make sense of their response after the fact. In these cases, regardless of how respondents choose to word their post-test introspections, they should be read as, "I must have been thinking…." That is to say, respondents are engaging in approximately the same reasoning to determine response categorization as researchers traditionally have. Higginbotham (2014) has demonstrated that researchers' intuitions about how to classify ambiguous associations matched respondents' reports approximately 90% of the time. This does not mean that researchers are accessing participants'

deepest mental processes; it may mean that participants are not doing so. For all of these reasons, the utility of conducting (otherwise time-consuming) post-test interviews of respondents is certainly open to debate.

Perhaps it is the uncomfortable combination of innovation and imperfection that most spurs on scientific research. Indeed, Fitzpatrick's paper – like most other research in most disciplines of study – provides evidence of both of these qualities. Yes, the categorization scheme and the use of post-test interviews are quite innovative, but as I have argued above, for reasons inherent to the ambitions of the enquiry – more than design flaws, as such – the validity and utility of either has yet to be proven. Also, given the essential role that the categorization of responses plays in the WA research process (see Figure 2.1), this study deserves the kind of scrutiny that a replication study can provide.

Concerning meaning-based response categories, the treatment of defining synonyms and specific synonyms as separate subcategories has already been raised. Certainly an argument can be made that synonymous associates represent a unique type of word knowledge and hence warrant a specific category of their own. It is not clear from Fitzpatrick's account, however, why she further parses these responses into the separate defining and specific subcategories. As I have explained above, these subcategories, at least as far as they are defined here (see Table 3.2), do not provide unambiguous choices for the researcher. Indeed, it is not clear if there is any utility in making distinctions between these two response types at all. An argument can also be made that they are actually the same thing as far as their representation in the mental lexicons of respondents is concerned. This

is illustrated by revisiting the example of the stimulus *contribution*: if a subject were to respond with *money*, there are no grounds upon which to state that the response is not a defining synonym *if the respondent's own cognitive/conceptual context makes it so*. In other words, a post-test interview may reveal that the subject had a church offering plate in mind. In the context of that subject's own cognition then, *money* is a *defining* synonym and should be categorized as such. The point is that WA research is not conducted as a means of labelling connections between referents in the real world, nor should it be about the connections between words in a given language. Cue-response pairs are used as a means of uncovering connections in the mental lexicon and in this way WA research allows us to infer lexical processes.

In practice, use of the defining synonym category is problematic for other reasons as well. As I have explained above, some of the example responses that Fitzpatrick provides appear to be far from definitive. For example, Fitzpatrick categorized all of the following as defining synonym responses to the stimulus *contribution*: *money, input, help, part, payment, sharing*, and *add*. While *money* may be the very definition of a contribution in contexts like the offering plate, this most certainly is not the case where the currency is in time or effort. By Fitzpatrick's definition then, should *money* not be categorized as a specific synonym? Likewise, a payment may be a kind of contribution, but that would make *payment* a subordinate response and thus class it as a member of the hierarchical/lexical set relationship category. Fitzpatrick may claim that post-test interviews helped to classify such ambiguous responses. One quickly realizes, however, the sheer volume of responses that may reasonably be categorized in more than one group given the manner in which the

defining and specific synonym categories are defined here. Moreover, the use of post-task interviews in many cases will not aid in disambiguating responses. This, as I have argued above, is especially true in the case of lower-level language learners.

The context association subcategory, in which responses provide a "conceptual context" for their stimuli, is also somewhat problematic for three reasons. First of all, it is not clear from Fitzpatrick's definition of the category precisely what "conceptual" means here. Are responses that provide a physical context for their stimuli to be assigned to this category? It is not clear whether *university* as a response to either *colleagues* or *lecture* fits into this category or not. A second problem is that objects or concepts residing within a given context, by definition, exist within meronymous relations with that context. *Instance* may be seen as a part or aspect of *time* just as a lecture or one's colleagues may be seen as a part or aspect of a university. It is difficult to imagine precisely what kind of post-test report would help to distinguish these responses as either context associations or as further examples of hierarchical/lexical set associations.

A third issue concerning the contextual association category is the fact that it only includes words that provide context for the stimuli that elicit them, not the other way around. That is, *university* provides a context for the stimulus *lecture* and is placed in this subcategory. The response *sentence* to the same stimulus however does not fit the criteria for inclusion and may be placed in the lexical set subcategory as a meronym of *lecture*. The meronymous aspect of the relation between this stimulus and response appears to be more salient when the context is

provided by the stimulus and not vice versa. In any case, it seems somewhat arbitrary that "*x* provides a context for *y*" is not part of the definition of this subcategory while both "*y* is a quality of *x*" and "*x* is a quality of *y*" define the quality association category. Without a theoretical basis for adopting this asymmetrical definition and with ease of use for the researcher in mind, it may be simpler to include both the *xy* and *yx* context relations under the same subcategory.

Given the problematic nature of this category's definition, as well as its inability to predict differences between NS and NNS subject groups (at least in the present study), I will suggest below that researchers may find it more advantageous to classify these associations as part of another subcategory. In fact, Fitzpatrick appears to have reached the same conclusion, as the context subcategory does not appear in subsequent studies (2007, 2009).

This asymmetrical relationship between stimuli and responses is also seen in the *position-based, different word class* subcategory of collocation. Responses in this subcategory include those that collocate with an affixed version of the cue (i.e., *x* + affix); e.g., when *voluntary* elicits *group* and the post-test interview reveals that the subject was thinking of the expression *volunteer group*. On the other hand, this subcategory does not include responses that, with an affix, collocate with their stimuli. Thus, when *method* elicits *science* and the participant reports to have been thinking of *scientific method*, it is categorized separately from the example above (presumably as a *yx* consecutive collocation). Without a theoretical or linguistic reason for classifying these two stimulus-response pairs separately, and for the

sake of ease of use of the categorization scheme, these two types of response might best be treated as a single subcategory.

A more fundamental issue that arises from an examination of responses in the different word class collocation subcategory concerns what precisely constitutes a legitimate word association response (i.e., a response not classified as erratic). If the purpose of word association methodology is to uncover links between stimulus words and response words in the lexicons of respondents, surely responses that are not based on the stimuli themselves do not inform us about the phenomenon we are attempting to investigate. Thus, *voluntary-group* tells us about the connection the respondent makes between *voluntary* and *volunteer* as well as the connection made between *volunteer* and *group*. These two associations would not be classified together: the former is a form-based association belonging to the derivational affix difference subcategory; the latter is a consecutive collocation. In that light, it appears that *group* has no direct connection to its stimulus. While in this case the researcher will see the indirect link and possibly try to accommodate it somehow, in other cases, the response may be unclassifiable or labelled as *erratic* under the current scheme. For example, someone responding with *orchestra* to *sympathy* may have done so because of an association with *symphony*, but the analyst cannot accommodate this within the protocol of analysis. As these are potentially interesting response types, however, I will suggest below that a new category be created for these: mediated responses.

Certain types of form-based associations are also inadequately classified under the current scheme. For example, responses that fall into the inflectional affix

difference subcategory can only be elicited by noun and verb stimuli (e.g., *injury-injuries*; *coincide-coincided*). Although it is certainly the case that the word class of a given stimulus affects the likelihood that a response will belong to a given word class (e.g., Aitchison, 2003; H. B. Nissen & Henriksen, 2006), inclusion in a given response category should not be solely dependent upon the grammatical class of the stimuli. One can imagine research findings in which idiosyncratic stimulus selection (an overabundance or scarcity of nouns and verbs, for example) leads to radically divergent results between studies. That said, despite utilizing a majority of nouns and verbs as stimuli in these studies (39 of 60), very few inflectional affix difference responses were elicited at all. Indeed, not a single response was classified as such in the current study. Given then that this subcategory holds little utility in distinguishing between NS and NNS response patterns, and may in fact tell us more about researcher stimulus selection than about the mental lexicons of respondents, I suggest below that it be eliminated from future second-language WA research.

Another form-based association subcategory that should be reconsidered in further research is the similar form association subcategory. This group consists of responses that associate with words having a similar form to their stimuli. Examples from the current study include: *consistent-include*, where the respondent was thinking of *consist*; and *equivalent-lateral*, where the respondent was thinking of *equilateral*. In both of these cases, there are no form-based connections between stimulus and response. As I have argued above in relation to responses in the different word class collocation subcategory, these responses are not directly related to their stimuli and may, under the current scheme, be labelled as erratic

responses. As these too are responses that may prove useful for researchers attempting to make inferences about the associative process, I will suggest they are included in my proposed *mediated response* category.

Finally, there are a number of issues concerning the way erratic responses are categorized and defined. For one, a certain number of respondents invariably produce responses that offer commentary on the circumstances or context in which the research is actually taking place. Examples of this type of association include *implicit-hot* when a respondent felt the room in which the research was taking place was too warm. Other associations have a semantic relation to the stimulus words themselves rather than the concepts that underlie the words. These include responses like *nothing* when "nothing came to mind" on seeing the cue, as well as associations like *incentive-difficult* "because it's a difficult word". Some subjects produce idiosyncratic responses with seemingly no connection to their stimuli, but reply that they had learned the two terms at the same time. These meta-associations are certainly meaning-based in the cognitions of the subject, but are not actually related to the meaning of the stimuli themselves. These responses are categorized as erratic under the current scheme, yet they offer a unique perspective on the associative process. I will suggest below that they be categorized as a separate subcategory in further research.

False cognates, one of the two erratic subcategories, include any response related to a false cognate of the stimulus in the first language of the respondent. Just as I have argued that response categories should not be dependent upon the specific grammatical classes of stimuli, neither should they be related to the first language

of the respondents. By definition, responses in this category can only be produced by nonnative speakers of the target language. Defined in this way, this category does not allow researchers to make comparisons between the responses of NS and NNS participants. Furthermore, response rates in this category do not allow comparisons between NNS subjects from differing L1 backgrounds. There are far fewer English-related cognates in the Japanese language, for example, than in the European languages spoken as a first language by some of Fitzpatrick's subjects in the original study. Also, as a number of the NNS respondents in the replication were studying third languages (French and German), they produced associations related to cognates in those other languages that could not be included in this category (i.e., responses related to false cognates in a third language). More usefully, then, this category should be expanded to include connections to third language cognates, as well as responses made in any non-target language.

3.4 Further research

3.4.1 The 2009 study

In the years that followed the study upon which this replication is based, Fitzpatrick returned to word association research (2007, 2009) utilizing a slightly modified version of the categorization scheme described above. This new scheme consisted of the same four major categories as the original, but included only 11 subcategories. This more refined version of the 2006 classification (i.e., from Fitzpatrick, 2009) appears in Appendix I. Modifications to the original (2006) scheme include the following:

- lexical set relationships and context relationships are now classified together.
- the phrasal collocation subcategories have been combined into the *other collocational association* category.
- derivational and inflectional affix differences are now classified together as *change of affix*.
- *similar form only* and *similar form association* are now classified together as *similar form not meaning*.
- the *erratic* category is now called *others* and includes a *blank* subcategory for non-responses.
- the *no link* subcategory is now the *erratic associations* subcategory. (In fact, the no link and blank subcategories were not distinguished in the 2007 study.)
- there are no *quality association*, *different word class collocation*, or *false cognate* subcategories

Fitzpatrick does not provide an explanation in either paper (i.e., 2007 or 2009) for why she made these revisions, but it may be that she had recognized concerns about the original categorization that are similar to those I have raised above. Two of these issues in particular are addressed well by the modified scheme detailed in Appendix I. Firstly, context associations have been combined with lexical set relations and meronyms. As I have argued above, the context association subcategory was problematic for three reasons: 1) the subcategory's definition was somewhat ambiguous, 2) contextual relations are, by their very nature, meronymous, and 3) the seemingly arbitrary decision to include only asymmetrical relations between stimulus and response in the definition (i.e., "*y* provides a context for *x*", but not vice versa). Combining the contextual and lexical set subcategories in the new categorization logically places contextual and meronymous relations together and successfully deals with the criticisms I have aired above.

The second issue to be addressed (albeit somewhat less elegantly than the manner in which problems with contextual associations were addressed above) concerns the different word class collocation subcategory. As I have indicated above, this category, somewhat arbitrarily, does not include collocations where the stimulus collocates with (*y* + affix), only those where the response collocates with (*x* + affix). More critically, this type of response is not actually a direct – or even a phrasal – collocation. In the case of *voluntary-group*, for example, the association is mediated by *volunteer* making it a form-based + position-based associative chain. Therefore, while I agree that the newer classification scheme rightly discards the different word class collocation subcategory, I do not feel that these types of response are best accounted for as members of the new 'other collocational association' subcategory. These stimulus-response pairs are clearly mediated by other links, and thus represent rather valuable data for the researcher. Since these are by no means typical collocational associates, I will argue below that a new subcategory should be created for classifying such responses.

A third issue addressed in Fitpatrick's newer scheme involves the inflectional affix difference subcategory. I have argued that responses in this category are too dependent upon the word class of stimuli and ultimately reveal little to help distinguish between the lexicons of native speakers and those of second language learners of English. Quite rightly, Fitzpatrick has combined this with derivational affix associations in her subsequent classifications. The utility of the new change of affix subcategory for distinguishing between NS and NNS responses, however, will have to be tested in future research.

3.4.2 A modified categorization scheme

While the 2007 and 2009 versions of the categorization scheme adequately deal with the issues described above, there are other criticisms I have raised that have yet to be addressed. Firstly, the definitions of the two synonym categories remain problematic. As I have illustrated in the example of the stimulus *contribution* above, in some cases the distinction between defining synonyms and specific synonyms is not clear. In fact, it is not clear that making such a distinction is warranted at all when respondents' cognitive contexts are taken into account. I have also argued that many specific synonym associations are in fact subordinate associations, representing examples of the stimulus. For all of these reasons then, I suggest that the synonym subcategories be combined into a single synonym subcategory in further studies and that researchers remain wary of the fine distinction between somewhat synonymous stimulus-response pairs and their potentially super-/subordinate relationship.

Another issue that has not been adequately addressed in the revised schemes is the *false cognates* subcategory (where *y* is related to a false L1 cognate of *x*). Indeed, the omission of this subcategory from the revised versions renders them somewhat more inexact than Fitzpatrick's original scheme. I have argued that this category should include all responses based on foreign language cognates regardless of the first language of the respondents. I also suggest that this category include any response given in a non-target language. Given the absence of this subcategory in the revised schemes, we may infer that Fitzpatrick feels these types of response can be included with erratic responses. While I agree that they should not be included among the three main categories (meaning-, position-, or form-

based associations), these types of response do not fit the definition of typical erratic responses. That is, they do in fact share discernible connections with their stimuli. As these include semantically or formally related foreign words, and responses mediated by semantically or formally related cognitions of foreign language cognates, as a group they fall outside of the scope of most monolingual WA research. For that reason, I suggest that foreign associates be included as a separate subcategory within the erratic response category.

Along with the *foreign associates* subcategory I've argued for above, I will suggest here two more subcategories to be included in the *other* supercategory that may prove useful in classifying word association responses in further research: mediated responses and meta-associations. It is hoped that their inclusion will provide greater clarity in making sense of response data and for distinguishing between NS and NNS response patterns.

The *mediated responses* subcategory would include all responses in which intervening cognitive steps ultimately account for the responses. This indirect relation between stimulus and response is the defining feature of responses in this category. Examples I have described above include: *voluntary-group* (where the intervening representation is *volunteer*), *consistent-include* (mediated by *consist*), and *equivalent-lateral* (where the intervening representation is *equilateral*). These responses are clearly not erratic, in the sense of having "no decipherable link". Instead there is a significant connection between stimulus and response mediated by one or more intervening cognitions. Future research that includes the classification of these responses may unravel the nature and importance of the

mediating "responses" to the associative process and may evoke a number of new research questions. For this reason, I suggest that a *mediated responses* subcategory be included as part of the *erratic associations* category in future research.

An additional subcategory to include in the *erratic associations* category is meta-associations. Unlike typical erratic associations, these responses are not necessarily indecipherably linked to their stimuli. Indeed, a response of *difficult* to a difficult word, or *nothing* when nothing comes to mind, have obvious, meaningful connections for the respondent. It is the fact that the subject is responding to his or her reaction to the stimulus, rather than responding to the stimulus itself that distinguishes these from more typical response behaviours. Also to be included in this subcategory are responses such as *hungry*, *hot*, or *tired* which may be unrelated to their respective stimuli, but reflect the context in which the respondent is completing the WA task.

Taking into account all of the issues I have raised above, I wish to present a modified 12-subcategory classification scheme to be utilized in further research of this type. This appears in Appendix II and reflects the following:

- a single subcategory for synonyms that includes responses with the same or similar meaning to their cues, and words that may be synonymous but only in certain contexts
- a *lexical and contextual set* subcategory that subsumes *meronymous, qualitative, and contextual relationships*
- a single subcategory for all phrasal collocations
- a single subcategory for all types of affixes
- subcategories for mediated and foreign language responses, as well as meta-associations within the *others* category

It must be conceded that the modified classification scheme presented here, just like Fitzpatrick's original scheme, may still prove difficult to implement in practice. The burden of accurately classifying a given response into one of 12 subcategories rather than the simpler, traditional four-category system is no less than enormous given the thousands of responses that require analysis in a typical second-language WA study. For this reason, researchers should recognize that they still have the option of categorizing responses into the four main categories alone when necessary. Moving back and forth from the specific categorization scheme to the general one offers the benefits of having a particularly detailed classification, yet saving time when such accuracy is not necessary. For example, in a study designed to investigate the extent to which members of binomial pairs are associated, it might be important to examine differences between *xy* collocations and *yx* collocations. *Spick* probably elicits *span* far more often than *span* elicits *spick* but evidence could not be assembled without separate subcategories for *xy* and *yx* collocations. For other studies, it may be simpler to treat all collocational responses as members of a single larger category. Indeed, given the small number of responses appearing in some of the subcategories (e.g., the different types of *others*), it may sometimes prove valuable for researchers to reserve the right to continue grouping responses according to broader categories than those proposed in the modified scheme presented here.

Finally, despite the concerns I have raised and the suggestions I have made for modifying future versions of the classification scheme, it is important to note that almost without exception, WA studies published prior to Fitzpatrick's (2006) paper relied on the traditional paradigmatic-syntagmatic-clang categorization.

Despite the long history of research in which this traditional classification was implemented, researchers have been adopting Fitzpatrick's scheme in their own research (e.g., Higginbotham, 2010) citing its potential to more accurately map the differences between the associations of native speakers and second language learners. It was in the spirit of fine-tuning a system that already makes a significant contribution to WA research that these criticisms have been raised and that the modified categorization scheme in Appendix II has been proposed.

3.5 Summary and conclusion

The replication study of Fitzpatrick's (2006) pioneering word association paper that employed a non-traditional response classification scheme attempted to address three research questions: 1) whether Fitzpatrick's findings would generalize to Japanese NNS respondents with lower levels of L2 proficiency, 2) how useful post-test interviews would prove in this type of research, and 3) how comprehensively Fitzpatrick's response categories would account for participants' responses. Overall, the results showed that Fitzpatrick's findings were generalizable to Japanese NNS respondents as identical effects were found in 10 of the 15 subcategories across studies. With regard to Question 2, I suggested that the impracticality of conducting post-test interviews and the inaccuracy of data elicited thereby, render them less than useful as a means of disambiguating WA responses. In fact, recent studies (e.g., Higginbotham, 2014) comparing researchers' intuitions with the results of post-test interviews showed that in the overwhelming majority of WA trials, researchers classified ambiguous responses in accordance with categorizations determined by post-test interviews. This

provides more evidence against the necessity of post-test interviews for disambiguating WA responses.

As for the response categorization scheme scrutinized here, the categories proposed in the original study, along with revisions to them in subsequent studies (Fitzpatrick, 2007, 2009) adequately accounted for the majority of data in this replication. This scheme is particularly useful in its flexibility. Researchers may choose not to employ the subcategory-level classifications depending on the focus of their studies, the specificity with which they need to examine the responses, and the number of responses elicited. However, I have proposed that three additional subcategories (foreign associates, mediated responses, and meta-associations) be included to better account for otherwise erratic responses. Like all subcategories, these may or not be necessary, depending on a researcher's area of study. Indeed, interested researchers may wish to reclassify these as categories unto themselves, separating them from the *other* category.

Finally, I have argued that the distinction between defining and specific synonyms is a false one. At the heart of this argument is the premise that for a WA categorization scheme to have utility for researchers attempting to infer the cognitive processes underlying associative links, the categories must be psychologically valid. As I have discussed in regards to models of the bilingual lexicon, psychological validity may be seen as the presence of relevant conceptual knowledge and links to lexical items. Thus, my argument is that, if respondents are responding with the first thing that comes to mind, there is no difference between the kinds of cognitive processes that may mediate *contribution-money* for one

subject and mediate *contribution-idea* for another. In the context of charities and work projects respectively, these responses may represent the very definition of synonyms. Examining this example in terms of the mental lexicon then, the BIA model would show substantial overlap between the concepts of *contribution* and *money* for the former respondent, and between the concepts of *contribution* and *idea* for the latter. In this sense, we can say that the links uncovered in these WA trials represent psychologically valid synonyms for these respondents and should be categorized accordingly.

One of the goals for WA researchers going forward, then, is that the classification of responses should reflect the psychological validity of each response for each individual respondent. Indeed, if response categories are unrelated to specific underlying processes and cognitions then there may be no utility in classifying responses at all. Likewise, what are we to make of differences in response patterns between subjects and across conditions if researchers cannot say that response categories capture something that is psychologically real for respondents? Also, psychologically valid categorizations are necessary to posit the notion of response preferences or "cognitive styles" (see Section 1.5) as reflected in response profiles. For all of these reasons then, we can see the critical role that response categorization plays in the WA research process (see Figure 2.1). In the next chapter, another essential aspect of the research process – the interpretation and conceptualization of response data – will be examined in detail.

Chapter 4. Profiles and proximities: The utility and validity of conceptualizing word association data as subject profiles

4.1 Introduction: From categories to profiles

The previous chapter presented a replication of Fitzpatrick's (2006) study in which she presented, for the first time, a two-tiered categorization scheme of WA responses. Among other findings, the categorization scheme, with some modification (see Appendix II), was found to comprehensively account for the responses of Japanese NNS respondents. This was despite the differences in L1 backgrounds and proficiency levels between the NNS groups in the original study and the replication. Among the modifications made to the original scheme at the point of analysis was the unification of the defining synonym and specific synonym category. The argument for this is that from the perspective of respondents' own lexicons, there is no difference between the two. Whatever the conceptual or cognitive context from which these responses are derived, the responses should be treated as synonymous with their cues. As I have argued, specific and defining synonyms cannot be separated from the perspective of the participant's conceptual frame. In this way, we may say that the modified categorization scheme in Appendix II represents a slightly more psychologically valid classification than the original. In precisely what way and to what degree we can say the modified scheme is psychologically valid is an issue I will return to in Chapter 7.

Having reached a point where the categorization scheme presented in Appendix II can be viewed as adequate for the tasks ahead, we can now turn to another issue

arising from Fitzpatrick's research: the utility and validity of interpreting WA responses as participant profiles.

4.2 WA responses as subject profiles

In another WA study, Fitzpatrick (2009) utilized a slightly modified version of the categorization scheme addressed in the previous chapter (see Appendix I). This time, her intention was to address some of the fundamental assumptions underlying L2 WA research. Among these is the notion that there are consistent, systematic differences between the responses of NS and NNS subjects and that these differences reflect the development of the L2 lexicon. Fitzpatrick cites an array of studies in which WA data was used in precisely this manner (e.g., N. Schmitt, 1998a; Söderman, 1993; Sökmen, 1993), but also points out that her own study (2006) failed to show a significant correlation between L2 proficiency and learner response profiles. That is, "learners did not appear to be moving towards more native-like response behaviour as their L2 proficiency increased" (Fitzpatrick, 2009, p. 43). A subsequent study also indicated that in NS subjects, responses varied considerably between subjects (i.e., within groups), yet remained consistent over time within individual subjects (Fitzpatrick, 2007). The 2009 study was thus conducted to determine if response-type preferences like these could also be observed in L2 participants.

At the same time, she wanted to determine whether, and to what extent, people's L2 preferences begin to reflect their own L1 preferences as their L2 proficiency increased. It is these two goals that grant the study particular relevance to the

current dissertation. In both cases, Fitzpatrick has conceptualized her data in terms of respondent "profiles". The conceptualization or interpretation of response data is an inextricable step in the WA research process (Figure 2.1, Step 6). It is interesting to note that in Fitzpatrick's (2006) study – replicated in the previous chapter – the term *profile* is not mentioned once. Indeed, I can find no evidence of WA responses ever being depicted as such in the literature prior to Fitzpatrick's 2007 and 2009 studies (the latter of which will be reviewed below). WA responses in all prior WA research are conceptualized as discrete points of data (typically, to be categorized), but the notion that together they constitute an individual sketch of a subject's response preferences is only a recent invention.

4.3 Association profiles in a second language: Fitzpatrick (2009)

Fitzpatrick's (2009) study involved administering two 100-cue WA tasks to a group of 37 native English speakers learning Welsh as a second language. Subjects rated their own Welsh language abilities on a 6-point scale from 'limited' (1) to 'expert' (6). As presented in Table 4.1, all but one respondent (of 37) claimed to be at least 'competent' in Welsh, and roughly half of the respondents (18) rated themselves as expert users. The 100 English cues were selected from the AWL (Coxhead, 2000) and the Welsh cues were translations of different AWL-words, but matched for frequency and word class with the English words. Participants were to respond to the English cues in English and, at a separate sitting, respond to the Welsh cues with Welsh words. A native speaker of Welsh was trained to categorize the Welsh responses and Fitzpatrick (a native English speaker)

categorized the English responses herself. This yielded what Fitzpatrick calls first- and second-language WA "profiles" for each participant.

Table 4.1. Self-ratings of Welsh proficiency (from Fitzpatrick, 2009, p. 45).

User level	No. of respondents
1. Limited	0
2. Modest	1
3. Competent	3
4. Good	8
5. Very good	7
6. Expert	18

In terms of the distribution of responses across categories, the results were broadly similar to prior studies (Fitzpatrick, 2006, 2007). Importantly, Fitzpatrick found high standard deviation scores within subcategories, indicating a great deal of variance within them. This accords with prior findings that, even in the case of NS respondents, responses are highly heterogeneous within groups. However, by applying a measure of Euclidian distance, Fitzpatrick was able to show that "the distance between an individual's Welsh and English profiles is significantly smaller than the mean distance between their Welsh profile and the other subjects' English profiles" (Fitzpatrick, 2009, p. 49). In other words, NNS respondents' L2 profiles were becoming more like their own L1 profiles, rather than native speakers' profiles as their L2 proficiency increased. Fitzpatrick concluded that the notion of native norms is "misleading" and that "we should look, instead, for learners' word association behaviours to move towards their own, individual, L1 behaviour as proficiency increases" (p. 51).

Like the 2006 study reviewed in Chapter 3, this paper represents a new approach to second language WA research. The appropriateness of treating WA responses as individual "profiles" (first introduced in Fitzpatrick, 2007) is borne out again here in the contrast between the stability of responses within individual subjects and the high standard deviation scores within subject groups. The utilization of proximity/distance measures was also unique among WA studies. The results called into question the hypothesis that non-native association responses become more like those of native speakers with increased second language proficiency. This finding – along with recent arguments against the use of native-speaker norms in WA research (e.g., Racine et al., 2014) – should compel researchers to re-examine the notion that native norms should serve as a standard by which NNS proficiency is to be measured. Indeed, it becomes necessary to reconsider whether there are reasons to consider native responses 'normative' at all.

Despite this intriguing finding and the innovative methodology that yielded it, there are a number of aspects of the study that should be given further consideration. One important aspect is the generalizability of its findings. This study was conducted with native speakers of English learning Welsh while living in Wales. It is fair to say, however, that the role of the Welsh language in Wales is unique among native languages. The number of communities in which the majority of the population speaks the language natively is declining. At the same time, however, the government continues to make efforts to maintain the presence of Welsh in the daily lives of residents. Among many other measures, there are bilingual street signs and public announcements, and a dedicated all-Welsh television channel. Fitzpatrick acknowledges that Wales is ostensibly a bilingual

society in which the Welsh language holds an "interesting status" and that, for this reason, it is "difficult to categorise as a first or second language for an individual speaker" (Fitzpatrick, 2009, p. 45). This raises the question as to precisely who the subjects are in this study. If it is as difficult as Fitzpatrick says to distinguish which language is a resident's L1 and which is the L2, then we might wonder if we can distinguish some of her proficient NNS subjects from NS subjects at all.

The language status of Welsh also raises the question of how generalizable these findings may be to L2 learners in other contexts. The status of Welsh in Wales may not be entirely unique in the world (cf. Catalan, Basque, Irish, etc.). Indeed, one can imagine the results here being pertinent to learners of officially recognized languages in English-dominant societies (e.g., French learners in predominately English-speaking parts of Canada). However, it is difficult to know how generalizable they would be beyond that. Welsh learners in Wales are not in a context typical of either EFL or ESL learners.

Another concern for this study is the manner in which the participants' language abilities were assessed. Firstly, subjects were ranked on the basis of their own self-assessments. This would certainly have introduced inaccuracy and perhaps bias into the study. Second, and more importantly, the fact that proficiency is measured as a discrete variable on only a 6-point scale may also be problematic in a study where proficiency is an essential element in at least one of the key hypotheses. A more refined measure of proficiency might yield more reliable data. As Table 4.1 shows, the self-assessments by the 37 participants are distributed over only a 5-point range (i.e., user levels 2 through 6). Moreover, half of the subjects (18 of 37,

specifically) evaluated themselves at level 6. This is problematic in terms of Fitzpatrick's own observation that in Wales it is difficult distinguish L1 from L2. One wonders if some of her most proficient NNS subjects might better be classified as NS, and whether they would in fact be classified that way by a more objective measure of language proficiency.

While the notion that individuals' L2 association profiles might become more like their L1 profiles with increased proficiency is an intriguing one, it needs to be re-evaluated in light of the concerns I have raised here. Indeed, one of the goals of this dissertation is to thoroughly investigate the validity of the *profiles* concept (see Section 1.5). In an attempt to address the concerns raised above and to examine the validity of the profiles concept, a partial replication of this study is presented below. Specifically, the replication is located outside the context of Welsh language learning and employs a more accurate and detailed assessment of subjects' L2 proficiency levels.

4.4 Getting the most out of the profiles concept: Cognitive styles

Before describing the replication, it is important to examine again precisely what a *profile* is. Fitzpatrick's initial use of "profiles" (2007), to describe individuals' response preferences, was in accordance with how the word might be used in common parlance. That is, response profiles as she described them are sets of data, potentially expressed in graphic form, that outline the details or features of subjects' responses. Indeed, it appears that the examination of profiles in graphic form (e.g., Figure 1.1) reveals the "shape" of the profile most easily. However, it is

actually the level of detail in the response categorization scheme (i.e., the presence of numerous subcategories) that yields the detailed "sketch" of how subjects have responded. Thus, in accordance with Fitzpatrick's definition, the collection and classification of a sufficiently large number of an individual's WA responses – distributed over a sufficiently detailed categorization scheme – will yield a profile. The research implications of interpreting WA data in this way are quite profound. As I have suggested in Section 1.5, a profile may be greater than the sum of its parts.

Fitzpatrick's "profile" is thus the quantitatively-driven outcome of a detailed categorization of an individual's WA responses. It is a potentially useful means of depicting or visualizing association data, rather than a reconceptualization of it.[1] In subsequent research, however, response profiles, and particularly how they are interpreted, appear to be undergoing a degree of conceptual drift. Higginbotham (2010; see also Figure 1.1), for example, who adopted a detailed categorization scheme based on Fitzpatrick's makes a number of interesting statements (italics added):

- This student *characteristically* gave responses that were from the same lexical set … (p. 384).
- … Student 4 *favoured* conceptual association responses (p. 385).
- An important point to come out of the case studies is that there is a variety of student profile *types* … (p. 385).
- … responses were categorized according to their two most dominant response *preferences* (p. 387).
- … students identified as being same lexical set *orientated* could be encouraged to make word families in their vocabulary notebooks (p. 388).

[1] This is not to imply that Fitzpatrick claimed her categorization scheme to represent a reconceptualization of association data. In fact, she agrees that her use of "profiles" is "numerical" (personal communication) rather than conceptual. In other words, profiles here are quantitatively-driven, shaped by the level of detail in the categorization scheme.

Albrechtsen et al. (2008) report a multifaceted study of vocabulary knowledge in which a WA task is combined with a variety of other measures. The researchers adopt the term "lexical profile" to refer to the composite of individuals' scores across these various lexical measures. Like Higginbotham's (2010), this study illustrates what profiles are by providing a number of examples of different *types of learners*. The point here is that both of these studies have gone beyond defining a profile as merely the representation of WA data distributed across a detailed categorization scheme or across a variety of lexical measures. The use of "profile" is drifting from strictly a graphic or quantitative depiction of any given person's response data to the illustration of specific *cognitive styles*, individual differences in response patterns.[2]

The reconceptualization of individuals' WA responses as not merely the quantitative conglomeration of a detailed categorization scheme (i.e., a 'profile' in Fitzpatrick's original use of the term), but as an underlying style,[3] brings with it a number of hypothetical possibilities. Indeed, any or all of the following may be true:

---

[2] The term *cognitive style* is not used by any of these authors and they do not cite prior studies that attempted to explore this concept. There is, however, a history of this kind of WA research dating back to the 1960s (e.g., Cramer, 1968; Moran, 1966). Indeed, this concept appears to have been a once fruitful line of research that has since been abandoned. It's difficult to say why WA researchers discontinued this type of investigation.

[3] Like Higginbotham (2010) and Fitzpatrick (2007), I will use the term "response preferences" interchangeably with "cognitive styles". This by no means implies that cognitive styles are a reflection of informants' *conscious* decisions to adopt a specific response strategy during WA tasks. Indeed, the assumption underlying these, and all WA studies to date, is that WA responses are "naturally" occurring, unfettered by conscious scrutiny.

- Responses to the same stimuli should remain relatively stable over time.
- Responses to different sets of stimuli that are matched according to relevant linguistic and/or cognitive criteria will exhibit similar styles.
- The same response preferences will be exhibited in an individual's L1 and L2 associations (and L3, L4, etc., assuming they are reasonably fluent in their nonnative languages).

But is this shift in how a 'profile' is conceptualized warranted? On the one hand – like the introduction of the classification scheme examined in the previous chapter – the reconceptualization of associations in terms of cognitive styles potentially yields a fresh perspective on WA response data. On the other, it is only as reliable as the evidence that there actually *is* a consistent style for a given person. With the implications of interpreting the data as either data profiles or as cognitive styles in mind, the following replication study was conducted.

4.5 Proximities and profiles: A replication of Fitzpatrick (2009)

4.5.1 Aims

In keeping with the aims of this dissertation, this replication study was conducted to establish whether WA data is more usefully interpreted in terms of data profiles or cognitive styles. As I have explained above, by my definition (and in keeping with Fitzpatrick's), a response profile is a detailed, quantitative depiction of response data. Cognitive styles, on the other hand, refer to stable properties in the individual's response preferences (over time, across cues, across languages, etc.) that can differentiate them from other people. Evidence of such stability across conditions justifies the conceptualization. This study is thus intended to seek evidence concerning the plausibility of the cognitive styles concept (as Higginbotham, 2010 appeared to be moving toward). At the same time, as this is

not an exact or identical replication of Fitzpatrick (2009), differences between the two studies will help to address the following research questions:

1. Do respondents learning English as a foreign language (Japanese learners of English in Japan, specifically) exhibit L1-L2 response profile proximities similar to respondents in second-language learning contexts (e.g., English-speaking learners of Welsh in Wales, as in Fitzpatrick, 2009)?

2. Do Japanese respondents' L2 profiles become more like their own L1 profiles as they become more proficient? More precisely (as this is not a longitudinal study), do the L2 profiles of proficient native-Japanese learners of English more closely model their L1 response profiles than do the respective profiles of less-proficient learners?

4.5.2 Method

4.5.2.1 Stimulus selection

To examine the differences between L1 and L2 response profiles, two 100-word lists of stimulus words were prepared. I have commented on some of the difficulties involved in selecting stimulus words for L2 WA research in Chapters 2 and 3 and they have been expanded upon elsewhere (e.g., Fitzpatrick, 2007; Fitzpatrick, 2009; Higginbotham, 2010; Meara, 1982). Among these issues – and of particular importance in the EFL context in which this research was conducted – is the issue of word frequency or difficulty: stimuli must be difficult or infrequent enough not to elicit only the highly predictable, or "canonical" responses (see Henriksen, 2008), which may be indicative of what Meara (1982) refers to as the "hard core" of the L2 learner's mental lexicon. That is, cues like *boy* and *black*, to which the majority of respondents respond with *girl* and *white* respectively, indicate the existence of strong, quasi-fixed links in the lexicon. Other, less predictable associations, however, are perhaps more representative of the emergent bilingual lexicon that L2 researchers wish to investigate. That is, the links between items in the lexicon are not necessarily as strong or as fixed. During

the course of lexical development some node words are learned temporarily, while others are consolidated permanently. Some associations are formed and strengthened while others remain salient for only a short time and are then forgotten.

At the same time however, stimulus items must be simple and frequent enough to be understood, or at least recognizable, by the respondents. This is particularly true in the case of WA studies designed to examine response profiles where comprehensive data is necessary. A large number of null responses will yield a very patchy and unreliable picture of a respondent's profile. This becomes even more apparent when using a detailed set of response subcategories such as the one devised by Fitzpatrick (2006), fine-tuned in the previous chapter, and to be adopted in the current study. Finding statistically significant differences between subject groups will require an abundance of responses in any given subcategory.

As in the replication study reported in the previous chapter and the study upon which it was based (Fitzpatrick, 2006), stimulus words were selected from the Academic Word List (AWL; Coxhead, 2000). This was done in order to avoid the predictable cue-response pairs described above. As the AWL does not include the most frequent 2000 English words, most words that would prompt canonical responses are absent. As the list's name implies, it consists of words drawn from a corpus of academic texts from a variety of fields.

Most previous WA research utilizing stimulus words from the AWL (e.g., Fitzpatrick, 2006, 2007, 2009) involved random selection from the entire AWL.

Stimuli in the current study, however, were selected from only the first four (of 10) sublists, i.e., the most frequent academic words. This measure was taken in order to minimize the risk of non-responses and erratic responses of the kind elicited in previous WA studies in which the relative infrequency of some AWL-word stimuli may have prevented NNS from giving a true association as a response (see Chap. 3, also Higginbotham, 2010). As a further precaution in this regard, AWL words were also examined for their appearance in the JACET 8000 word list (Japan Association of College English Teachers; see Uemura & Ishikawa, 2004). The JACET 8000 list was created to provide a kind of academic word list for EFL contexts in Japan. Using a measure of log-likelihood, its creators compared word frequencies in the British National Corpus (BNC; see Leech, Rayson, & Wilson, 2001) with a subcorpus which included materials to which EFL learners in Japan were more likely to be exposed (English from textbooks, standardized examinations, children's literature, etc.). For the current study, words were selected only if they belonged to the first four sublists of the AWL and then also belonged to the first 3000 words of the JACET 8000. These criteria yielded 204 words which were then systematically divided into two matched lists and reduced to 100 words each as follows.

First, the words were divided as evenly as possible using word class as the criterion. The sole examples of an adjective/verb (*appropriate*) and a preposition (*despite*) to appear in the sample of 204 words were eliminated, along with one adjective (*apparent*) and one verb (*demonstrate*) selected at random. This yielded a total of 200 words comprised of even numbers of items within each of six word class categories (noun, noun/verb, verb, adjective, adjective/noun, and adverb).

This allowed them to be split into two sets with matching quantities within each category. The word class categories and the number of members within each are listed in Table 4.2.

Table 4.2. Word class distribution within stimulus sets.

Word class	Frequency
Noun	40
Noun/verb	21
Verb	19
Adjective	11
Adjective/noun	5
Adverb	4
Total	100

Not only were these word sets balanced in terms of word class membership (per Table 4.2), but they were also systematically divided so that word family membership (see Bauer & Nation, 1993) would be balanced across tasks. That is, members of the same family (e.g., *achieve-achievement, communicate-communication*) would not appear in the same stimulus set.

AWL sublist membership, determined by the relative frequency of the words within the AWL, was another factor contributing to which set stimulus words were assigned. Words were thus sorted in a manner that would, as evenly as possible, maintain the same number of stimuli from each AWL sublist in each stimulus set. This can be seen in Table 4.3. Table 4.4, which combines the data from Tables 4.2 and 4.3, shows how precisely word class category membership and AWL sublist membership were balanced across the two stimulus groups. Membership of the respective Levels of JACET 8000 was not scrutinized while dividing the words into stimuli sets, but was examined post hoc. As can be seen in Table 4.5, with the

exception of the noun/verb category, the two stimulus sets were quite well balanced in this regard as well. Undoubtedly, this is attributable to similarities in the frequency rankings between the AWL and the JACET 8000 lists.

Table 4.3. Frequency of AWL items in each stimulus set.

Stimulus Set	AWL Sublist 1	AWL Sublist 2	AWL Sublist 3	AWL Sublist 4	Total
Stimulus Set 1 (L2)	31	30	22	17	100
Stimulus Set 2 (L1)	33	29	20	18	100

Table 4.4. Frequency of AWL items in each stimulus set as a function of word class membership.

Word Class	Stimulus Set	AWL Sublist 1	AWL Sublist 2	AWL Sublist 3	AWL Sublist 4	Total
Noun	1	13	13	8	6	40
	2	12	13	8	7	40
Noun/verb	1	6	6	4	5	21
	2	6	6	4	5	21
Verb	1	8	5	4	2	19
	2	8	5	4	2	19
Adjective	1	3	3	3	2	11
	2	4	3	2	2	11
Adjective/noun	1	1	1	2	1	5
	2	2	1	1	1	5
Adverb	1	0	2	1	1	4
	2	1	1	1	1	4
Total	1	31	30	22	17	100
	2	33	29	20	18	100

Table 4.5. Frequency of JACET 8000 items in each stimulus set as a function of word class membership.

Word Class	Stimulus Set	JACET 8000 Level 1	JACET 8000 Level 2	JACET 8000 Level 3	Total
Noun	1	9	21	10	40

	2	9	21	10	40
Noun/verb	1	6	9	6	21
	2	2	18	1	21
Verb	1	4	12	3	19
	2	3	9	7	19
Adjective	1	3	5	3	11
	2	2	7	2	11
Adjective/noun	1	1	3	1	5
	2	1	2	2	5
Adverb	1	1	2	1	4
	2	0	1	3	4
Total	1	31	30	22	100
	2	33	29	20	100

Finally, the first set of 100 stimulus words became the stimuli in the English (i.e., L2) association task. These appear in Appendix III. In order to examine the differences between L1 and L2 response profiles, the other set of 100 words (Appendix IV) were translated into Japanese to serve as stimuli in the L1 task.

4.5.2.2 Participants and procedure

The subjects in this research were 44 university students of English as a foreign language studying at a medium-sized private university in Japan. All were native speakers of Japanese and had attained a mean English proficiency score of 659.9 on the standardized TOEIC test (range = 555 to 855; *SD* = 80.4). Participants received the L1 and L2 association tasks at separate sittings two to three weeks apart. Approximately half of the subjects received the L1 task followed by the L2 task. The remainder received the tasks in the opposite order.

The words were listed in a computer-generated random order (see Appendix V for an example). Participants were instructed in their L1 to respond to each cue by

writing the first word that came to mind (Japanese for the Japanese (L1) list; English for the L2 list) in a space provided beside each word in the list. They were informed that there were no right or wrong answers. This was to minimize the influence of participants' adopting a response strategy in accordance with how they felt they were supposed to respond. Subjects were also asked not to concern themselves with correct spelling so as to minimize the possibility that they would change their answers to those that were easier to write, rather than the first thing that came to mind. Data from participants who failed to complete more than 30% of the L2 task were eliminated on the grounds that it would not provide an accurate enough representation of their association profile. Six participants were eliminated in this way, resulting in a sample size of 38.

4.5.2.3 Response categorization

Responses were categorized according to the taxonomy created in light of the results of the replication of Fitzpatrick's (2006) study (see Appendix II). As explained above with reference to the 2009 study, this classification scheme was adopted because its incorporation of detailed subcategories was expected to yield a richer depiction of learners' association profiles than could be achieved through traditional categorization schemes. Moreover, the results of the study presented in the previous chapter support the claim that this scheme does in fact comprehensively categorize WA responses. Japanese responses in this study were classified by a native Japanese-speaking collaborator who received training in the use of the response categories detailed in Appendix II, and who was provided with Japanese examples of each type of association. All potentially ambiguous responses

(i.e., those that the collaborator had any doubts about) were discussed with the researcher until agreement was reached.

4.5.3 Results

Following Fitzpatrick (2009, p. 49), the relative distance between L1 and L2 profiles was calculated using a measure of Euclidean proximity. In other words, within-subject proximities (i.e., the distance between individual respondents' own L1 and L2 profiles) were calculated by first squaring the difference between the proportions (i.e., the percentages) of L1 and L2 responses for each response subcategory. Then the square root of the sum of these 12 values was calculated. The resulting figure thus represents how similar a respondent's Japanese (L1) response profile was to his/her English (L2) profile. The mean within-subjects proximity score for all 38 participants can be seen in Table 4.6. A between-subjects proximity measure was calculated similarly, but in this case subcategory comparisons were made between individuals' L2 scores and all other subjects' L1 scores. The mean for these 1406 (38 x 37) cases also appears in Table 4.6.

Within-subjects and between-subjects means were compared by way of an independent *t*-test. Results showed a significant difference between these scores ($t = 2.52$, $df = 1442$, $p < .05$). That is, respondents' native profiles and their L2 profiles were significantly closer than were their L2 profiles to other respondents' L1 profiles. These results closely resemble Fitzpatrick's (2009) findings reproduced in Table 4.7 ($t = 4.679$, $df = 1367$, $p < .001$). Therefore, the answer to the first research question is *yes*: the L1 and L2 response profiles of EFL learners (namely, Japanese

learners of English in Japan) do exhibit similar proximities as those of Welsh learners in Wales.

Table 4.6. Proximity scores.

	n	*Mean*	*SD*
Within-subjects proximity score	38	28.9	9.62
Between-subjects proximity score	1406	33.9	12.05

Table 4.7. Proximity scores from Fitzpatrick (2009, p. 49).

	n	*Mean*	*SD*
Within-subjects proximity score	37	28	11.08
Between-subjects proximity score	1332	38	13.46

To address the second research question – concerning whether learners' L2 profiles become more similar to their L1 profiles as L2 proficiency increases – the correlation between within-subjects proximity scores and TOEIC scores was examined. The results showed a moderate negative correlation (r_s = -.354, p < .05). This finding too replicates that of Fitzpatrick's 2009 study (r_s = -.370, p < .05) and provides support for the conclusion that learners' L2 response profiles increasingly approximate their L1 profiles as L2 language proficiency improves.

4.5.4 Discussion

It appears then that there is indeed evidence that we are justified in using *profile* to refer to an individual's tendency or predisposition to respond to WA cues in a specific way. While researchers may continue to use *profile* to refer to the graphic, or "numerical", representation of a subject's responses as distributed across categories, they should recognize that they are simultaneously referring to both

the representation and the underlying cognitive style of the respondent that it represents. The concept has received validation in the fact that subjects' individual L2 responses correspond more closely to their own L1 responses than to the L1 responses of their cohorts. This adds weight to the hypothesis that language learners' personal response profiles are stable whether they are examined through native language associations or through those of a second language. The fact that individuals' L2 responses become more like their own L1 responses as L2 proficiency develops also points to the strength of individual profiles across languages.

Although the present findings are too limited to demonstrate it, they are consistent with the possibility that the L2 mental lexicon is restructured and reorganized as L2 proficiency develops, and that this reorganization gradually approximates the structure of the L1 lexicon. It may also be the case that as the associational links between nodes in the L2 network strengthen and consolidate, an individual's preferred (L1) processes for access and retrieval during WA tasks become less hindered by missing or weak links. Thus, L2 association behaviour increasingly conforms to the individual's L1 preferences.

While either of the explanations above provide a reasonable account of the current findings, I wish to propose an alternative explanation: as the L2 lexicon develops and transforms with gains in L2 proficiency, so too is the L1 lexicon transformed. The consolidation of *L1-L2* links and *L2-concept* links (see Kroll & Stewart, 1994) may weaken or otherwise alter *L1-concept* links. Thus, while L2 association behaviour is transformed in accordance with growth in L2 proficiency, L1

associations may also undergo modification. Language attrition (also called *language loss*, or *language forgetting*), is just one of several well-documented phenomena observed in the L1s of bilinguals (see Higby & Obler, 2015; Isurin, 2013; Schmid, Köpke, Keijzer, & Weilemar, 2004). While it is beyond the scope of this dissertation to explore empirically how these changes might be reflected in WA behaviours, there is evidence that the L1 lexicon undergoes some amount of change as the L2 develops.[4] One can see how a within-subjects study of L1 and L2 profile strength over the course of L2 development might reveal the degree to which each of these competing hypotheses accounts for the current findings. Further research intended to more precisely determine the mechanisms underlying the current results would benefit from being conducted longitudinally.

4.6 Conclusions

While one of the purposes of this study had been to examine the legitimacy of respondent profiles as a means of interpreting WA responses, this study has also served as a first test of the modified response classification introduced at the end of the last chapter (see Appendix II). As expected, it appears to have served its purpose, providing a detailed framework for classifying the various types of responses elicited during L2 WA studies. What has also become clear, however, is that the distribution of responses across such a detailed classification yields frequencies in certain subcategories so small that statistical analysis is rendered impractical. As a case in point, *clang* and *other* responses in prior WA research tended to account for only 1 to 2% of total responses. Membership in these two

[4] For studies examining other specific ways in which the L1 is altered simultaneously with the development of the L2, see Cook (2003).

categories (now *form-based* and *other*) is currently divided among six subcategories. Comprehensiveness of the categorization scheme appears to have come at the expense of statistical analysability of the data.

A related issue, alluded to in my comment about clang/form-based responses, is that in practical terms there are very few differences in the manner in which responses were classified here and in the ways they would have been classified using a traditional categorization scheme. Responses traditionally placed in paradigmatic, syntagmatic, and clang categories would now be categorized as meaning-based, position-based, and form-based, respectively. That is, the distribution of responses across categories would be identical, for the purposes of most research, regardless of which scheme is employed. Perhaps the only substantial exception to this occurs where some traditionally syntagmatic responses are now considered conceptual associations (i.e., members of the meaning-based, rather than the position-based category). Syntagmatic responses were often defined as a kind of collocational response in that cues and responses could co-occur in the same sentence, but not replace each other, due to a difference in word class. An example of this is *institution-working*. In the new categorization scheme (Appendix II), *working* would fall into the conceptual subcategory because the cue and response are semantically related, yet they do not constitute an obvious collocation. Arguably, this classification more clearly defines the position-based category as the repository of collocations. However, in the absence of any other substantial differences between the traditional and the current schemes, it may also be argued that the creation of the new categorization has merely relabelled, rather than reconceptualised, response data.

While the utility of the new categorization scheme may be debatable, the same cannot be said about response profiles. The results of the replication study presented here almost duplicate exactly those of Fitzpatrick (2009). Therefore, I believe WA researchers are justified in referring to individual respondents' results as profiles. Types of responses – across a subject's L1 and L2 at least – appear to remain relatively stable, illustrating a cognitive style unique to that individual. The question is: how strong is one's response style?

The research presented in the next two chapters will help to address this question as well as the following. At what point does experimental manipulation cause informants to cease to respond according to their own style? How different need the cues be across sessions? How much time needs elapse?

Chapter 5. Priming vs. profiles: Putting cue order, cue strength, and response preference in context

5.1 Introduction

In Chapter 3, I examined the ability of Fitzpatrick's (2006) two-tiered categorization scheme to comprehensively and inclusively cover the types of responses made by NS and NNS respondents. With minor adjustments (see Appendix II), I found that the scheme adequately classified the varieties of responses elicited in L2 WA studies and I was able to successfully adopt the revised scheme for the research presented in Chapter 4, even if it was not directly needed for the profile analyses.

I also focused on WA responses in Chapter 4. There, however, the purpose was to examine whether it is valid to characterize WA responses in terms of individual response profiles. Prior studies (Fitzpatrick, 2007, 2009) appeared to justify interpreting the data in this manner as results had indicated that response patterns within groups of native-speaking respondents were not homogeneous and that the relative proximity of an individual's L1 and L2 response profiles correlated with L2 proficiency. These findings may seem like ample evidence of the utility of conceptualizing responses as profiles. However, as I have argued in Section 1.5, the consequences of invoking profiles as a concept are potentially profound. As explained, response profiles rendered through the use of a detailed categorization scheme allow researchers a window into respondents' cognitive styles. In other words, the distribution of an individual's WA responses across a

detailed set of categories – especially when depicted in graphic form (see Figure 1.2; also Higginbotham, 2010) – can reveal a unique perspective on a subject's response preferences. These cognitive styles, depending on their strength, should predict the response patterns of individual subjects on future WA tasks, regardless of which cues are used or which language the cues are presented in. The findings presented in Chapter 4 support this.

The impact of these findings extends to the interpretation of research that has attempted to establish native-speaker WA norms (e.g., Postman & Keppel, 1970), research that has utilized such norms for the purposes of assessing L2 proficiency (e.g., Kruse et al., 1987), and research utilizing WA methods for the diagnosis of psychological or neurological disorders (e.g., Gollan et al., 2006; Merten, 1993). These consequences, were the rationale for replicating Fitzpatrick's (2009) study. As the results of the replication reported in the previous chapter indicate, her main findings were confirmed. It appears that learners' L2 response patterns begin to more closely resemble their L1 patterns with increased proficiency in their second language. This evidence seems to suggest that the conceptualization of WA responses in terms of profiles and cognitive styles is not only tenable, but that it offers potentially fruitful avenues for future research.

One of the key findings of the prior research into association profiles – and a key assumption going forward – is that individual response behaviour is relatively consistent. This has been demonstrated with different L1 cues over time (Fitzpatrick, 2007) and across respondents' first and second languages, whether they be English and Welsh (Fitzpatrick, 2009), or Japanese and English (Chapter 4).

One of the purposes of the current study, then, is to investigate precisely how stable response profiles remain under specific experimental conditions and to determine whether certain manipulations can steer respondents from their usual preferences.

We can infer from prior studies that some types of cues are more likely to elicit specific types of responses than are others. In the case of L1 associations, for example, Deese (1962) found that nouns and high-frequency adjectives typically produce paradigmatic responses (i.e., responses belonging to the same word class as their cues), while low-frequency adjectives, verbs, and adverbs are more likely to elicit syntagmatic responses. Since Deese's early work, many researchers have observed similar word class effects in both L1 (e.g., Entwisle, 1966a; Fillenbaum & Jones, 1965; McNeill, 1963; Stolz & Tiffany, 1972) and L2 studies (e.g., H. B. Nissen & Henriksen, 2006; Racine, 2008, 2011b). Besides word class, other manipulations of association stimuli that may impact upon response profiles include the use of loanwords and cognates (Racine, 2011c; Van Hell & De Groot, 1998), selecting cues of varying frequencies (Higginbotham, 2010), concreteness (Van Hell & De Groot, 1998), affectivity or familiarity (see Cramer, 1968 for a summary of these findings), presenting cues in participants' first or second language (Kolers, 1963), or in written vs. oral format (Kudo & Thagard, 1999; Linton & Brotsky, 1969; Racine, 2008). Given this large body of evidence, it would seem that several factors influence response behaviour. The cumulative impact of such variables or the influence of any one of these will potentially affect individuals' response behaviour, undermining any clear indication that there is a 'usual' preference, and what it might be.

5.2 The influence of cue order

One largely overlooked area of WA research involves the manipulation of cue order. This oversight may stem from the fact that WA researchers' main focus lies in the connections between stimuli and responses, rather than the connections between stimuli themselves and the possible influence these connections may have on responses. Take for example two lists of WA task cues. As *bank* is a homograph representing a financial institution as well as the land adjacent to a river or lake, we should expect a certain percentage of responses to reflect each of these interpretations. Now consider the impact if, in one list of WA cues, the word *bank* is preceded by the word *wallet* and in another it is preceded by the word *sea*. Intuitively, we can expect research participants to respond more often to the former with words related to banking and finance, and to the latter with, for example, *river* or *lake*. Indeed, a wide variety of vocabulary recognition studies (see Neely, 1991 for a review) shows that information preceding a recognition task has clear and substantial effects on the results. The same holds true in vocabulary production tasks where the effects may include substitution errors (e.g., Garrett, 1992, 1993) or interference effects (Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990). Broadly, these effects are referred to as *priming effects* and are documented throughout the cognitive psychology and psycholinguistic literature.

5.3 Types of priming effect

One of the most frequently investigated priming effects is that of *semantic* priming. This is where subjects respond more quickly to a word when it is preceded by a

semantically related word than when it is preceded by an unrelated one. This effect is seen in lexical decision tasks and naming tasks (e.g., Collins & Loftus, 1975; Cree, McRae, & McNorgan, 1999; McClelland, 1987; McRae & Boisvert, 1998) and is considered to be a reflection of lexico-semantic organization in the lexicon. That is, these effects demonstrate links between semantically related words. For example, after being exposed to the word *zebra*, you should be able to more quickly identify the word *lion* than if you had seen or heard the word *table*. Depending on the type of relation between primes and targets, semantic priming may be divided into three subcategories (see McDonough & Trofimovich, 2009, p. 62): *associative* priming (where primes and targets are related semantically, but not part of the same semantic set; e.g., *sugar-sweet*), *category* priming (where primes and targets share hierarchical or cohyponymic relations within semantic categories; e.g., *bird-eagle, tulip-rose*), and *mediated* priming (where the words do not share direct semantic links, but are mediated by another word or concept; e.g., *stripes-lion* which presumably is mediated by the word *tiger* or the idea of a tiger). Semantic priming effects are said to be stronger under *self-priming* conditions – where primes and targets are both produced by the same speaker (Jaeger & Snider, 2007 in Taylor, 2012).

Another type of priming is *syntactic* (or *structural*) priming (e.g., Arai et al., 2007; Branigan et al., 2000; McDonough, 2006). This refers to the increased likelihood that an individual will use a recently encountered syntactic structure rather than an equally viable alternative. McDonough and Trofimovich (2009, p. 98) provide the example of a listener encountering a double-object dative (e.g., *Susie baked her friends a cake*) and then producing another in identical form (*John bought his*

mother a bicycle) rather than utilizing some other acceptable structure such as a prepositional dative (*John bought a bicycle for his mother*). It has been argued that proficient L2 learners possess a "particular aptitude" to acquire structures more efficiently via this process (McDonough & Mackey, 2008).

A third type of priming is *auditory* (e.g., Bassili, Smith, & MacLeod, 1989; Ellis, 1982; Schacter & Church, 1992). This refers to the facilitation of processing spoken words or word combinations based on the listener's prior experience. Put simply, we process spoken words more easily if we have heard them already in conversation. This type of priming is also referred to as *form* priming (e.g., Forster & Veres, 1998) when stimuli are presented visually rather than aurally. Together, syntactic and auditory/form priming are referred to as *repetition* priming, as they appear to indicate a language user's sensitivity to repeated exposure to language forms. When repeated primes are identical to their targets, it may also be referred to as *identity* priming. Priming – whether semantic or repetition priming – may be considered *masked* if the presentation of primes is obscured either by the immediate presentation of the target (typically less than 80 ms later), or by the presentation of symbols (e.g., ######). These methods are intended to reduce visibility of the primes. Subjects in masked priming trials may be unaware of their exposure to primes altogether.

Finally, if it is not obvious already, it should be noted that there are clear parallels between the types of priming described above and the WA categories I have been investigating in this dissertation. The similarities are between semantic priming and meaning-based responses, between syntactic priming and position-based

responses, and between visual and auditory priming and form-based responses. Indeed, even the *mediated associations* subcategory that I proposed in Chapter 3 (see Appendix II) shares features with mediated priming and even shares its name. The connections between each of these types of priming and the various types of WA response are undoubtedly deeper than mere surface similarities. I will address the nature of these relationships and their implications for the current study in Section 5.5.

5.4 A primer in priming: Elgort (2011)

Before continuing, it is useful to review a pertinent paper that illustrates the application of priming methods to the study of L2 lexical acquisition. Elgort (2011) describes three lexical decision task (LDT) experiments designed to determine the extent to which the deliberate learning of vocabulary would trigger the acquisition of formal-lexical and lexical-semantic representations. The priming methods adopted in her study are relevant to the ones I will employ in this chapter and again in Chapter 6.

In the learning phase of each of the three experiments, Elgort asked subjects to study a series of pseudowords using word cards (with the pseudoword on one side and a pseudo-meaning on the other). The pseudowords were all pronounceable nonwords, seven to nine letters in length. Each differed from its respective base word by only one letter (e.g., *infecent, indecent*) and all observed the spelling and pronunciation rules of English. Base words were chosen so that they had no

orthographic neighbours and were of relatively low frequency. These were used as targets in all three of the experiments that followed.

The first experiment utilized a masked, form-priming paradigm in which trials consisted of the presentation, via computer, of a row of hash marks, a prime and a target word for about a half-second each. The targets consisted of real English words and nonwords.[1] Prime words consisted of real words, nonwords, and the pseudowords studied previously. Real word primes were either orthographically related or unrelated to their targets. The results replicated the findings of Forster and Veres (1998), showing a significant prime lexicality effect. That is, subjects responded more quickly to real word targets when they were preceded by orthographically related *non*word primes than they did when the prime was an orthographically related word. More importantly, pseudoword primes showed no significant effect on response times (RTs), suggesting that the formal-lexical representations of the pseudowords "were established and integrated in the mental lexicon of the participants" (p. 382).

In the second experiment, a masked, repetition priming paradigm was employed. That is, primes were either unrelated (neither orthographically nor semantically related) or were identical to the target words. Words preceded by identical primes were recognized more quickly than when preceded by unrelated words, presumably because their representation had already been activated. However, this effect was not observed for nonword targets. Elgort attributed this to the fact

[1] Nonwords were constructed in the same manner as pseudowords. The difference was that there were no pseudo-meanings associated with the nonwords.

that being previously unseen forms, no lexical entries exist for nonwords in the lexicon, so preactivation did not occur. As for the pseudowords, if lexical representations of the pseudowords had been acquired during the prelearning stage, those stimuli should be more quickly recognized than nonwords (which had not been studied by the participants). In fact, this is precisely what Elgort found. A robust repetition effect was found for the studied pseudowords, similar to the real word targets used in the experiment. Like the results of Experiment 1, this appears to demonstrate that the learned pseudowords behave similarly to previously acquired vocabulary. This finding has important implications for vocabulary learning which I will elaborate on below.

The third experiment was designed to establish if a semantic priming effect could be elicited. This effect occurs when a target is recognized more quickly when preceded by a semantically related word. If the lexical-semantic representation of a pseudoword like *reatangle* – learned by participants as "an overhang that projects over a window or outside door and serves as protection from the rain and snow" (p. 391) – existed in the lexicons of the participants, then it would semantically prime a target word such as *balcony*. The results of this experiment too did show a positive priming effect (i.e., a mean difference in RT of 22 ms) when pseudowords were semantically related to their targets. Semantically related real word primes demonstrated the same effect, but the effect for real words was significantly stronger (a mean difference of 37 ms) than the effect for the pseudoword primes. Elgort concludes that while the lexical-semantic representations of the pseudowords had been acquired (in most cases), the representations of some pseudowords, for at least some of the participants, were

unstable. This resulted in competition among semantic neighbours in the lexicon during activation and thus an inhibitory effect (see Dagenbach et al., 1990).

Finally, as a further test of the automaticity of the priming effects demonstrated in Experiments 2 and 3, Elgort employed a coefficient of variation measure. Doing so she was able to demonstrate that differences between reaction times (RTs) in trials where priming effects had been exhibited for non-real words (i.e., pseudoword trials) and those where no priming had occurred, were not attributable to participants' simply becoming faster at recognizing related forms and meanings. She found that changes in RTs remained constant for primed trials, indicating a qualitative change in the processes underlying the lexical decision task. Namely, she argued, participants relied less on controlled lexical processes and had exhibited more automaticity in the task.

Before addressing the relevance of this study to the current dissertation, I should say that this research makes quite a significant contribution to our knowledge about second language acquisition and ultimately to the teaching of English (and other languages) as a second or foreign language. The results contribute to the growing body of knowledge (e.g., N. C. Ellis, 2002; R. Ellis, 2002a, 2002b) showing that the deliberate study of language – in this case focusing on the form and meaning of new vocabulary – can go beyond immediate declarative knowledge to true acquisition. There are still those who debate the merits of "natural" teaching methods versus classroom methodologies that focus on language forms and student output (e.g., Racine et al., 2013). However, the current study provides quite strong support for the hypothesis that differences do not exist between the

processes and representations of "acquired" vocabulary and deliberately learned items presumed by some researchers to be qualitatively different (see Krashen, 1981, 1982; Krashen & Terrell, 1983). As Elgort concluded, deliberate learning is an efficient, convenient, and effective method of acquiring L2 vocabulary and "the hypothesis regarding the learning/acquisition dichotomy is not justified" (p. 399).

While I believe this study is a valuable addition to the lexical acquisition literature, there are still a number of concerns to address. First of all, it should be noted that this is very much a psycholinguistic study. This is not classroom research. We may acknowledge that the experimental design is solid and that the research was conducted systematically, but we may not be fully justified in extrapolating the findings beyond the laboratory. Real language learning involves consolidation of new lexical items in an environment where there may be repeated exposure to the words in question. It is not clear from strict laboratory studies (particularly those involving pseudowords) how further exposure would influence the results. Another issue here involves aspects of word knowledge. Encounters with new lexis in the classroom involve exposure to a variety of forms and contexts. Thus, the manner in which words are acquired and the specific aspects of word knowledge that are acquired may be just as varied. Laboratory studies like this one, however, often treat "acquisition" as an either/or equation typically operationalized as the acquisition of form-meaning pairs.

As for its relation to this dissertation, Elgort's study provides excellent examples of the kinds of cognitive psychological and psycholinguistic methods – in this case, priming, RT measures, and LDTs – that may be brought to bear in lexical research.

In particular, her study describes in detail the types of priming effects that may be observed during cognitive tasks, such as the lexical decision trials employed here, and it details the methodologies employed. All of these will prove useful in the studies I will present in this and the next chapter.

5.5 Priming, word association, and response profiles

I noted in Section 5.3 the similarities between the WA response types investigated in this thesis and the many type of priming effects that have been investigated by cognitive psychologists and psycholinguists. The relationship is more than just a resemblance. Indeed, some researchers may argue that WA itself is actually just a form of priming in which the cue primes, or facilitates, the production of a specific response over other possible responses. This is not an unreasonable viewpoint but, for the purposes of this dissertation, where a strong focus has been placed on methodology, I will treat them as two separate methods in the ways they are implemented, and as two separate mechanisms in how they affect respondents. The key difference between WA and priming methods is that in WA, barring some type of additional experimental manipulation, cues elicit only those responses that subjects are naturally predisposed to respond with. Priming, on the other hand, is precisely one of the experimental methods (i.e., manipulations) that may cause subjects to veer from their usual response behaviour. The priming effect, applied in a traditional manner might reveal facilitation (i.e., faster than normal RTs) or interference (i.e., slower RTs). For the current studies, however, priming will be measured in terms of changes in response type, rather than response time. The distinction I have made between the two types of methodology yields an

opportunity to combine them in a manner that researchers have rarely (if ever) attempted. This particular way of combining WA research with priming methodologies may prove to be very fertile experimental soil indeed.

The methodology I am about to describe not only brings together WA and priming methods, it will also act as a further test of the response profile or cognitive style concept that I examined in the previous chapter. Essentially, response profiles will be pitted against a priming condition (in this case, brought about by the manipulation of cue order). If the manipulation of cue order leads to deviation from responses expected on the basis of subjects' response profiles then we may conclude that their strength is rather relative to the WA task at hand. On the other hand, if – despite the manipulation of cue order – no priming is detected in terms of responses deviating from those predicted by the profiles, we may conclude that profile strength is particularly durable even in the face of the experimental manipulation I am introducing here.

5.5.1 Aims of the study

The purpose of this study, then, was to determine whether subjects' responses to specific cues could be manipulated. This entailed knowing what their response profiles were *without* such manipulation, and then applying an intervention that would test their resilience. Thus, it was necessary to create subsets of cues known to elicit specific types of associations from a given participant. The basic method was to use the results from a prior WA task to select a series of cues that would be placed in a specific order for a second task, so as to let one response type cue the next. For example, if a given respondent has replied with meaning-based

responses to a number of cues, these cues can be placed in consecutive order (Series 1) for the second task. Another series of cues to which the respondent has previously replied with position-based responses (Series 2) can then be presented immediately after Series 1. The experimental hypothesis is that the processes involved in responding to the first series of cues will have an effect on the manner in which the second series of cues is processed, and that these differences in processing will be reflected in the types of responses elicited. In other words, cues may elicit responses unlike those predicted by the participant's response profile (i.e., as measured in a previous WA task) if the participant has repeatedly engaged in a different type of association process immediately prior to the experimental trials.

This experimental design entails separately developing test materials for each participant. The design is illustrated in Table 5.1. Cues are placed in random order for Task 1.[2] Cue order for Task 2 is manipulated based on the responses elicited from the specific individual during Task 1. Cues that initially elicited a meaning-based response (Cues 1, 2, 5, 6, and 8) are used as a series of primes (described above as Series 1). Cues that had elicited position-based responses during Task 1 (Cues 4 and 7) are used to elicit target responses (Series 2 above; Table 5.1). If

[2] It should be clear by now that one of the basic premises of the current research is that WA responses may be subject to priming effects depending on the order in which cues are presented. Readers may also infer then that priming may inadvertently occur in the association task described here (just as it may in any WA task, for that matter), potentially confounding or masking the results of the second task. For this reason – as described in the procedure section below – two, different randomly ordered task forms were administered in Task 1. While not completely eliminating the possibility that individual subjects' responses may still be influenced by inadvertent priming, this measure at least serves to limit the effects of specific cue orders across subjects.

repeatedly engaging in a meaning-based WA process does create a priming effect, the subject's responses to Cues 4 and 7 would be predicted to become meaning-based responses. If this turns out to be the case, then there is evidence that despite the apparent strength of association response profiles demonstrated in the previous chapter, cue order is potentially stronger. That is, priming overrides profile. If so, then cue order is an essential aspect of the WA process that researchers must take into consideration when designing their studies.

Specifically, this study utilizes the method described above to address the following research questions:

1. Do priming effects override response profiles? That is, will repeatedly engaging in the same association process (priming) alter subsequent response processes such that respondents no longer reply in accordance with their own response history (i.e., their profile)?

2. To what extent might priming effects be observable in the results of first- and second-language respondents respectively? I.e., is a certain level of proficiency/automaticity necessary for priming manipulations to have an effect on responses?

3. If priming effects are observed, are there recency effects? That is, is it more likely that target responses elicited immediately after the primes will show the priming effect, or will the effect be spread more generally across the entire series of target trials?

Table 5.1. An example of a priming methodology for a two-task WA experiment.

Cue Order for Task 1	Responses
Cue 1	meaning-based
Cue 2	meaning-based
Cue 3	form-based
Cue 4	position-based
Cue 5	meaning-based
Cue 6	meaning-based

Cue Order for Task 2	Predicted Responses in Task 2
Cue 1	meaning-based
Cue 2	meaning-based
Cue 5	meaning-based
Cue 6	meaning-based
Cue 8	meaning-based
Cue 4	?

Cue 7	position-based
Cue 8	meaning-based

Cue 7	?

5.5.2 Method

5.5.2.1 Subjects

To answer these questions two English WA tasks were administered to two groups of respondents. One group consisted of 90 native speakers of English (NS) studying at a university in the United Kingdom. Twenty-three of these subjects were unavailable to undertake the association task at Time 2 resulting in a sample of 67. Responses to a demographic question about their language background confirmed that all NS respondents were either native speakers of English (n = 64) or that despite English not being their native language, they did "not have any problems using English" in their daily lives (n = 3). The other group consisted of 94 native speakers of Japanese (NNS) who were studying at a university in Japan. Six of these subjects were unavailable to take part at Time 2. The data from a further nine subjects was discarded as there were not enough position-based responses (fewer than six) to create the second word association task. That is, subjects placed in the meaning-based priming condition need to have responded to at least six Task 1 cues with position-based responses so that these may serve as target and control cues in Task 2 (see Section 5.5.2.3 for a thorough explanation of the Task 2 design). Removal of these 9 participants' data left a total of 79 NNS participants.

5.5.2.2 Cue selection for Task 1

The first WA task consisted of a list of 100 cues. Two different forms were created so that instructions would appear in respondents' first languages (i.e., Japanese or

English). As one of the premises of this study was that cue order potentially influences response behaviour, two versions of each form were created (i.e., with cues presented in different random orders). One of the English forms can be seen in Appendix V. Cues were selected from lists of word association cues utilized in prior studies (Fitzpatrick et al., 2015; Jenkins, 1970; H. Moss & Older, 1996) with the conditions that 1) the majority of primary responses did not belong to any single response category, and 2) that they would be understandable by Japanese learners who had achieved low-intermediate to intermediate English proficiency. These two criteria and the selection process are elaborated on below.

> <u>Criterion 1:</u> *The majority of primary responses did not uniformly belong to a single-response category.* That is, if the top 50% of responses to a given cue all belonged to either the meaning- or position-based response category, the cue was not included in the study. Table 5.2 illustrates two cues that were rejected for this reason. As can be seen, the most frequent 52.0% of responses to *frighten* (*scare, children, ghost, fear,* and *terrify*) were all meaning-based responses. Likewise, the top 81.0% of responses to billiard (*ball* and *table*) were position-based responses. Conversely, if the primary responses to a given cue belonged to a diversity of response categories, the cue was seen as acceptable for the current study. Examples of these cues (*shoulder* and *voice*) can be seen in Table 5.3, where it is clear that both meaning- and position-based responses were among the top 50% of primary responses. This selection criterion – that the majority of primary associations to an individual cue should not belong to a single response category – is similar to the criterion employed in previous chapters where cues that elicit dominant primary responses (e.g., *mother-father, black-white*) were also eliminated. As the strength of these links can be witnessed in both English associations and in those of a variety of other languages (e.g., Meara, 1982), they may prove ineffective in L2 WA studies where potential differences between learners' and native users' lexicons may go undetected. By the same token, these types of cues exhibiting high association strength may remain strongly linked to their most common responses regardless of contextual influences such as cue order and may not, therefore, be susceptible to priming effects. While this hypothesis is yet to be tested, it is noteworthy that associative strength itself has been utilized as a measure of relatedness in the selection of primes and target

words for previous studies in semantic priming (e.g., Meyer & Schvaneveldt, 1971).

Criterion 2: All cues were selected so as to be understood by Japanese learners of English. Thus from all cues matching criterion 1 above, only those that belonged to the first three levels – the first 3000 words – of the JACET 8000 (Uemura & Ishikawa, 2004) word list were selected. As described in the previous chapter, JACET 8000 is a list of the most frequent English words likely to be encountered by learners of English in Japan. Texts included in the corpus from which they were derived include textbooks, standardized examinations, children's literature and other sources.

Table 5.2. Examples of cues to which the majority of primary responses belong to a single response category (responses and response percentages from H. Moss & Older, 1996).

Cue: **Frighten**

Primary Responses	Percentage of total responses	Response category
Scare	30.4	meaning-based
Children	6.5	meaning-based
Ghost	6.5	meaning-based
Fear	4.3	meaning-based
Terrify	4.3	meaning-based
…		

Cue: **Billiard**

Primary Responses	Percentage of total responses	Response category
Ball	42.9	position-based
Table	38.1	position-based
Cue	4.8	meaning-based
Snooker	4.8	meaning-based
…		

Table 5.3. Examples of cues to which the top 50% of primary responses belong to two or more categories (responses and response percentages from H. Moss & Older, 1996).

Cue: **Shoulder**

Primary Responses	Percentage of total responses	Response category
Cry	23.4	position-based
Arm	10.6	meaning-based
…		

Cue: **Voice**

Primary Responses	Percentage of total responses	Response category
High	9.5	position-based
Over	9.5	position-based
Speech	9.5	meaning-based
…		

5.5.2.3 Cue selection for Task 2

Sixty cues were used in Task 2. Thirty of the cues were new to the subjects (i.e., had not been used in Task 1) and were selected using the same criteria as the words in Task 1 (i.e., they were likely to produce a variety of primary response types and likely to be understood by the NNS respondents). These cues were distractors and were the same for all subjects, appearing in the same order on all forms. The other 30 cues were to serve as experimental cues in this task and were selected from cues already used in Task 1. Selection of these cues and the positions in which they appeared on the form depended upon the responses of individual respondents elicited in Task 1. That is, Task 2 forms were customized for each individual respondent.

A "blank" version of the English form for Task 2 appears in Appendix VI. The cues are numbered for ease of explanation. Numbers did not appear on the actual experimental forms. All forms had the individual respondent's name pre-printed at the top, followed by instructions identical to those that appeared on the Task 1 form. The *experimental block* of cues appears in the last 13 positions on the form (cues 48 to 60). While this basic format was identical for all subjects, each had been placed in either a position-based priming condition or a meaning-based priming condition. Subjects in the position-based priming condition had responded with position-based responses to at least 10 cues in Task 1, and these cues were used as primes in task 2. The final three positions were cues that had elicited meaning-based responses during Task 1 (i.e., the *targets*). Thus, the experimental manipulation was to have subjects repeatedly engage in the position-based response process for the 10 prime trials, and then measure the amount of change (i.e., from meaning-based responses during Task 1 to position-based responses during Task 2) in responses to the three target cues. Likewise, subjects in the *meaning-based priming* condition received 10 primes (cues 48 to 57), to which they had responded with meaning-based responses during Task 1, followed by three target cues (58 to 60) to which they had responded with position-based responses during Task 1.

A control block of cues (cues 6-18) was also included in the Task 2 list. The three control cues (cues 16 to 18), like the target cues in the experimental block, had been utilized in Task 1 and had elicited the same types of responses. That is, for subjects in the meaning-based priming condition, these were cues to which the

respondents had offered position-based responses in Task 1. For subjects in the position-based priming condition, these were cues that had elicited meaning-based responses during Task 1. Comparisons between the control cues and target cues would yield a determination as to whether the target responses had changed in Task 2 due to the influence of the primes that preceded them, or whether this was attributable to a more general instability in subjects' responses over time. The 10 fillers that preceded the control cues (cues 6 to 15) were randomly selected items from cues included in Task 1 and were not dependent on subjects' responses to them during the first task. The decision to use Task 1 cues (as opposed to cues not previously encountered by the subjects) in these pre-control slots was to rule out the possibility that any differences in changes of response behaviour to the targets and controls was that the primes had been encountered previously.

Another precaution taken in the research design of Task 2 was that the experimental block appeared in the final positions of the survey and the control block appeared almost at the beginning. This was to counter another possible confound: that the effects of the experimental block may continue to influence other responses, in particular, the responses to the control cues. The choice not to place the control block in the very first slots of the form was yet another precautionary measure against the possible confounding influence that subjects may recognize the manipulation. That is, subjects may have recognized that there were blocks of previously encountered cues amid the new cues and, in some manner, might respond differently than they otherwise would. It was therefore deemed necessary to begin with five previously unseen cues to mask the presence of the first block. Likewise, four filler items taken from Task 1 were placed in

randomly selected slots throughout the form (cues 20, 29, 37, and 41). This was to give the impression that the cues they had seen before were randomly distributed across the Task.

No formal experimental measures were taken to determine whether these precautions had in fact prevented subjects from recognizing the existence of the experimental blocks. However, upon completion of the second task, respondents were asked as a group whether they had recognized the presence of cues from Task 1. While some subjects had indeed recognized certain cues from Task 1, none of them had recognized that they had appeared in blocks. Nor were subjects able to guess what the purpose of the study had been until after they had been debriefed.

The decision to adopt the methodology of Task 2 (where changes in responses to control cues were to be compared with responses to target cues) was based on two main premises. The first was that repetition priming effects may be induced by manipulating WA cue order (hence, they may be induced unintentionally when cue order is not taken into account by researchers preparing WA research forms). The second premise was that individual research participants' responses remain relatively constant over time. This premise finds support in the findings of the previous chapter (as well as in the results of Fitzpatrick, 2007, 2009) where strength or stability of individual response profiles was demonstrated. That said, however, for the findings of the current study to be considered reliable it was necessary to contrast the effects of the priming manipulation on the target cues with changes in responses to control cues. The latter would indicate general instability in response behaviour over the course of the two tasks, and reduce the

confidence with which changes in the primed responses could be attributed to the priming itself.

5.5.2.4 Procedure

The tasks were completed at the subjects' respective universities (i.e., a university in the UK for the NS group and a university in Japan for the NNS group) during class hours. All instructions on the task were written in the subjects' native language. In Task 1, respondents were randomly assigned to receive one of the two forms (i.e., different cue orders) in keeping with the premise described above that cue order potentially affects response outcomes. No time limit was imposed. All subjects were able to complete the task within about 30 minutes (approximately 20 minutes for the NS group). Responses to Task 1 were categorized using the four main WA categories outlined in Appendix II: meaning-based, position-based, form-based, and others. The decision not to include the many subcategories examined in Chapter 3 was due to the necessity of having enough responses in two individual categories to create the Task 2 forms.[3] Task 2 was administered six to eight weeks after Task 1.

As noted earlier, for Task 2, subjects were randomly selected to be in one of two conditions. One of these was a *meaning-based prime* condition. Subjects in this group received as primes 10 cues to which they had responded with meaning-based responses in Task 1. In accordance with the experimental design described

[3] See the discussions in Sections 3.3.2 and 4.6 concerning the interaction between the employment of subcategories, the number of elicited responses, and statistical validation. See also Fitzpatrick et al. (2015) where a recently completed components analysis statistically justifies the use of these categories.

above, target cues and control cues were those to which these subjects had responded with position-based responses during Task 1. Subjects placed in the *position-based prime* group received 10 cues to which they had responded with position-based responses in Task 1 as primes. Target cues and control cues were those to which they had responded with meaning-based responses during Task 1.[4] Filler cues in both conditions were randomly selected from the remaining Task 1 cues. Task 2 responses to the cues in the control and experimental blocks were categorized using the same criteria as in Task 1, resulting in the same four types of response: meaning-based, position-based, form-based, and others.[5]

5.5.3 Results

To address the first research question – whether priming effects might influence WA subjects' responses – participants' responses to the three target cues were examined. Each trial where the manipulation had elicited a target response that matched the primes (i.e., target responses that were meaning-based after being exposed to meaning-based primes and target responses that were position-based in the position-based priming condition) was given a score of 1. For example, if a subject placed in the position-based priming condition had responded to *limit* with *restriction* in Task 1 (a synonymous, meaning-based response) and then replied to

[4] As explained in Section 5.3.1 above, three NNS participants had not produced at least six position-based responses during Task 1. As experimental conditions could not be met (i.e., it was impossible to make a customized Task 2 form for these subjects), they were eliminated from the study.

[5] As demonstrated in a wealth of prior WA research findings, subjects (regardless of whether they are native or nonnative speakers) produce very few form-based responses (typically about 1% of total responses) and even fewer *others*. For this reason, it was decided a priori not to include either of these types when designing the experimental conditions.

the same cue with *time* (a *yx* collocation, position-based response) in one of the target trials in Task 2, the trial would be given a score of 1. This resulted in a *raw priming score* ranging from 0 to a maximum of 3 points for each subject. A score of 0 to 3 was also tallied for the responses to the control cues using the same criteria (i.e., a *raw control score*). A comparison of the means of these two scores across the cohort would thus act as a measure of whether the primes had had an effect (prime score > control score) or not (prime score ≤ control score).

The mean prime and control scores for both the NS and NNS groups (and their standard deviations) appear in Table 5.4 along with the results of a paired samples *t* test. The results show that there was no significant difference between the prime and control scores for NNS subjects (t = 1.085, p = .281, df = 78). This indicates than the priming manipulation had no significant impact on the NNS group. A statistically significant effect was found, however, in the comparison of prime and control scores for the NS group (t = 2.784, p = .007, df = 66) indicating that a priming effect had been induced in the case of NS subjects.

Table 5.4. The comparison of prime and control scores of the NS and NNS groups.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
NNS (*n* = 79)	0.97 (.97)	0.84 (.85)	1.085
NS (*n* = 67)	1.24 (.96)	0.85 (.82)	2.784**

** $p < .01$

A further analysis was conducted on the data from the NS subjects to determine whether this priming effect was attributable to just one or both of the two experimental conditions (meaning-based priming and position-based priming). As

can be seen in Table 5.5, a statistically significant difference (t = 2.440, p = .020, df = 35) can be seen between the mean prime score and the mean control score for the 36 NS subjects who were in the meaning-based priming condition. This indicates that when subjects were presented a series of primes to which they had responded with meaning-based responses in Task 1, they were likely to respond to subsequent cues with meaning-based responses despite having responded to these targets with position-based responses during Task 1. Table 5.5 also shows that the same effect was not observed in the results from the 31 NS subjects in the position-based priming condition (t = 1.366, p = .182, df = 30).

Table 5.5. The effect of meaning-based and position-based priming conditions on NS respondents.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
Meaning-based (*n* = 36)	1.47 (.94)	0.94 (.83)	2.440*
Position-based (*n* = 31)	0.97 (.91)	0.74 (.82)	1.366

* p < .05

To attempt to answer the second research question – whether any priming effects elicited were temporally related to the presentation of the primes (recency effects) or whether they were distributed relatively evenly across the three target responses – it was necessary to tally individual priming scores for each of the three target response positions. The results of these tallies showed no temporal effects in either the NS or NNS data. That is, the amount of priming remains relatively constant across the three target positions (from Target 1 to 3: NNS = 28, 24, 25; NS = 29, 29, 24) with no obvious decrease, either sudden or incremental. For ease of comparison, these figures are presented divided by *n* in Table 5.6.

Table 5.6. Priming effects as a function of target position.

	Target 1	Target 2	Target 3	Total
NNS (*n* = 79)	.35	.30	.32	.97
NS (*n* = 67)	.43	.43	.36	1.22
NS – meaning (*n* = 36)	.44	.47	.53	1.44
NS – position (*n* = 31)	.42	.39	.16	.97

Perhaps more relevant is the data obtained from NS subjects in the meaning-based priming condition specifically – the only condition to show a significant priming effect. Here too, however there is no obvious time-related effect. The 36 subjects in this condition made 16 meaning-based responses to the first target cue, 17 to the second, and 19 to the third. This absence of recency effects signals a variety of possibilities for the data. First of all, it is possible that recency effects were simply not observable over the course of just three target trials.[6] That is, all three targets may be showing the effects of their relatively close temporal relations to the primes, but had subsequent targets also been presented (Target 4, Target 5, etc.), the effect would diminish or disappear completely. It should be noted here that the decision to include three target trials was a somewhat arbitrary one. Had there only been one such trial, priming of the target may have been observed in more than just the NS/meaning-based cell. Likewise, if five or 10 target trials had been included, then it may have been possible to observe the attrition of the effect over a series of trials. In any case, given the time resources available to conduct this

[6] It is interesting to note that Table 5.6 does show some indication of a temporal effect in the NS responses from the position-based priming condition. While the priming effect seems stable over the first two targets (13 and 12 occurrences), the effect seems to have tapered off by the time subjects encounter Target 3 (only 5 occurrences). However interesting, this finding is presumably a spurious one, as NS respondents did not exhibit a significant amount of priming according to the analyses reported here.

study – and given that this particular combination of WA task and priming measure represents a new, and hence exploratory, methodology – three experimental trials seemed a justifiable starting point for this research. Of course, further studies exploring possible priming effects over varying numbers of experimental trials are warranted.

A second possibility is that the kind of priming measured here is rather subtle. That is, while it may or may not have a statistically significant impact on target cells immediately following the experimental manipulation, the influence of the primes may be affecting responses to specific cues throughout remaining trials. Only responses that are "vulnerable" to the effect would show change across tasks. Vulnerability in this case, could be determined by participant factors (e.g., the extent to which learners have knowledge of possible meaning-based or position-based links to the cues) or by aspects of the cues themselves (e.g., some cue words may simply have fewer links to other lexical items belonging to specific association categories, or something about the cues fails to elicit specific types of responses). I will return to these possibilities in the Discussion below.

5.5.4 Discussion

According to these findings, the answer to the research question, 'Do priming effects have a significant influence on WA responses?' is a qualified 'yes': priming effects may indeed impact word association responses under some circumstances. While priming was not demonstrated in the results of the NNS subjects, the NS participants – at least those in the meaning-based priming condition – were influenced by the experimental manipulation, changing their previously position-

based responses to meaning-based ones during the second task. While it is not clear that such effects would have a profound influence on all WA research studies, the fact that a priming effect was observed in at least one of the experimental conditions should be a cause for concern for researchers. The implication here is that cue order is a potentially confounding factor in WA research and that countermeasures are necessary. An easily implementable solution to this problem is to conduct WA research utilizing multiple forms for each WA task administered (as per Racine, 2008, 2011b, 2011c). In other words, while individual respondents' associations may still be subject to the influences inherent in the specific task forms to which they have been assigned, the impact of specific cue orders on overall group results will be diminished. I suggest that, in WA tasks administered on written forms, the use of even two differing cue orders within subject groups would be a means of eliminating the influence that any one specific cue order may have on the results. Of course, computer-based WA research utilizing cue presentation software easily allows researchers to randomize cue order for all subjects.

One of the interesting outcomes of the current study is the fact that the priming effect occurred only in the case of NS subjects, and only for those NS subjects in the meaning-based priming condition. One reason for this may be that, despite the efforts described above to select cues that the NNS participants would know, cue words were not firmly consolidated in the second-language lexicons of this group. I have commented in Chapter 4 (in relation to the criteria for cue selection; 4.5.2.1) and elsewhere (in relation to the measurement of productive L2 vocabulary; Racine, 2011a) about the instability of the developing lexicon. New lexical items

are constantly entering the lexicon, some to be permanently consolidated in memory, others to be forgotten soon afterwards. Some new items are particularly salient for a period of time, later to be replaced by other items of greater use to their storeowner. Aspects of word knowledge for each entry in the lexicon may undergo similar cycles of acquisition and loss. That is, a given meaning of a word may soon be forgotten, a certain collocation may become temporarily salient, etc. It is within this turbulent environment – affecting both knowledge of cues and of potential responses – that L2 WA research attempts to find its answers. Indeed, we may say that all L2 WA research results are fraught with noise that potentially masks or eliminates experimental effects.

Lexical processes may also be a determining factor in the absence of priming effects in the NNS respondents. There is reason to believe that a certain threshold of L2 proficiency must be attained before the processes involved in lexical access can occur automatically and with the degree of accuracy necessary to make them susceptible to priming effects. Indeed, the results of a number of studies involving bilingual subjects suggest that such automaticity of process is required to elicit both form and semantic priming (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Frenck-Mestre & Prince, 1997; Kroll & Stewart, 1994). Frenck-Mestre and Prince (1997), for example, were able to induce three types of semantic priming effects in native speakers and in proficient L2 learners. None of these priming effects could be induced in a group of less proficient L2 learners. The researchers described the results in terms of "autonomous" semantic processing, meaning that native speakers and highly proficient learners exhibit automatic and unstoppable activation when encountering a prime. This activation automatically spreads to

semantically related words, indicated by a priming effect when related targets are encountered.

The stability of native speakers' response profiles reported by Fitzpatrick (2006, 2007, 2009) and in the results presented in the previous chapter, is presumably generated from lexicons consisting of entries well-consolidated in memory, firm links between them, and accurate, primarily automatic processes of access. It is likely then, given the turmoil of the developing L2 lexicon, and in the absence of automatic semantic processes, that this is why priming effects could not be induced in this relatively non-fluent NNS group. As NS associations are derived from a stable, well-consolidated L1 lexicon, they may have proven less vulnerable to the influence of the experimental manipulation. This strength of cognitive style is at the heart of WA research involving response profiles. The "competition" between experimental manipulation and strength of response preference is a theme I will return to in Chapter 7.

5.5.4.1 Cue effects

The explanation above provides an account for the fact that NNS data did not display the effects of the priming conditions. But why is it that only the meaning-based priming condition led to a priming effect for the native-speaking subjects? The answer to this question may lie once again in the nature of the cues and their potential responses rather than in aspects of the subjects and their response profiles. As noted in the introduction to this chapter, a number of studies have shown links between cue type and response type. These include the effects of word class (e.g., Deese, 1962; Entwisle, 1966a), word frequency (Higginbotham, 2010), concreteness (e.g., Van Hell & De Groot, 1998), and a variety of other cue

characteristics. That said, it is important to note that the majority of responses from the vast majority of respondents in perhaps all published WA studies to date are meaning-based. Typically, when presented with a known word, the first thing that comes to mind is semantically related. Responses in the current study are no exception. As experimental cues utilized in Task 2 were selected based on the responses they elicited during Task 1, this factor has considerable implications for the experimental design utilized here.

One implication concerns the filler cues used in the control block. As fillers were chosen randomly from Task 1 cues, they would have already elicited many meaning-based responses from most subjects at that time. Therefore, there was an increased possibility that inadvertent 'priming' had affected responses to the three control cues. In other words, in the meaning-based priming condition – where control cues had elicited position-based responses during Task 1 – there would have been an increased likelihood of control responses being primed by the fillers, i.e., eliciting meaning-based responses during Task 2. However, since *priming* in this study is defined as a function of the difference between inadvertently 'primed' responses to the control cues and primed responses to the target cues, it is a testament to the robustness of the priming effect that it was found in the meaning-based priming condition for NS respondents at all – inadvertent priming would have reduced the chances of an effect. At the same time, in the position-based priming condition – where control and target cues had elicited meaning-based responses during Task 1 – the preponderance of cues that had elicited meaning-based responses during Task 1 among the fillers in the control block should have increased the likelihood that responses to control cues remained meaning-based in

Task 2. Since, once again, priming is operationalized here as the difference between inadvertently 'primed' responses to the control cues and primed responses to the target cues, one might expect a large priming effect to be observed in this condition. Yet there was no significant priming in this condition at all.

To make sense of these two seemingly contradictory findings it is necessary to re-examine the two aspects of the word association process that this chapter set out to compare: response profiles and cue effects. A "response profile", as described in the previous chapter, is the general response pattern, characteristic of an individual respondent, across WA tasks. This may be demonstrated over time (WA Task 1, WA Task 2, etc.) or across languages (L1 WA Task, L2 WA Task). "Cue effects" here refers to the tendency of a given WA stimulus to continue to produce the same type of response (meaning-based or position-based, etc.) regardless of individual differences among respondents or the broader contexts in which the research may take place. This may be referred to as *cue strength* and its existence is perhaps one of the underlying assumptions in the production, publication and ultimately the utility of WA norms lists such as Moss and Older's (1996), those included in the Postman and Keppel (1970) volume, and others.

A second look at the results for NS subjects described above allows us to compare these two elements of the WA process: the relative strength of respondent preferences (i.e., response profiles) and cue strength. Table 5.5 shows that there was more overall priming (as a result of the priming manipulation and inadvertently in the form of 'primed' control responses) observed in the results of

the subjects in the meaning-based condition (mean prime score of 1.47; mean control score of 0.94) than there was in the results from the position-based priming condition (0.97 and 0.74, respectively). If there had been no difference between the mean control scores for these two subject samples, this finding might imply that cue effect had no influence here. That is, regardless of what response type had been elicited by a given cue in a previous task, respondents were just as likely to change their response during a later task. The fact that more control responses became meaning-based in the meaning-based priming condition (M = 0.94) than became position-based in the position-based priming condition (M = 0.74) may be due to the relative strength of cues to elicit meaning-based responses. In other words, even under conditions where it would be difficult to detect priming – such as the control condition described here where filler cues had perhaps inadvertently elicited meaning-based responses during Task 1 – most WA stimuli continue to elicit meaning-based responses and, importantly, those that hadn't during an initial task, could be manipulated to do so in a subsequent one. This finding points to the relative strength of cues to elicit certain types of responses (mostly meaning-based) despite experimental manipulation. Generally speaking, this evidence makes a contribution toward our answer to the first research question: under some conditions at least, cue effects such as priming do influence individual responses despite the relative strength of individuals' response profiles.

5.5.4.2 Word knowledge

Another consideration that may account for this discrepancy is a lack of position-based word knowledge, namely, knowledge of collocations. This may certainly be the case for NNS subjects who, in previous research, appear to show deficits in

their collocational knowledge when compared to semantic knowledge for individual L2 words (e.g., Bahns & Eldaw, 1993; Barfield, 2006). But it is difficult to draw strong conclusions about subjects' knowledge of collocations in their first language. If it were possible to extensively quantify these types of lexical knowledge, it may prove to be the case that NS subjects simply have fewer collocations available for a given cue than they have meaning-based associative knowledge. For the purposes of WA research – that is, based on the response categorization schemes outlined in Appendix II – collocational knowledge (i.e., position-based responses) includes merely the words that may precede or follow the cue, either directly or indirectly, in a known phrase. Compared to meaning-based knowledge (which includes synonyms, along with all lexical set, conceptual, and contextual relatives), collocational associations appear to be a relatively small body of knowledge from which respondents may draw their responses during WA tasks. If the relative 'quantities' of these types of word knowledge could be measured empirically, it might be possible to explain the differences in NS responses described above. Indeed, this may account for the predominance of meaning-based responses in almost all WA studies to date. Another way of framing this account is that the percentage of associations of each type may ultimately be a product of the percentage of available, known words relating to the cue words in that way.

While this explanation is intriguing, there has been relatively little research conducted to quantify these types of word knowledge thus far. Schmitt (2010) acknowledges this lack of empirical evidence, but speculates that collocational knowledge is likely acquired only after a word's spoken and written form, meaning,

and grammatical characteristics have already been learned. Even when a vocabulary item is "mastered", collocational knowledge may still lag behind most other types of lexical knowledge. The volume of word knowledge available to a respondent may then be subject to an order-of-acquisition effect. Further studies – both L1 and L2 studies – may help to more conclusively assess the account for the findings I have described above.

5.5.5 Limitations and further research

One minor issue for this study concerns the decision to select cues based on the primary responses of participants in previous research as reported in established WA norms lists (such as H. Moss & Older, 1996). Word association norms lists, particularly position-based response data, may reflect subjects' culture-specific knowledge. That is, they are likely to include references that are well known and temporally salient among the native-speaking population from whom the normative data was collected. Such references may not necessarily be understood outside of that population. Higginbotham et al. (2015) provide the example of *submarine* to which the primary response in Moss and Older's (1996) norms list is *yellow*. This is a reference to the title of a Beatles song and subsequent animated film. While The Beatles remain relatively well known in Japan, it is difficult to know precisely how associated these two words might be by the NNS group sampled here in comparison to the UK-based NS group. As the Moss and Older norms list was created from the responses of subjects in the UK, this may be a potential confound. As many position-based responses, in particular, are based on cultural references – including famous quotations, song titles, band names, slogans from recent television commercials, etc. – it is difficult to predict the relative

contribution of this kind of word knowledge to the responses of subject populations from different cultural backgrounds or, indeed, ages. Researchers attempting to select stimuli to which all experimental groups might be able to produce a wide variety of primary responses may wish to pay close consideration to specific responses within the original norms lists.[7]

As described above, a more serious concern for this study is the fact that a preponderance of Task 1 responses was meaning-based. The filler cues preceding the control cues, while randomly selected from Task 1 cues, did not feature items that had elicited a random distribution of meaning-, position-, and form-based responses from their respective subjects. Indeed, some subjects may have received 10 filler cues to which they had responded with only meaning-based responses during task 1. One way of alleviating this problem in future studies would be to control for this by intentionally selecting filler cues that had produced a variety of response types during the first task (i.e., equal numbers of cues that had produced each of meaning-based, position-based, and form-based responses) and then randomly order them within the control block in Task 2. This solution, however, is still somewhat problematic for two reasons. First of all, it may be the case that only the final filler items in the block have an effect on the subsequent control responses. Depending then on the type of response that Filler item 10 elicited in Task 1, inadvertent priming may influence subjects' response processes during Task 2. There is also a second issue that would not be resolved by implementing

[7] My colleagues and I (Higginbotham et al., 2015; Racine et al., 2014) also argue for the use of only the most recently generated WA norms. Just as responses will differ between cultures (e.g., Son et al., 2014), responses collected in different years will also differ, as age cohorts may have different associations.

the proposed solution: the methods implemented here do not guarantee that subjects respond to the cues in the prime block as they did in Task 1. While the relative strength of individual response profiles was demonstrated in the previous chapter and elsewhere (e.g., Fitzpatrick, 2007, 2009), there is never a 100% guarantee that a given response – or even response type – will be elicited from the same respondent by the same cue during a second encounter. Fitzpatrick et al. (2015, p. 38) report that only about 25% of words given as responses were repeated on a second task administered three months later. For this reason, future investigations of the relation between priming and WA should be based around a methodology that better guarantees that the experimental manipulation has actually taken place.

Based on these concerns, a modified method might be beneficial. One possibility is to no longer make the experimental manipulation in Task 2 completely dependent upon responses from the first task. As an alternative, Task 2 could include a series of stimuli to which subjects are instructed to provide a specific type of response (see the discussion of restricted WA tasks in 2.3.2). This would ensure that respondents had actually engaged in the type of processing necessary to meet the conditions of the experiment. Responses to subsequent cues – to which participants would respond in the traditional manner (i.e., with the first response that came to mind) – would then indicate the influence of the restricted WA task, if any. In either case, researchers would be confident in the validity of the experimental manipulation – that participants were in fact subject to the correct experimental conditions. Two such experiments will be presented in the next chapter.

5.6 Summary

In this study, a priming effect was discovered for NS participants who had been assigned to the meaning-based priming condition. For this reason, it was suggested that cue order can be a potential influence and that researchers should take that into account when designing WA studies. The same effect was not found for NS subjects in the position-based priming condition. Nor was it found for NNS subjects in either of the experimental conditions. The latter finding was attributed to a lack of stability in L2 word knowledge and a lack of fluency in the lexical processes of the language learners. A closer examination of the method revealed that the desired experimental manipulation may not have been achieved and a follow-up study has been suggested to address the issue. Two experiments based on this suggestion will be presented in the in the next chapter.

Chapter 6. Restricted associations as procedural primes

6.1 Introduction

It should be clear from the evidence presented in the previous chapter that cue order is an aspect of research methodology that deserves thorough consideration when designing WA studies. Depending on the experimental conditions employed, cue order could impact upon subjects' responses – hence, upon their response profiles – and render a skewed view of their cognitive styles. It should also be clear that aspects of the WA process can be primed just as can aspects of semantic knowledge, forms, phonology and other elements of language (see 5.3). As I pointed out, however, what is not clear from the results of the prior study is whether the attempted manipulation – to prompt subjects to repeatedly engage in a specific type of associative process (meaning- or position-based) – actually worked (see 5.5.5).

The study introduced in this chapter therefore revisits the experimental design to explore potential improvements to the method for inducing priming effects. The aim of these improvements is to increase the level of confidence we can have in the findings as reliable indicators of the hypothesised processes. The research questions here continue to probe the extent to which priming manipulations can move responses from those which accord with participants' own response preferences to those we would expect to accord with the priming manipulation. Once again, differences between NS and NNS are recorded and analysed. To address the potential unreliability inherent to the priming methods adopted in the

prior study, the current study is redesigned as a 'restricted' word association experiment.

6.2 Restricted word association

Restricted word association (e.g., Miron & Wolfe, 1964; Ramsey, 1981; Riegel et al., 1967; Riegel & Zivian, 1972) – also referred to as *controlled* association (e.g., De Deyne & Storms, 2008; Smolentseva, Sozinova, Vasenina, & Levin, 2014) or *directed* association (Francis, 1972) – differs from conventional (or *free*) association methods in that responses are not elicited from subjects by asking them to simply "respond with the first thing that comes to mind". As in Riegel et al. (1967; reviewed in Chapter 2), respondents may be required to perform any of a variety of lexical or semantic tasks (e.g., providing a response that has a superordinate or coordinate relation with the cue, or identifying a function or quality of the object represented by the cue). As in free association, restricted association is implemented in order to observe links between stimuli and responses, and from observing these links, to infer connections within the mental lexicons of respondents. Unlike free association, however, responses in restricted association studies are likely to be elicited via conscious effort on the part of subjects. Responses (or failures to respond) may thus be evidence of what *can* be elicited, rather than what simply happens to be elicited, as in conventional WA methods. The manner in which restricted association trials will be implemented in the study presented below, however, is quite different from that of, for example, Riegel and his colleagues (Riegel et al., 1967; see 2.3.2). The current study uses

restricted associations as a means of inducing the experimental manipulation – i.e., to create a priming condition.

Essentially, as in the priming study in Chapter 5, a free association task (Task 1) is administered as a means of establishing subjects' response preferences to specific cues. A number of weeks later, a second WA task is administered, using a combination of the same cues and some new ones. This Task 2 includes a set of restricted association trials to see if they direct the participants towards specific association types in subsequent free-choice trials. Changes in response types to the same cues across tasks can be therefore be attributed to the manipulation.

6.3 Method

6.3.1 Participants and cues

NS subjects in this study were undergraduate students studying at a university in the United Kingdom. NNS subjects were Japanese students studying English at a university in Japan. As in the prior study, subjects who failed to complete the task correctly, or who did not complete at least two thirds of the Task 1 trials, were removed from the data. This resulted in the elimination of two NS subjects' and eight NNS subjects' data, yielding a sample of 53 NS subjects and 85 NNS subjects.

The 100 cues for Task 1 were identical to those utilized in Task 1 of the study reported in the previous chapter. That is, they were selected such that 1) the majority of primary responses would not belong to a single response category (according to the norms data from H. Moss & Older, 1996), and 2) they would be

known to the majority of NNS subjects.[1] As in the study described in the previous chapter, two versions of the task sheets were created in each of the native languages of the subject groups with the cues appearing in two different randomly selected orders. One of the English forms appears in Appendix V.

6.3.2 Task 1 Procedure

The tasks were completed at the subjects' respective universities (i.e., in the UK for the NS group; in Japan for the NNS group). Respondents were randomly assigned to receive one of the two Task 1 forms (i.e., each with a different randomly selected cue order). No time limit was set, and all subjects were able to complete the task within 30 minutes (20 minutes for the NS group). Responses were categorized, as in the previous study, for their reuse in Task 2. As this would be the first application of the categorization scheme developed from the results of the study described in Chapter 3 (i.e., the one presented in Appendix II) to a restricted association task, the decision was made to begin by categorizing responses using only the four main WA categories (*meaning-based, position-based, form-based responses,* and *others*), rather than the more finely detailed subcategory classification. It was assumed that significant, yet subtle, findings might not be observable using the more detailed classification scheme.[2] It was reasoned that, depending on the results of this study, the more detailed category/subcategory categorization scheme could be applied to measure the influence of restricted association tasks in subsequent studies.

[1] See Section 5.5.2.2 for a detailed explanation of the cue selection process.
[2] See the discussion concerning categories and effect sizes in Sections 3.5 and 3.6.

6.3.3 Task 2 Procedure

Task 2 was administered six to eight weeks after the first task. As with the experiment reported in the last chapter, subjects were randomly selected to be in one of two conditions. One of these was a *meaning-based priming* condition, in which they received a meaning-based restricted association task. Here, experimental trials (i.e., the target and control trials) consisted of cues to which they had responded with position-based responses in Task 1. As in the study reported in the preceding chapter, this aspect of the experimental design yields data that allows a comparison between random change in response types over time (the control trials) and change believed to be due to the experimental manipulation (target trials). The other condition was a *position-based priming* condition (where subjects received a position-based restricted association task) in which the experimental cues were those to which they had responded with meaning-based responses in Task 1.

The role of the restricted association trials in this procedure can be seen in the description of the meaning-based priming condition outlined in Table 6.1. As illustrated, Task 2 consists of two parts. Part 1 is a meaning-based, restricted association task. This is followed by a free association task in Part 2 in which the experimental and control cues are embedded. The experimental cues (in this case, cues that elicited position-based responses during Task 1) immediately follow the restricted association task. The prediction is that if subjects have repeatedly responded with meaning-based responses during the restricted task, a priming effect will be observable in these trials (i.e., responses to these cues which had been position-based in Task 1, may now be meaning-based). The experimental

trials are followed by 20 more free association trials. These trials are intended to reduce the influence of the priming manipulation on the subsequent control trials. As in the experiment reported in the preceding chapter, the priming effect is measured via the comparison of changes in response between the experimental and control trials. An example of the English form for this condition appears in Appendix VII.

Table 6.1. Task 2 research design and trial presentation order for the experimental-first, meaning-based priming condition.

	Type of Task	Number of Trials	Type of Cues
Part 1	Restricted Association Task (meaning-based)	20 trials	New cues (not seen in Task 1)
Part 2	Free Association Task	3 experimental trials	Cues that elicited position-based responses in Task 1
		20 trials	New cues
		3 control trials	Cues that elicited position-based responses in Task 1

The research design described in Table 6.1 is referred to as the *experimental-first*, meaning-based priming condition because the experimental cues appear before the control cues. To ensure that any effects caused by the restricted association task did not influence responses to the control trials, the design was counter-balanced with a *control-first* condition. This reorganization of the elements results in a three-part, rather than a two-part structure. This tripartite design is illustrated

in Table 6.2. Subjects in this condition receive three control trials embedded within a free association task in Part 1. They then take part in the restricted association task. This is followed immediately by the experimental trials, embedded in a second free association task (Part 3). Responses to the final 17 trials in Part 3 were inconsequential to the data collected in this study. These trials merely appeared on the page to camouflage the importance of the three experimental trials that preceded them. In effect, they were to draw participants' attention from the three experimental cues which would have been overly salient had they appeared alone as the final section of the task. How they appeared to participants can be seen in the English version of the form in Appendix VIII.

Table 6.2. Task 2 methodology and trial presentation order for the control-first, meaning-based priming condition.

	Type of Task	Number of Trials	Type of Cues
Part 1	Free Association Task	20 trials	New cues (not seen in Task 1)
		3 control trials	Cues that elicited position-based responses in Task 1
Part 2	Restricted Association Task (meaning-based)	20 trials	New cues
Part 3	Free Association Task	3 experimental trials	Cues that elicited position-based responses in Task 1
		(17 trials)	(New Cues)

As explained above, the experimental-first and control-first conditions are necessary to counter-balance each other. In the experimental-first condition (Table

6.1, Part 2), a block of 20 distractor trials consisting of cues to which subjects have not been previously exposed is used to create distance between the priming task and the subsequent control trials, to minimize the risk of them being influenced by that priming. In the control-first condition, there is, of course, no risk of this. However, these 20 distractor trials need to be included in the design, to maintain balance in task length and difficulty across conditions. They are retained in their position immediately before the control trials (Table 6.2, Part 1), so that any potential influence of them in the experimental-first condition is replicated in the control-first condition.

Participants were assigned randomly within a 2 x 2 matrix of conditions (meaning-based or position-based priming, and experimental-first or control-first task orders). Instructions for the free association tasks were essentially the same as those that the participants had encountered in Task 1 six to eight weeks earlier. As can be seen in Appendix VII and VIII, participants in the meaning-based condition received the following instructions for the restricted association task:

> For each word below, write a related word. For example, if the word is *cat*, you might write *pet, animal*, or *lion*. Any answer is okay as long as it is related to the meaning of the word printed here. Write only one word for each.

In the position-based restricted association task, subjects received the following instructions:

> For each word below, write a word that comes before or after that word. For example, if the word was *dog*, you might write *hot* (as in *hot dog*)*, or house* (as in *doghouse*). Any answer is okay as long as it might appear before or after the word printed here. Write only one word for each.

All instructions appeared in the individual subject's first language.

6.4 Results

As in the study presented in the previous chapter, changes in the responses to the control cues served as a measure of general instability in response preferences over time. It is the comparison of changes between the target trials and the control trials that determines the extent to which the restricted task has influenced response preferences. Therefore, as in the previous study, the results were quantified by calculating a score of 0 to 3 for each set of experimental trial responses. That is, a score of 1 was assigned to each experimental trial where a response was elicited in accordance with the priming manipulation (i.e., cues that had elicited meaning-based responses in Task 1 later elicited position-based responses in Task 2 and cues that had elicited position-based responses in Task 1 later elicited position-based responses in Task 2). Thus, for those in the meaning-based condition, if they gave meaning-based responses for all three experimental cues it would yield a raw priming score of 3, suggesting that priming may have occurred. The same assignment was used for the control trials, resulting in a raw control score ranging from 0 to 3 points each for each subject. It is the comparison of these two scores that determines if priming has actually occurred. In other words, support for the hypothesis relies on the experimental cues generating significantly higher scores across participants than had control cues.

To determine whether or not this difference existed, a paired-samples *t* test was conducted to compare the two scores for both NS and NNS participants. As the results in Table 6.3 indicate, both subject groups exhibited higher mean

experimental scores than control scores. While the results of the *t* test approached significance for the NNS group (p = .065), indicating that some kind of priming effect may have occurred, the results of both tests failed to show statistically significant effects.

Table 6.3. Comparison of prime and control scores of the NS and NNS groups.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
NNS (n = 85)	0.95 (.86)	0.75 (.84)	1.871*
NS (n = 53)	1.34 (.92)	1.08 (.98)	1.729*

* non-significant effects (p < .05)

Table 6.4 provides a closer examination of the NNS respondents' scores. There are larger prime scores than control scores in both the meaning-based and position-based priming conditions. The *t* test confirms that the difference in the position-based condition is a significant one (t = 2.142, p < .05, df = 42). As can be seen in Table 6.5, these findings are paralleled in the results of the NS group's *t* tests. Here too, higher prime scores than control scores are seen in both conditions. Again, there is a significant priming effect in the case of the position-based condition (t = 2.060, p < .05, df = 26).

Table 6.4. The effect of meaning-based and position-based priming conditions on NNS respondents.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
Meaning-based (n = 42)	1.17 (.96)	1.05 (.94)	0.696
Position-based (n = 43)	0.74 (.69)	0.47 (.63)	2.142*

* p < .05

Table 6.5. The effect of meaning-based and position-based priming conditions on NS respondents.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
Meaning-based (*n* = 26)	1.65 (.89)	1.58 (.99)	0.359
Position-based (*n* = 27)	1.04 (.85)	0.59 (.69)	2.060*

* p < .05

6.5 Discussion

The significant effect reported here – that both NS and NNS subjects' meaning-based responses were susceptible to position-based priming – does not support the findings of the study reported in the previous chapter. Those results had shown that only NS subjects' responses were significantly influenced by priming conditions and only in the meaning-based priming condition. What might account for the difference between these two sets of results? One possibility is that the previous design was unreliable in achieving a priming effect, while the present design was successful, revealing here a pattern regarding the relative susceptibility to priming of meaning versus position-based responses. Indeed, as I argued in the discussion section of the previous chapter, there is reason to believe that the method I adopted in that study mistakenly links the very existence of the priming manipulation (an independent variable) to response stability over time (a dependent variable). While there is reason to believe WA responses remain relatively stable over time or across tasks (e.g., Fitzpatrick, 2007; Fitzpatrick & Playfoot, 2011; Fitzpatrick et al., 2015; Higginbotham, 2010), there is no data suggesting that specific responses or even response types will remain *identical* given the same stimuli at different times. For this reason, it is fair to say that

subjects in the second task of the prior study did not engage consistently in the types of processes they had when receiving the same cues in the first WA task. In other words, those subjects had not consistently experienced the priming manipulation and, for this reason, the current study represents an improvement on the prior method.

Another interpretation of the disparity between these two sets of findings is that priming effects are too weak to alter WA responses consistently or that they are simply too subtle to be reliably captured by the current methods. In either case, it is difficult to predict with certainty under what conditions a priming effect might be observed or whether one will be observed at all. This subtlety of effect returns us to a recurring theme in this research: the notions of response preference and cognitive styles (see, Section 1.5 and Chapter 4). Indeed, we may choose to characterize "choice" of response in the current studies as the result of competition between subjects' natural preference to respond with a specific type of response and the tendency to respond in accordance with the experimental manipulation – in this case, priming. While it is not feasible to fully address this cognitive style vs. external influence hypothesis here, it is a theme I will return to in the next chapter while attempting to model the WA process.

Accepting the assumption that the current findings are more reliable than the data from the experiment reported in the last chapter (i.e., these were not subject to the methodological flaws discussed in the previous chapter) – then it becomes quite difficult to account for them. Subjects' meaning-based responses – within both the NS and the NNS groups – seem to be susceptible to position-based priming. This

runs contrary to the argument that meaning is at the core of associative memory. Such a semantic "default setting" is how I accounted for the findings reported in the last chapter. My argument was that meaning-based links are central to the lexicon. "Centrality" here may be defined in terms of their salience or ease of activation among other types of links, or simply in terms of the sheer numbers of connections. In either case, this would account for the fact that position-based responses were, to an extent, influenced by meaning-based priming methods. It also accounts for the fact that the majority of all word associations (elicited from adults at least) are meaning-based.

This explanation, however, does not account well for the current finding – that subjects' meaning-based responses were susceptible to priming, changing into position-based responses, while the reverse was not true (at least not at a statistically significant level). In other words, if most links in the mental lexicon (or the strongest/most salient ones) are semantic in nature, then we would expect specific cues that elicited meaning-based responses to continue to do so regardless of any influence the priming conditions might have on the association process. Just as the results reported in the previous chapter begged a re-examination of the methods employed, in light of the findings reported here, similar scrutiny should be applied now to the current method. In short, perhaps the revised method used in this study did not satisfactorily solve the methodological problems of the previous one.

In the current study, the priming conditions were achieved via the restricted instruction tasks. Like most other experimental tasks, the results were dependent

upon subjects' ability to understand the instructions and complete the task as directed. As we have no reason to believe that subjects did not understand how to complete the task here, it may be the case that the instructions themselves are at fault. For ease of reference, I will duplicate the instructions for the restricted association tasks here from the Method section above. These are the instructions for the meaning-based priming condition:

> For each word below, write a related word. For example, if the word is *cat*, you might write *pet, animal*, or *lion*. Any answer is okay as long as it is related to the meaning of the word printed here. Write only one word for each.

and for the position-based priming condition:

> For each word below, write a word that comes before or after that word. For example, if the word was *dog*, you might write *hot* (as in *hot dog*)*, or house* (as in *doghouse*). Any answer is okay as long as it might appear before or after the word printed here. Write only one word for each.

The instructions in the case of the meaning-based condition are quite broad: "Any answer is okay as long as it is related to the meaning of the word." Indeed, the passage begins even more broadly, asking participants to merely write a "related" word. The purpose of this set of instructions was to get participants to engage in meaning-based association processes specifically. One must consider the possibility that at least some subjects may have associated "freely" (i.e., thought of the first word that came to mind) and then evaluated their responses in terms of whether it met the criteria specified in the instructions. Others may have intentionally engaged in a specific meaning-based process. For example, some may have adopted a strategy of responding with synonyms – or possibly antonyms, superordinates, or hyponyms, etc. – as these types of responses are semantically-

related to their cues. Given the many types of response these instructions may have elicited, it is difficult to say what the data in this condition really represent.

By contrast, the position-based priming instructions are quite specific; requiring respondents to think of a collocate that might precede or follow each cue. While more specific than the meaning-based instructions, here too participants may have adopted one or both of two particular strategies in order to fulfil the instructions. Once again, we cannot be certain of what the data really represent here. It would seem that the experimental instructions did not lead respondents to adopt one specific strategy to complete the restricted association tasks. Moreover, this lack of specificity was unequal across conditions. That is, instructions to respondents in the position-based priming condition appear to prompt the use of two response strategies, while those of the meaning-based condition seem to lead to the adoption of any number of semantic association strategies.

To rule out the possibility that the current findings were unduly influenced by this lack of specificity, a second experiment was conducted employing new instructions.

6.6 Supplementary experiment: Calibrating conditions

Participants in the supplementary study were selected from the same populations as in the experiment above. That is, NS were students from a university in the UK (n = 60) and NNS were Japanese students of English at a university in Japan (n = 38). None of them had taken part in the initial experiment. The same cue words

were utilized in the two tasks and WA trials were ordered according to the same counterbalanced research design described in Tables 6.1 and 6.2.

The main differences between this experiment and the one reported above were in the instructions. Rather than attempting to elicit semantically related responses in general in the meaning-based priming condition, the new instructions specified that respondents should provide a synonym for each cue:

> For each word below, write a word that has the same or similar meaning. For example, if the word is *hello*, you might write *hi*. For a word like *man*, you might write a word like *boy* or *male*. Any answer is okay as long as it has a similar meaning as the word printed here. Write only one word for each.

Similarly, the instructions for the position-based priming instructions – which had elicited both *xy* and *yx* collocations in the prior study – were adapted to ensure that only *xy* collocations would be elicited here:

> For each word below, write a word that might come after that word. For example, if the word was *sun*, you might write *glasses* (as in *sunglasses*). If the word was *hot* you might write *dog* (as in *hot dog*). Any answer is okay as long as it might appear after the word printed here. Write only one word for each.

Despite these changes, the results mirrored those of the initial restricted association study reported above and the cue-order study from the previous chapter, at least to an extent. While some amount of priming occurred for both NNS and NS subject groups – in the sense that mean prime scores are larger than their corresponding control scores (Table 6.6) – the differences between them were not statistically significant. Tables 6.7 and 6.8 show that this was the case for both NS and NNS subject groups.

Table 6.6. Comparison of prime and control scores of the NS and NNS groups.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
NNS (*n* = 38)	0.50 (.76)	0.39 (.55)	0.850
NS (*n* = 60)	0.48 (.68)	0.30 (.98)	1.794

Table 6.7. The effect of synonym-based and position-based priming conditions on NNS respondents.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
Synonym-based (*n* = 21)	0.48 (.75)	0.29 (.46)	1.284
Position-based (*n* = 17)	0.53 (.80)	0.53 (.62)	0.000

Table 6.8. The effect of synonym-based and position-based priming conditions on NS respondents.

	Mean Prime Score (*SD*)	Mean Control Score (*SD*)	*t* value
Synonym-based (*n* = 37)	0.32 (.48)	0.22 (.42)	1.071
Position-based (*n* = 23)	0.74 (.86)	0.43 (.73)	1.432

6.7 Discussion: Methodology and further research

The aim of the studies presented here and in the previous chapter was to examine the effects of priming methods on response types in L2 WA tasks. The reason for the various experimental designs employed was to establish an original and reliable method of achieving a priming manipulation. In discussing the results of the study in the previous chapter, I argued that the experimental manipulation failed to reliably produce priming conditions at all. Having thus called into question that study's concept validity (i.e., that the method may not have measured

what was intended), the restricted association tasks described in this chapter offer a demonstrable improvement in experimental design. Although these methods failed to yield significant results (at least in the supplementary study), we can feel confident that subjects actually engaged in the kinds of processing demanded by the task.

One issue that has not yet been considered concerns an artefact of the restricted association task methodology, namely, the placement of the task instructions. The instructions for Part 2 in the experimental–first condition (see Appendix VII) and the instructions for Part 3 in the control–first condition (Appendix VIII) must intervene between the priming trials and the experimental trials. One can see how their placement, while necessary to the experimental design, may dilute the impact of the restricted task on subsequent trials. Perhaps reading these instructions, their existence on the form, the turn of the page, etc. are enough to "reset" subjects' approach to subsequent trials. The impact of the priming manipulation then may be too weak to overcome this "instruction effect".

While statistically significant priming effects were not observed in the current studies, it is important to recognize that in all cells in all three experiments (reported here and in Chapter 5) the *direction* of the priming effect is positive. That is, for both NS and NNS subjects, in both the meaning-based/synonym-based conditions and in the position-based conditions, the mean control scores are lower or equal to the mean prime scores (see Tables 5.5, 6.4, 6.5, 6.7, and 6.8). From these findings, it is reasonable to conclude that the priming manipulations did have an impact on response preference – albeit a subtle one.

It is also reasonable to propose that the priming methods used here for the first time might be refined in future studies to produce stronger or weaker priming effects. I have already noted one variable in particular that may be adjusted in further studies: the number of target cues. As I stated in the previous chapter, the decision to examine differences in response between three control trials and three target trials was somewhat arbitrary. If priming effects in word association are subtle but long-lasting, we may more easily detect them through the analysis of a greater number of target trials. On the other hand, if the effects of priming on WA are short-lived, then perhaps a single trial is enough to detect it. The inclusion of additional trials in the analysis would serve only to dilute the effect. Further experimentation may provide evidence for increasing or decreasing the number of experimental trials in subsequent studies.

There remains another, more traditional methodological option for further studies on the effects of priming on WA: to adopt reaction time as the measurement of priming. It may be the case that the priming conditions created in this study do not significantly affect the content (i.e., the category) of one's WA responses, but that the production (i.e., response latency) of responses is hindered by the repeated activation of the priming procedures. If this is the case, interference created by the priming conditions would be observable in longer reaction times even if response types were not altered. During trials where responses do appear to have changed in accordance with the priming manipulation, RTs may also reveal the effects of facilitation where participants respond more quickly to primed cues than they would during baseline (i.e., non-primed) trials. This type of experimental design

might provide insight into the conflict that I have suggested exists between response preferences and priming manipulations. A long RT may indicate the existence of such a conflict, where the response is the "choice" between primed or naturally preferred alternatives, and the extended RT includes the processing time required to "decide" between them. The principles behind this idea will be explored in greater depth in the chapter below.

Finally, it is important to note that the interpretations of the results of this study and those reported in the previous chapter are all based on the premise that respondents are uniform in their approach to the task. An alternative interpretation would be that some respondents are more susceptible to priming manipulations than others. The relatively large standard deviation scores in the tables above suggest that such variation may have existed within subject groups in the current studies.[3] In other words, some respondents exhibited a large amount of priming (e.g., scoring 3 on the priming score and only 0 or 1 on the control score), while others displayed no such effect. Thus, results from each type of respondent essentially cancel each other out, rendering the overall results statistically non-significant. If this is the case, further research may reveal that significant priming effects are observable when subject groups reach a particular size. Another interesting possibility for this line of research is to determine precisely what factors relate to individual subjects' susceptibility to priming and other experimental manipulations. Such studies might reveal that there are specific "types" of respondent who either continue to respond in accordance with their

[3] In fact, there is an accumulating body of evidence that there is great diversity even within NS subject groups. See Racine et al. (2014) for a review of the literature.

own cognitive styles, despite manipulation, or veer from their characteristic responses in accordance with that manipulation. Further research may help to pinpoint what types of personality or cognitive factors determine this susceptibility.

6.8 Conclusion

In the discussion above I have introduced some suggestions about how further fine-tuning of these methods might take us closer to observing reliable priming effects. What is becoming clear is that there is probably a complex relationship between the responses actually given by participants in experiments and several other factors, including their underlying response preferences (i.e., their cognitive styles), and the precise presentation of the cues and instructions. Furthermore, because priming is by definition a lasting effect, the number of cues, the quantity of distractor material and the order of presentation could all potentially play a role in how individuals respond. In other words, these are dynamic variables that are difficult to control. The final chapter explores in more depth the nature of these relationships and their implications for other aspects of WA experimental design and theory.

Chapter 7. Modelling the determinants of word association

7.1 Back to basics

At the start of this thesis, a simple definition of word association was given: the presentation of cues and the recording of responses. However, as indicated in the results presented in the previous chapters, there are numerous assumptions intrinsic to this elementary model and there is much more to be added in light of the current findings. In this chapter, I will re-examine the basic elements of the WA process, summarize the conclusions drawn from the current findings concerning each of these elements, and identify some lessons relevant to future research. Finally, I will bring these conclusions together to form the foundation of a wholly new model of the WA process. offer some caveats for future research,

7.2 Elements of the WA process

7.2.1 Cues

I focused a great deal of attention on the selection of cue words in my empirical studies reported above. Indeed, it may be axiomatic that researchers should carefully consider which cues they present to research participants and how they present them. Conclusions based on the current findings provide support for this long-held assumption:

1. Linguistic features of cues (semantic, collocational, formal, etc.) are a determining factor in the types of responses they elicit. Possible responses are therefore restricted to words that relate to cues in these specific ways.
2. Frequency and difficulty of cues also determine types of responses.

The first conclusion refers to cue strength. As I have stated in Chapter 5, if we wish to explore the architecture of the mental lexicon via WA research, we should choose cues that do not have strong primary responses (e.g., *king-queen, black-white*). The reason being, as Meara (1982) pointed out, that such cues represent "core" vocabulary, typically learned in the early days of acquiring a first or second language, and hence tell us very little about newly acquired words in the periphery of the developing lexicon. They and their primary responses are strongly linked associations in the mental lexicon and we find very little variation in possible responses. However, cue strength does not refer only to the degree to which cues elicit strong primary responses. More broadly, it refers to any limiting effect that linguistic features of the cue may have on potential responses. Figure 7.1 illustrates the relationship between a cue and possible responses. In the figure, potential responses are shown to be restricted to those that are related either semantically (meaning-based responses), collocationally (position-based), or formally (form-based) to the cue. If this diagram were proportionally representative, the shaded area on the right side (i.e., potential responses) would be smaller if the cue had, for example, fewer synonyms, antonyms, rhymes, or if it existed in fewer formulaic sequences.

Note that the cue itself does not exist entirely within the field representing the respondent's word knowledge. This implies that there are aspects of word knowledge for this cue that the respondent does not know. In this way, there will be a certain degree of proportionality between the shaded areas on the left and right of the figure. The more aspects of the cue the respondent knows (represented by the extent to which the cue bubble is shaded on the left of the figure), the more

potential responses are at her disposal (represented by more shading on the right). By the same token, if the cue were completely unknown to the subject, it would be located outside of the respondent's knowledge field. Thus, there would be no shading on the left and no potential responses on the right. In this way, we can say that the second conclusion – that cue frequency and difficulty are factors that also contribute to the determination of responses – is also illustrated in the figure. As above, these factors are represented as respondent knowledge. The extent to which a subject knows the cues and has potential responses at her disposal is often determined by these words' frequencies and levels of difficulty.[1] Thus, regardless of the existence of many potential responses within the target language as a whole, the limits of subjects' word knowledge places further restrictions on the number of possible responses. This is another factor determining WA responses.

In fact, there are many characteristics of words that may influence the perception and comprehension of cues and the production of responses. While it is beyond the scope of this dissertation to treat these other factors in depth, readers are referred to a plethora of psycholinguistic literature investigating these word features. Among these are studies taking word length into account (e.g., Elgort, 2011; New, Ferrand, Pallier, & Brysbaert, 2006). It has been demonstrated that words greater than eight letters in length may be beyond the threshold of visual acuity (New et al., 2006). One can imagine how this might influence the salience of cues' formal features during WA trials. Presumably, longer words would need to be processed

[1] Research into lexical frequency profiles, while focused primarily on written vocabulary production, shows that lexical knowledge and use, broadly speaking, correspond with word frequency. See, for example, Laufer (1994, 2005); Laufer and Nation (1995); Meara (2005); and Muncie (2002).

piece by piece, possibly letter by letter. If so, this increased attention to form might increase the likelihood of form-based responses in comparison to shorter cues.

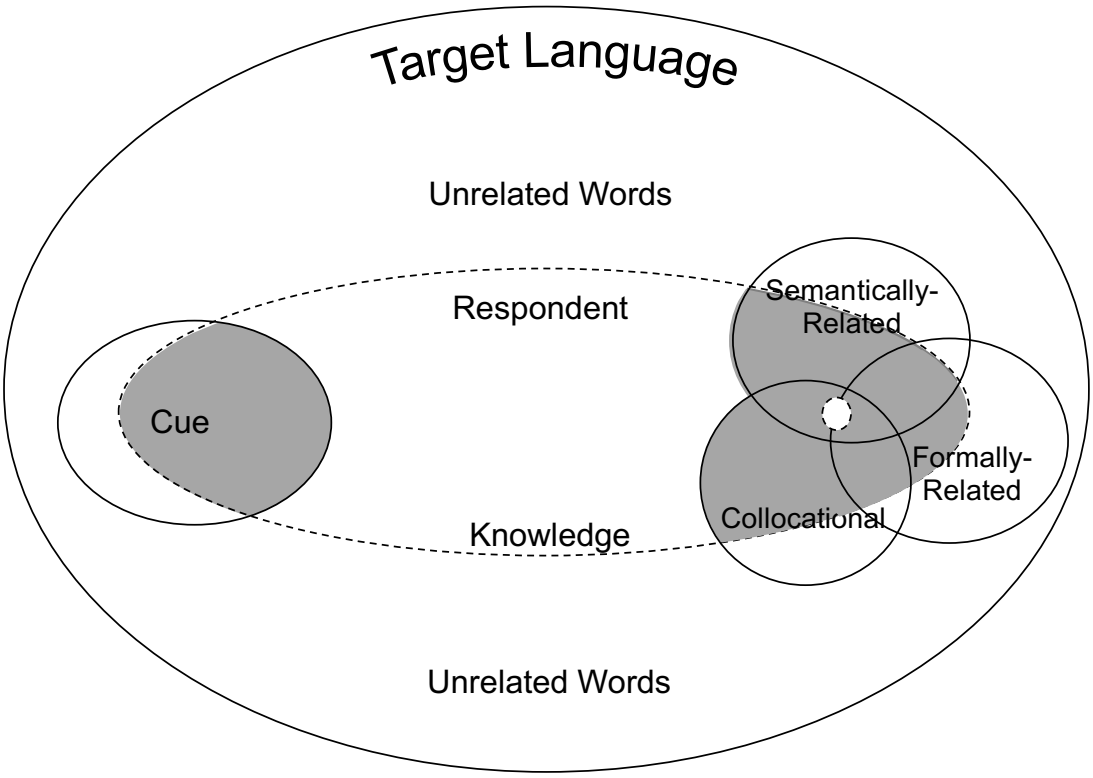


Figure 7.1. A partially-known cue and potential responses within the contexts of respondent knowledge and the language as a whole.

Another consideration for cue selection in future WA studies is the existence of orthographic neighbours. These are words that differ from the cue word in regard to only one letter. For example, the cue *food* has nine orthographic neighbours (*fool, fold, good*, etc.), whereas *donation* and *soldier* have none (see Masterson, Stuart, Dixon, & Lovejoy, 2010). Forster and his colleagues (Forster & Davis, 1984; Forster, Davis, Schoknecht, & Carter, 1987) have shown that the existence of these neighbours can inhibit priming in the case of word forms. This is not of direct concern to the current results, as form priming was not attempted here. However,

we may speculate that the number of "semantic neighbours" a cue has – words with related meanings – influences the extent to which semantic priming may be induced. Precisely how semantic neighbours should be defined and operationalized, and whether their existence actually inhibits semantic priming is an issue to be examined in future studies.

Finally, it should be clear from Figure 7.1 that cue selection and response options are both bound by the contexts of their language. One obvious repercussion of this is that response options are finite. More importantly for L2 WA methodology – in particular, for the adoption of restricted WA tasks – is that specific types of responses may be limited within specific languages. For example, in English, we cannot expect synonymous responses to cues such as *fork* and *rose*. These words, as they are most often used at least, represent a utensil and a flower that do not go by any other names.[2] Similarly, in a restricted association rhyming task we cannot expect too many responses from a cue like *orange* (which is perhaps famously known as the only English word that does not have a rhyme). A different language may have more or fewer possible responses given the same tasks and the equivalent set of cues. This is certainly a caveat for future bilingual and cross-language WA studies.

7.2.2 Subjects

Concerning respondents participating in WA studies, we may surmise the following:

[2] Of course, these may also be interpreted as *the fork in the road* and the past tense of *rise* respectively, to which synonyms might more easily be associated.

1. Individual WA respondents have cognitive styles and response preferences. Depending on the methodological parameters of the research in which they are participating, they may or may not respond in accordance with these preferences.
2. With increased proficiency, subjects' L2 responses begin to accord with their L1 responses.

I have demonstrated the first of these conclusions in the case of priming effects, induced both by cue order and by restricted association tasks. As the results showed, the influence of these manipulations can have an impact on response style. In the case of certain subject groups, under certain conditions at least, respondents did veer from their natural response preferences. My findings have also demonstrated support for the second conclusion as originally proposed by Fitzpatrick (2007).

Just as the psycholinguistic literature adds to our understanding of cue effects, it also provides some answers as to why – in many cases – we did not observe priming effects when respondents were tasked with L2 association trials. At its core, this is an issue of proficiency and automaticity. Favreau and Segalowitz (1983), for example, have shown that a certain level of L2 proficiency needs to be attained before learners are capable of automatically accessing lexical knowledge. Bijeljac-Babic et al. (1997), also Van Hell and Dijkstra (2002), argue that only when subjects are able to accurately and automatically process lexical information can priming effects be induced. This is certainly a consideration for future L2 WA studies.

So far, our conclusions about WA research participants and the cues to which they respond point to a number of dichotomies. On the one hand, we expect responses to reflect the individual's natural preferences, but on the other, the repertoire of response options is limited by the individual's knowledge of the cues and the language as a whole. Importantly, there can also be conflict between response preferences and the parameters of the research design, whether it be via specific (restricted) association tasks, priming manipulations, or cue selection. For these reasons, no model of the determinants of WA behaviour can be considered complete without including experimental methodology as an essential element.

7.2.3 Methodology again

As I have stated above, there is at least one essential element missing from the superficial "*cue* → *respondent* → *response*" model of WA. That element is research methodology. The sections immediately above – indeed, this dissertation as a whole – suggest that methodology, and in particular, the manner in which responses are characterized, is in itself a determining factor in the production/elicitation of WA responses. Yes, cue strength is a strong determinant of response, but responses may be altered by the experimental manipulation of cue order, and of course by selecting cues based on specific features. Likewise, response preferences and cognitive styles are contributing determinants to response behaviour, but, depending on the experimental manipulation applied, subjects may not always respond in accordance with their predispositions. In other words, no model of the determinants of WA behaviour can be complete without the inclusion of the very methods by which those behaviours are observed and analysed.

Once again, our focus on methodology brings us back to the centrality of categorization to the research process.

1. The way we conceptualize the mental lexicon – and language itself – determines how we categorize WA responses. At the same time, the way we categorize WA responses fundamentally determines how we conceptualize WA research findings.
2. Psychological validity is an essential criterion in the assessment of any classification of WA responses. Only if categories accurately identify distinct types of response will they provide insight into underlying lexical processes.

One of the ramifications of these conclusions is perhaps obvious. Researchers must pay due attention to the criteria by which they select or adapt a categorization scheme. Throughout this dissertation – and, in particular, in Chapter 3 which focussed on the effectiveness of the categorization scheme first introduced by Fitzpatrick (2006) – the attempt has been made to select and adapt response categories that would best suit the needs of the study at hand while still representing the underlying lexical processes that participants, presumably, engaged in. In other words, there would no point in adopting a position-based category, for example, if we did not believe that the retrieval process of collocational responses differed from that of, say, form-based responses.

Perhaps the most important ramification stemming from these conclusions is the importance of the relationship between psychological validity and the way response categories are defined in WA research. Without psychologically valid response categories – that is, without categories that accurately label the underlying processes taking place in the mental lexicon during WA trials – we cannot reach any reasonable conclusions from response data at all. In this way,

psychological validity plays an essential role in the WA research model I propose below.

7.3 A dynamic model of the word association process: The DMWA

The model I am proposing here is called the Dynamic Model of Word Association (DMWA). The model incorporates the three elements described above (i.e., cues, respondent factors, and categorization) and depicts them as interacting within relationships that are constantly in flux. In this way, the model is "dynamic" in the everyday sense of the word; it portrays relationships between its constituent elements in a constant state of change. It is also a "dynamic" model in the sense of the word as used in the field of physics. That is, the interactions between elements in the model can be regarded as forces in conflict. Response behaviour is thus the product of the interaction of the forces exerted by the cue (i.e., cue strength; its tendency to produce a single type of response) and by the predispositions of the respondent (e.g., cognitive style). These two forces meet within the context of WA trials where the trials themselves exert their own force on possible responses. In other words, the influences of the methodology – the demands of the task – are the third force that comes into play in determining WA responses. All WA responses are thus determined by the interaction of these three forces.

At the end of a WA trial, after the interactions of these three forces have yielded a response, one more important element of the model comes into play: categorization. What makes the categorization of responses so essential to the WA process is that it is the strength of the scheme, as measured in its psychological

validity, that allows researchers to draw firm conclusions about what WA responses actually mean.

In the following sections a number of simulations will demonstrate how these elements interact to determine WA responses. I will argue that, by accounting for the relative strengths of the forces described above, within the context of a psychologically valid categorization scheme, this model can predict WA behaviour.

7.3.1 Forces and charges

Beginning with Figure 7.2, a number of diagrams help to illustrate how this model works. In each, we see the contributing forces of respondent factors and cue strength on the left side of the figures. To recapitulate, *respondent factors* refers to cognitive style or response profile, i.e., the respondent's predisposition to respond in a particular way. *Cue strength* refers to the capacity of the cue to elicit particular types of responses. Each of these is marked with either a positive or negative icon. In the case of respondent factors, a positive "charge" here may mean, for example, a tendency to respond with meaning-based responses. By the same token, a positive charge for cue strength would mean that the cue has a strong likelihood of eliciting a meaning-based response. In the same way, a negative charge would mean a tendency to respond with, say, position-based responses, in the case of respondent factors, or the tendency to elicit position-based responses, in the case of cue strength. Note that, for the sake of simplicity, I am using positive charges to represent the forces directed toward meaning-based responses and negative charges to represent forces aimed toward position-based responses in all of these simulations. In fact, these charges may represent forces aimed towards any

specific type of response, like those comprising the many subcategories described in the current studies. To simulate trials in which three or more types of force collide, researchers may find it useful to adopt *M* for meaning-based, *P* for position-based, and *F* for form-based forces (rather than binary pluses and minuses). Similarly, one could adopt say *M+* for meaning-based forces and use *M-* to represent all opposing forces.

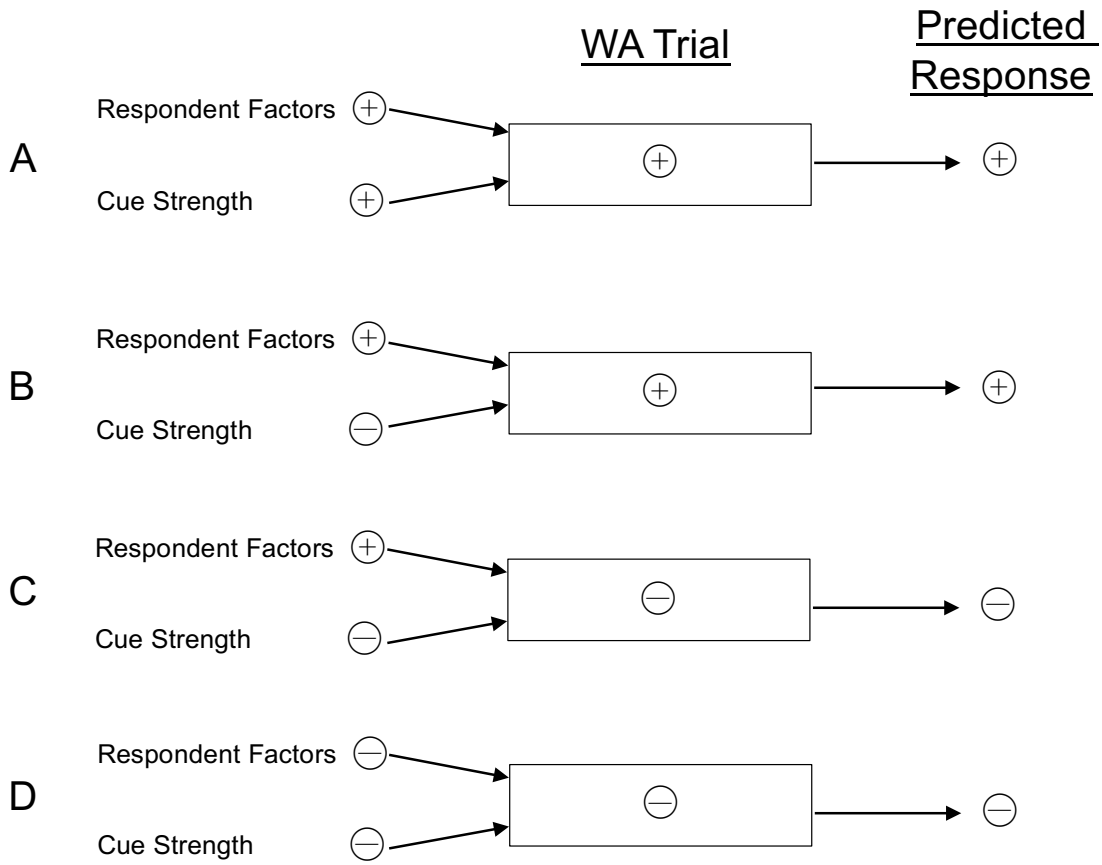


Figure 7.2. Simulations of congruent and incongruent forces within the DMWA.

Methodology appears near the centre of Figure 7.2 in the form of the WA trial. Just as the forces exerted by respondents and cues are charged, so too can the methodology be. In this case, a positively charged methodology would be one in which the researcher attempts to elicit meaning-based responses from her

subjects. As seen in the current studies, this might be achieved through the intentional manipulation of cues (e.g., by selecting those with a higher likelihood of eliciting meaning-based responses, or by ordering them in such a way as to prime meaning-based responses; see Chapter 5). It might also be achieved through other types of experimental manipulation, such as the restricted association tasks employed in Chapter 6.

7.3.2 Congruity

One of the key features of this model is its ability to make predictions based on the congruity, or incongruity, of charges among the contributing forces. Examples A and D in Figure 7.2 show the results of WA trials in which all three of the response-determining forces are congruently charged. In example A, the subject has a tendency to produce, say, meaning-based responses, the cue tends to elicit meaning-based responses, and the experiment itself is designed to promote the elicitation of meaning-based responses. For this trial, therefore, we may predict with reasonable certainty that the response will be meaning-based. This appears on the right side of the figure as a positively charged response.

Examples B and C in Figure 7.2 represent trials in which there exists incongruity between the contributing – in this case, competing – forces. Example B illustrates a subject who tends to respond with, for example, meaning-based responses. Here she is presented with a cue that tends to produce position-based responses, yet the trial is embedded within an experiment designed to elicit meaning-based responses. Assuming for the moment that all three forces exert an equal amount of influence on the response process, then the DMWA predicts that the response will

be a meaning-based one. The outcome of their interaction is thus determined mathematically: two positives + one negative = one positive (i.e., a meaning-based response). In the section below I will use the DMWA to simulate the WA process where unequally strong forces are brought to bear on the determination of responses.

7.3.3 Varying strengths of determining forces

The DMWA can account for differences in strength among the response-determining forces. Example E in Figure 7.3 represents a subject with a tendency to respond with meaning-based responses and a methodology intended to prompt meaning-based responses. Respondent factors and the WA trial are thus marked with positive charges. At the same time however, this particular trial involves the presentation of a cue with a very strong tendency to elicit position-based responses. This is denoted by two negative charges for cue strength.

As I have explained above, the use of positive and negative charges in the diagrams are a means of simply representing any number of forces at work during the WA process. Likewise, the use of one or two charges here is also a shorthand means of expressing the relative strength of forces and should not be taken as an exact quantitative value. This is demonstrated by applying the mathematical calculation of charge from the section above to the current example. In this case, the formula would be: one positive (respondent) + one positive (methodology) + two negatives (cue). An overall charge of zero (i.e., neither positive nor negative) should mean that it is impossible to predict what the response would be in this case. However, I am suggesting in Example E, that we may predict the resulting response to be

position-based. My reasoning is this: assuming that the strength exerted by any one of the contributing forces is strong enough, it becomes unsusceptible to other influences. An example of this would be a cue such as *billiard* whose most common primary responses are 81% position-based (H. Moss & Older, 1996; see also Table 5.2). In this case, the cue's capacity to elicit position-based responses may be almost insurmountable. Regardless of the strengths of respondent tendencies or experimental manipulation, these forces are unlikely to exert enough influence to override the strong bonds presumed to pre-exist between the cue and its possible (position-based) responses within the mental lexicon of the respondent.

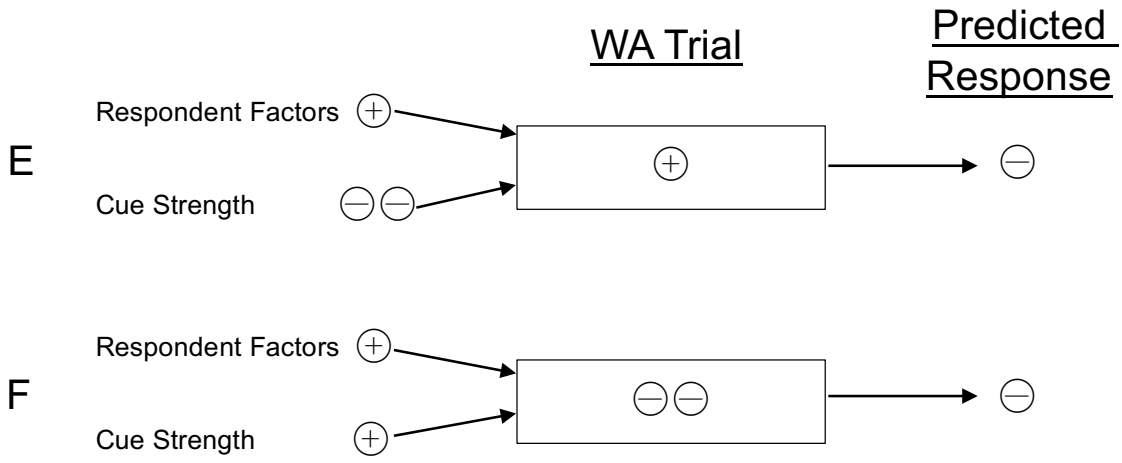


Figure 7.3 DMWA simulations of WA trials with forces of varying strength.

Note that this account also provides an explanation in support of the long-held assumption that WA researchers should avoid utilizing cues which exist in strongly bonded word pairs such as opposites (e.g., *black-white*) or binomial chunks (*salt and pepper*). The DMWA illustrates that a given force – in this case, cue strength – may simply be too strong to be overcome by the opposing forces operating within the WA response process. Users of the DMWA who wish to more accurately

portray the relative strengths of WA forces may choose to adopt a scale of numerical values (e.g., *-5* to *+5*) in place of the shorthand, single- or double-charges I am using here.

Example F in Figure 7.3 illustrates another trial in which unequal forces are brought to bear on the response process. In this case, the respondent has a predisposition to provide meaning-based responses. At the same time, the cue is expected to elicit a meaning-based response. Both of these forces are marked with positive charges in the figure. The WA trial itself however is marked with two negative charges suggesting that the methodology is intended to strongly elicit position-based responses. To choose a very obvious scenario, this could be an example of a restricted association task where the respondent is asked to respond to the cue with a collocation. In this case, it would be very unlikely for the respondent to provide a meaning-based response.

In the case of a free (i.e., non-restricted) association task however, Example F may represent a trial in which the respondent is in a priming condition, having undergone experimental manipulation intended to increase the likelihood that she will respond to the cue with a position-based response. As seen in Figure 7.3, the expected outcome, according to the DMWA, is a position-based response. Such a prediction, however, appears to contradict the results reported in Chapters 5 and 6 where, more often than not, the priming manipulation did not have a statistically significant influence on subject responses. I will argue below that these findings are not actually contradictions.

7.3.4 DMWA prerequisites

The example above states that in the case of a strong priming condition, the DMWA predicts that we would elicit responses in accordance with the prime regardless of whether the forces of respondent factors and cue strength were congruent with that of the priming or not. While this may appear to be a fault in the model's ability to make predictions, the current studies provide evidence that the DMWA's accuracy as a descriptor of the WA process and as a predictor of responses relies heavily on the existence of three important methodological preconditions. Only by taking these prerequisites into account can we begin to resolve the discrepancies between the current findings and the predictions made by the DMWA.

7.3.4.1 Valid categories

The first prerequisite for successfully applying the DMWA to the association process is that researchers must employ a valid categorization scheme. I have stressed repeatedly the importance of categorization to the WA research process and I will re-emphasize it here: only when the categorization scheme accurately and comprehensively accounts for all possible responses, *while remaining psychologically valid*, can we reliably predict response outcomes. For this reason, we can say that all of the simulations and predictions made by the DMWA are premised upon a psychologically valid categorization scheme. Without such a scheme, accurate response predictions are unlikely and there are no grounds on which to evaluate the model. As I have defined it above, psychological validity refers to a categorization scheme's ability to accurately label the underlying categories and processes presumed to be at work in the mental lexicons of respondents engaged in WA trials. What should be apparent from this definition is

that, if a psychologically valid categorization of responses has been adopted, then any change in response behaviour – whether it be over time, across languages, between subject groups, or due to experimental manipulation – can be attributed to changes or differences within the mental lexicon. To make reasonable inferences about the mental lexicon is one of the fundamental purposes underlying the current studies – indeed, underlying much of WA research in applied linguistics. For this reason, a psychologically valid categorization scheme remains a key prerequisite for these studies.

7.3.4.2 Strong (enough) forces

Beyond categorization, the priming methods employed in the three studies reported in Chapters 5 and 6 also raise issues for the descriptive and predictive power of the DMWA. Taken in order, these studies reveal a progression from a rather rough priming methodology – that may have failed to create the intended experimental conditions (Chapter 5) – to a relatively refined one (in the supplementary study of Chapter 6). The underlying issue here is whether any of these three methods actually created priming conditions strong enough to justify claims made about them. As a specific example, in the position-based priming condition in the supplemental study in Chapter 6, responses from neither the NS group, nor the NNS group, showed any significant priming effects. I have already detailed some of the possible reasons for this finding in the discussion in that chapter, but I wish to draw attention to one point in particular.

One of the purposes of conducting the priming studies reported in Chapters 5 and 6 was to test the validity of the *response profile* concept empirically. I had already argued in the discussion section of Chapter 4 that subjects' response tendencies

(across languages, etc.) are stable enough to warrant conceptualizing them as "profiles". The priming studies were intended to test more precisely how strong profiles actually are. In essence, these studies pit the robustness of response profiles against another robust phenomenon, priming.[3] However, from my descriptions of how the priming conditions were induced in these studies, it should be clear that the methods were atypical in comparison with those used in prior research. In the current studies, cue order was manipulated and restricted association tasks were used as a means of priming subject responses. This is very different from the timed presentation of related words seen in almost all priming studies to date (see for example Elgort, 2011, reviewed in Section 5.4). The question is, then: Can we be certain that the priming methods employed in these studies were strong enough to say that we really have tested the relative strengths of profiles vs. priming. If we wish to make solid claims about the durability of response preferences under the influence of experimental manipulation, more reliable, well-established methods need to be employed.

7.3.4.3 Traditional measures

Finally, a third aspect of the current methods to be re-examined in light of the DMWA is the manner in which the priming scores were measured. Just as the methods used to induce the priming effects were atypical in comparison to prior research, so too was the manner in which these effects were scored and measured. My interest, from the very beginning of these studies has been in human vocabulary. How words are linked and organized in the mind, how lexical

[3] For an idea of how robust priming effects are, consult the enormous body of psychological and psycholinguistic literature on this topic, or see Kinoshita and Lupker (2003); Lucas (2000); McDonough and Trofimovich (2009); Neely (1991); and Trofimovich and McDonough (2011) for overviews and examples.

representations in first and second languages are stored in memory, and why people respond the way they do in WA trials, were three of the broad, fundamental issues I wished to explore when I began this research. This interest in lexis accounts for my focus on changes in response type – rather than changes in RT – as an indicator of priming effects. This methodological choice was also supported by my wish to examine and assess WA categorization schemes.

Despite these reasons for choosing response type as the unit by which to measure priming, there is also reason not to do so. Just as the methods used to induce priming effects were untried, so too were the means employed to measure the effects. Scores for the priming effects here were based on the differences between three potentially primed trials and three control trials. As I have stated in the priming chapters, the choice of three experimental trials was at least somewhat arbitrary. Perhaps this kind of priming "resets quickly". That is, maybe a strong priming effect can be captured in a single trial and the use of three is merely dissipating that effect. It could also be the case that this kind of priming effect is very subtle and does not "take effect" immediately. If this is the case, then we can expect to capture possible effects over, say, 10 or 20 trials. We may miss the effect altogether if experimental trials are limited to only three per subject. Precisely how many trials researchers need to observe to best capture this type of priming – or even the more conventional type of priming effect – is an empirical question that deserves scrutiny in future studies.

To reiterate, it should be noted that the many examples of robust priming effects so well documented in the research literature are based almost entirely on the

measurement of reaction times. For this reason, I suggest that, alongside traditional methods of inducing priming effects, traditional reaction time measurements be utilized to measure them.

7.4 Summary and conclusion

As I have already summarized my approach, my methods, my findings, and the limitations of my work in each chapter as I reported them, I will offer only a broad overview of the dissertation as a whole here. To summarize, then, the studies reported in this dissertation have contributed to the L2 WA research literature in the following ways:

- several lines of empirical evidence have been presented, revealing and reflecting upon the inherent complexity of the WA process itself and the intricacies of WA research.

- some of the fundamental issues of WA research methodology have been investigated, including the ways in which response data are categorized (Chapter 3) and conceptualized (Chapter 4).

- new methods have been introduced (e.g. the use of cue order as a means of eliciting priming effects; Chapter 5) and seldom-used methods have been adopted and adapted for new purposes (e.g., restricted association tasks; Chapter 6).

These studies, their results, and in particular, my reflection upon their weaknesses, and on the inherent difficulties of this line of research, have culminated in the Dynamic Model of Word Association that I have introduced above.

Finally, I should add that I have attempted to address weaknesses in the current studies in each chapter. Some of these are attributable to my inexperience as a researcher, or simply to my mistakes. Others, undoubtedly, I have yet to discover.

It is my sincere wish that these findings and the model they helped to bring about, will receive the attention and scrutiny that I have attempted to bring to the prior studies upon which the current studies are based. WA research is fraught with the kinds of difficulties that arise when human subjects are involved and the phenomena to be observed are as fleeting as cognitive processes, or as nebulous and notional as the mental lexicon itself. An approach to further research that balances critical scrutiny and a wish to innovate will serve the field well.

References

Aitchison, J. (2003). *Words in the mind* (3rd ed.). Oxford: Blackwell.

Aitchison, J. (2012). *Words in the mind* (4th ed.). West Sussex: Wiley-Blackwell.

Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language*. New York: Palgrave Macmillan.

Aldridge, M., Fontaine, L., Bowen, N., & Smith, T. (2018). A new perspective on word association: how keystroke logging informs strength of word association. *WORD, 64*(4), 218-234. doi: 10.1080/00437956.2018.1535365

Arai, M., van Gompel, R., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*(3), 218-250. doi: 10.1016/j.cogpsych.2006.07.001

Arthur, A. Z., & Freemantle, A. (1966). Extreme response bias and associative commonality. *Psychonomic Science, 6*(8), 399. doi: 10.3758/bf03330956

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*(1), 101-114.

Barfield, A. W. (2006). *An exploration of second language collocation knowledge and development.* University of Wales, Swansea.

Bassili, J. N., Smith, M. C., & MacLeod, C. M. (1989). Auditory and visual word stem completion: Separating data-driven and conceptually driven processes. *The Quarterly Journal of Experimental Psychology, 41A*, 439-453.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279. doi: 10.1093/ijl/6.4.253

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory and Cognition, 25*(4), 447-457.

Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: The case of the *Euralex French Tests*. *Applied Linguistics, 21*(4), 490-516. doi: 10.1093/applin/21.4.490

Bøyum, M. (2016). *A new word association test for Norwegian.* (Unpublished M.A. Thesis), University of Oslo.

Branigan, H., Pickering, M., Stewart, A., & McLean, J. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition, 28*(1297-1302).

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13*(3), 359-371.

Brysbaert, M., Verreyt, N., & Duyck, W. (2010). Models as hypothesis generators and models as roadmaps. *Bilingualism: Language and Cognition, 13*(3), 383-384.

Carroll, J. B., Kjeldergaard, P. M., & Carton, A. S. (1962). Number of opposites versus number of primaries as a response measure in free-association tests. *Journal of Verbal Learning and Verbal Behavior, 1*(1), 22-30. doi: 10.1016/S0022-5371(62)80015-2

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1029-1040. doi: 10.1037/a0015901

Churchill, E. (2007). A dynamic systems account of learning a word: from ecology to form relations. *Applied Linguistics, 29*(3), 339-358.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.
Cook, V. (Ed.). (2003). *Effects of the second language on the first*. Clevedon, UK: Multilingual Matters.
Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. doi: 10.2307/3587951
Cramer, P. (1968). *Word association*. New York: Academic Press.
Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: simulating semantic priming. *Cognitive Science, 23*, 371-414.
Dagenbach, D., Carr, T. H., & Barnhardt, T. M. (1990). Inhibitory semantic priming of lexical decisions due to failure to retrieve weakly activated codes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 328-340.
Davis, B. J., & Wertheimer, M. (1967). Some determinants of associations to French and English words. *Journal of Verbal Learning and Verbal Behavior, 6*(4), 574-581. doi: 10.1016/S0022-5371(67)80019-7
De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics, 13*(1), 1-24.
de Bot, K. (2015). *A history of applied linguistics: From 1980 to the present*. London: Routledge.
De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*(1), 198-205. doi: 10.3758/BRM.40.1.198
Deese, J. (1962). Form class and the determinants of association. *Journal of Verbal Learning and Verbal Behavior, 1*, 79-84.
Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Erlbaum.
Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*, 175-197.
Dijkstra, T., van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica, 38*(3/4), 177-196.
Du, W., & Gao, Y. (2013). Study on the organizational pattern of mental lexicon for Chinese ESL college students of science. *International Journal of English Linguistics, 3*(4), 50-62. doi: 10.5539/ijel.v3n4p50
Dunn, S., Bliss, J., & Siipola, E. (1958). Effects of impulsivity, introversion, and individual values upon association under free conditions. *Journal of Personality, 26*(1), 61-76. doi: 10.1111/j.1467-6494.1958.tb01571.x
Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367-413.
Ellis, A. W. (1982). Modality-specific repetition priming of auditory word recognition. *Current Psychological Research, 2*, 123-128.
Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, R. (2002a). Does form-focused instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition, 24*(2), 223-236.

Ellis, R. (2002b). The place of grammar instruction in the second/foreign language curriculum. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 17-34). Mahwah, NJ: Lawrence Erlbaum.

Entwisle, D. R. (1966a). Form class and children's word associations. *Journal of Verbal Learning and Verbal Behavior, 5*(6), 558-565. doi: 10.1016/S0022-5371(66)80091-9

Entwisle, D. R. (1966b). *Word associations of young children*. Baltimore: Johns Hopkins.

Entwisle, D. R., Forsyth, D. F., & Muuss, R. (1964). The syntagmatic-paradigmatic shift in children's word associations. *Journal of Verbal Learning and Verbal Behavior, 3*, 19-29.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *American Journal of Psychology, 74*(3), 361-372.

Ervin, S. M., & Osgood, C. E. (1954). Second language learning and bilingualism. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A Survey of Theory and Research Problems* (pp. 139-146). Baltimore: Waverly Press.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition, 11*(6), 565-574.

Feldman, L. B. (2003). What the repetition priming methodology reveals about morphological aspects of word recognition. In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 124-138). Oxford: Oxford University Press.

Fillenbaum, S., & Jones, L. V. (1965). Grammatical contingencies in word association. *Journal of Verbal Learning and Verbal Behavior, 4*, 248-255.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51*, 1-22. doi: 10.1016/j.jml.2004.01.004

Fishman, J. A., & Cooper, R. L. (1969). Alternative measures of bilingualism. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 276-282. doi: 10.1016/S0022-5371(69)80076-9

Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook, 6*, 121-146. doi: 10.1075/eurosla.6.09fit

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics, 17*, 319-331. doi: 10.1111/j.1473-4192.2007.00172.x

Fitzpatrick, T. (2009). Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners* (pp. 38-52). Bristol, UK: Multilingual Matters.

Fitzpatrick, T., & Izura, C. (2011). Word associations in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition, 33*, 373-398. doi: 10.1017/S0272263111000027

Fitzpatrick, T., & Munby, I. (2014). Knowledge of word associations. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 92-105). Basingstoke, UK: Palgrave Macmillan.

Fitzpatrick, T., & Playfoot, D. (2011). *Investigating the effects of age on lexical retrieval behavior*. Paper presented at the British Association for Applied Linguistics, Bristol.

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. doi: 10.1093/applin/amt020

Fitzpatrick, T., & Racine, J. P. (2014). *Using learners' L1 word association profiles as an alternative to native speaker norms*. Paper presented at the AILA World Congress, Brisbane, Australia.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 680-690.

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: repetition of partial activation. *The Quarterly Journal of Experimental Psychology, 39A*, 211-251.

Forster, K. I., & Veres, C. (1998). The prime lexicality effect: form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*(2), 498-514.

Francis, H. (1972). Toward an explanation of the syntagmatic-paradigmatic shift. *Child Development*, 949-958.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *TRENDS in Cognitive Sciences, 8*(2), 87-93.

Frenck-Mestre, C., & Prince, P. (1997). Second language autonomy. *Journal of Memory and Language, 37*, 481-501.

Galton, F. (1879). Psychometric experiments. *Brain, 2*, 149-162.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Dent & Sons.

Garrett, M. F. (1992). Disorders of lexical selection. *Cognition, 42*, 143-180.

Garrett, M. F. (1993). Errors and their relevance for models of language production. In G. Blanken, J. Dittman, H. Grim, J. Marshall & C. Wallesch (Eds.), *Linguistic disorders and pathologies* (pp. 72-92). Berlin: de Gruyter.

Gewirth, L. R., Shindler, A. G., & Hier, D. B. (1984). Altered patterns of word associations in dementia and aphasia. *Brain and Language, 21*, 307-317.

Glaser, W. R., & Düngelhoff, R. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640-654.

Glucksberg, S. (1984). The functional equivalence of common and multiple codes. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 100-104. doi: 10.1016/S0022-5371(84)90529-2

Gollan, T. H., Salmon, D. P., & Paxton, J. L. (2006). Word association in early Alzheimer's disease. *Brain and Language, 99*(289-303).

Goral, M., Spiro, I., Albert, M. L., Obler, L. K., & Connor, L. T. (2007). Change in lexical retrieval skills in adulthood. *The Mental Lexicon, 2*(2), 215-238.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: types of distractors and types of associations. *The Modern Language Journal, 85*, 567-577.

Guo, T., Misra, M., Tam, J. W., & Kroll, J. F. (2012). On the time course of accessing meaning in a second language: An electrophysiological and behavioral investigation of translation recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*(5), 1165-1186. doi: 10.1037/a0028076

Harris, R. (2001). *Saussure and his interpreters*. Edinburgh: Edinburgh University Press.

Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language* (pp. 22-62). Basingstroke: Palgrave Macmillan.

Higby, E., & Obler, L. K. (2015). Losing a first language to a second language. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 645-664). Cambridge: Cambridge University Press.

Higginbotham, G. (2010). Individual learner profiles from word association tests: The effect of word frequency. *System, 38*, 379-390. doi: 10.1016/j.system.2010.06.010

Higginbotham, G. (2014). *Individual profiling of second language learners through word association.* (Unpublished Ph.D. dissertation), Swansea University, UK.

Higginbotham, G., Munby, I., & Racine, J. P. (2015). A Japanese word association database of English. *Vocabulary Learning and Instruction, 4*(2), 1-20. doi: 10.7820/vli.v04.2.higginbotham.et.al

Hirsch, K. W., & Tree, J. T. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics, 14*, 1-44. doi: 10.1016/S0911-6044(00)00002-6

Isurin, L. (2013). Memory and first language forgetting. In J. Altarriba & L. Isurin (Eds.), *Memory, Language, and Bilingualism* (pp. 319-348). Cambridge: Cambridge University Press.

Jenkins, J. J. (1960). Commonality of association as an indicator of more general patterns of verbal behavior. In T. A. Sebeok (Ed.), *Style in Language* (pp. 307-329). Cambridge, MA: MIT Press.

Jenkins, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), *Norms of word associations* (pp. 1-38). New York: Academic Press.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Jiang, N. (2015). Six decades of research on lexical presentation (sic) and processing in bilinguals. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 29-84). Cambridge: Cambridge University Press.

Jung, C. G. (1918/1969). *Studies in word-association* (M. D. Eder, Trans.). London: Routledge & Kegan Paul.

Keatley, C. W. (1992). History of bilingualism research in cognitive psychology. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 15-49). Amsterdam: Elsevier Science.

Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Psychiatry, 67*, 37-96.

Kinoshita, S., & Lupker, S. J. (Eds.). (2003). *Masked priming: The state of the art*. New York: Psychology Press.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In J. Aitken, R. W. Bailey & N. Hamilton Smith (Eds.), *The Computer and Literary Studies* (pp. 153-165). Edinburgh: Edinburgh University Press.

Kolers, P. A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior, 2*(4), 291-300. doi: 10.1016/S0022-5371(63)80097-3

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom*. Harlow, UK: Pearson Education.

Kroll, J. F., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 169-199). Mahwah, NJ: Lawrence Erlbaum Associates.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*(2), 137-171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*, 149-174. doi: doi:10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. L. Nicol & T. Langendoen (Eds.), *Language Processing in Bilinguals* (pp. 49-71). Cambridge: Blackwell.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381.

Kruse, H., Pankhurst, M., & Sharwood Smith, M. (1987). A multiple word association probe in second language acquisition research. *Studies in Second Language Acquisition, 9*(2), 141-154. doi: 10.1017/S0272263100000449

Kudo, Y., & Thagard, D. (1999). Word associations in L2 vocabulary. *University of Hawai'i Working Papers in ESL, 17*(2), 75-105.

Lambert, W. E. (1956). Developmental aspects of second-language acquisition: I. Associational fluency, stimulus provocativeness, and word-order influence. *Journal of Social Psychology, 43*(1), 83-89. doi: 10.1080/00224545.1956.9919202

Lambert, W. E., & Moore, N. (1966). Word-association responses: Comparisons of American and French monolinguals with Canadian monolinguals and bilinguals. *Journal of Personality and Social Psychology, 3*(3), 313-320. doi: 10.1037/h0023046

Lambert, W. E., & Rawlings, C. (1969). Bilingual processing of mixed-language associative networks. *Journal of Verbal Learning and Verbal Behavior, 8*(5), 604-609. doi: 10.1016/S0022-5371(69)80111-8

Language Teaching Review Panel. (2008). Replication studies in language learning and teaching: questions and answers. *Language Teaching, 41*(1), 1-14.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal, 25*(2), 21-33.

Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world: A response to Meara (2005). *Applied Linguistics, 26*(4), 582-588.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322. doi: 10.1093/applin/16.3.307

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. Harlow, UK: Longman.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, H. (2011). An investigation into the L2 mental lexicon of Chinese English learners by means of word association. *Chinese Journal of Applied Linguistics, 34*(1), 62-76.

Linton, M. L., & Brotsky, S. J. (1969). Administrative procedures in the word association test: method of stimulation and method of responding (oral vs. written). *Journal of Verbal Learning and Verbal Behavior, 8*, 123-128.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7*(4), 618-630. doi: 10.3758/BF03212999

Manschreck, T. C., Merrill, A. M., Jabbar, G., Chun, J., & DeLisi, L. E. (2012). Frequency of normative word associations in the speech of individuals at familial high-risk for schizophrenia. *Schizophrenia Research, 140*, 99-103.

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129*(3), 361-368. doi: 10.1037//0096-3445.129.3.361

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173-193.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology, 101*(2), 221-242. doi: doi:10.1348/000712608X371744

Mauranen, A. (2011). Learners and users: Who do we want corpus data from? In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A taste for corpora* (pp. 155-171). Amsterdam: Benjamins.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3-25). Hillsdale, NJ: Erlbaum.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375-407.

McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition, 28*, 179-207.

McDonough, K., & Mackey, A. (2008). Syntactic priming and ESL question development. *Studies in Second Language Acquisition, 30*, 31-47.

McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. New York: Routledge.

McNeill, D. (1963). The origin of associations within the same grammatical class. *Journal of Verbal Learning and Verbal Behavior, 2*, 250-262.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 558-572.

Meara, P. (1978). Learners' word associations in French. *Interlanguage Studies Bulletin, 3*(2), 192-211.
Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching, 13*(3-4), 221-246.
Meara, P. (1982). Word associations in a foreign language. *Nottingham Linguistic Circular, 11*(2), 28-38.
Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*(1), 32-47.
Meara, P. (2009). *Connected words: word associations and second language vocabulary acquisition*. Amsterdam: Benjamins.
Meara, P. (2012). The bibliometrics of vocabulary acquisition: An exploratory study. *RELC Journal, 43*(1), 7-22. doi: 10.1177/0033688212439339
Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System, 28*, 19-30.
Meara, P., & Jones, G. (1990). *Eurocentres Vocabulary Size Test 10Ka*. Zurich: Eurocentres.
Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41*(3), 211-253.
Merten, T. (1993). Word association responses and psychoticism. *Personality and Individual Differences, 14*(6), 837-839.
Merten, T., & Fischer, I. (1999). Creativity, personality and word association responses: associative behaviour in forty supposedly creative persons. *Personality and Individual Differences, 27*, 933-942.
Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognising pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 80*, 227-235.
Milton, J., Daller, H., Malvern, D., Meara, P., Richards, B., & Treffers-Daller, J. (2008). Guest editorial. *Language Learning Journal, 36*(2), 135-138. doi: 10.1080/09571730802389959
Miron, M. S., & Wolfe, S. (1964). A cross-linguistic analysis of the response distributions of restricted word associations. *Journal of Verbal Learning and Verbal Behavior, 3*(5), 376-384. doi: 10.1016/S0022-5371(64)80006-2
Moran, L. J. (1966). Generality of word-association response sets. *Psychological Monographs: General and Applied, 80*(4), 1-25.
Moss, H., & Older, L. (1996). *Birkbeck word association norms*. East Sussex, UK: Psychology Press.
Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*(4), 863-883.
Muncie, J. (2002). Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts. *System, 30*(2), 225-235. doi: https://doi.org/10.1016/S0346-251X(02)00006-4
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.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Laberge & S. J. Samuels (Eds.), *Basic processes in reading: perception and comprehension* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological Bulletin, 84*(1), 93-116.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*, 45-52.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics, 16*(3), 389-408.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*(1), 1-32.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*(4), 273-281. doi: 10.1080/17470216508416445

Orita, M. (2002). Word associations of Japanese EFL learners and native speakers: Shifts in response type distribution and the associative development of individual words. *Annual Review of English Language Education in Japan, 13*, 111-120.

Paribakht, T. S., & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*, 9-29.

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring yes-no vocabulary tests: Reaction time vs. nonword approaches. *Language Testing, 29*(4), 489-509.

Politzer, R. L. (1978). Paradigmatic and syntagmatic associations of first year French students. In V. Honsa & M. J. Hardman-de-Bautista (Eds.), *Papers on Linguistics and Child Language* (pp. 203-210). Berlin: Mouton.

Postman, L., & Keppel, G. (Eds.). (1970). *Norms of word association*. New York: Academic Press.

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23-28. doi: 10.1016/S0022-5371(84)90489-4

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*(1), 36-57.

Racine, J. P. (2008). Cognitive processes in second language word association. *JALT Journal, 30*(1), 5-26.

Racine, J. P. (2011a). Feeding, breeding, extinction, and river length: Problems with treating words like species - Comments on Meara and Olmos Alcoy (2010). *Reading in a Foreign Language, 23*(2), 235-237.

Racine, J. P. (2011b). Grammatical words and processes in the L2 mental lexicon: a word association perspective. *Studies in Foreign Language Teaching, 29*, 153-197.

Racine, J. P. (2011c). Loanword associations and processes. *OTB - The Tsukuba Multi-lingual Forum, 4*(1), 37-44.
Racine, J. P. (2013). The history and future of word association research. *Dokkyo Journal of Language Learning and Teaching, 1*, 55-73.
Racine, J. P. (2014). Reaction time methodologies and lexical access in applied linguistics. *Vocabulary Learning and Instruction, 3*(1), 66-70. doi: 10.7820/vli.v03.1.racine
Racine, J. P., Benevides, M., Graham-Marr, A., Coulson, D., Browne, C., Poulshock, J., & Waring, R. (2013). Vocabulary acquisition, input, and extensive reading: A conversation. *The Language Teacher, 37*(4), 56-60.
Racine, J. P., Higginbotham, G., & Munby, I. (2014). Exploring non-native norms: A new direction in word association research. *Vocabulary Education and Research Bulletin, 3*(2), 13-15.
Ramsey, R. M. (1981). A technique for interlingual lexico-semantic comparison: The lexigram. *TESOL Quarterly, 15*(1), 15-24. doi: 10.2307/3586369
Randall, M. (1980). Word association behaviour in learners of English as a foreign language. *Polyglot, 2*(Fiche 2), B5-D1.
Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371. doi: 10.1177/026553229301000308
Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in Language Assessment* (pp. 41-60). Mahwah, NJ: Erlbaum.
Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209-227). Amsterdam: John Benjamins.
Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77-89. doi: 10.2307/3585941
Riegel, K. F., Ramsey, R. M., & Riegel, R. M. (1967). A comparison of the first and second languages of American and Spanish students. *Journal of Verbal Learning and Verbal Behavior, 6*(4), 536-544. doi: 10.1016/S0022-5371(67)80013-6
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.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605.
Saussure, F. d. (1916/2006). *Writings in general linguistics* (C. Sanders & M. Pires, Trans.). Oxford: Oxford University Press.
Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 915-930.
Scheffler, P. (2015). Lexical priming and explicit grammar in foreign language instruction. *ELT Journal, 69*(1), 93-97.
Schmid, M. S., Köpke, B., Keijzer, M., & Weilemar, L. (2004). *First language attrition: Interdisciplinary perspectives on methodological issues*. Amsterdam: John Benjamins.
Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.
Schmidt, R. W., & Frota, S. N. (1986). Developing basic conversational ability in a second language: a case study of an adult learner of Portuguese. In R. R. Day

(Ed.), *Talking to learn: conversation in second language acquisition* (pp. 237-326). Rowley, MA.: Newbury House.
Schmitt, L. M., Ankeny, L. D., Sweeney, J. A., & Mosconi, M. W. (2016). Inhibitory Control Processes and the Strategies That Support Them during Hand and Eye Movements. *Frontiers in Psychology, 7*.
Schmitt, N. (1998a). Quantifying word association responses: What is native-like? *System, 26*(3), 389-401. doi: 10.1016/S0346-251X(98)00019-0
Schmitt, N. (1998b). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 281-317. doi: 10.1111/1467-9922.00042
Schmitt, N. (2010). Key issues in teaching and learning vocabulary. In R. Chacón-Beltrán, C. Abello-Contesse & M. M. Torreblanca-López (Eds.), *Insights into Non-native Vocabulary Teaching and Learning* (pp. 28-40). Bristol: Multilingual Matters.
Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning, 64*(4), 913-951. doi: 10.1111/lang.12077
Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition, 20*, 17-36.
Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture-word interference studies. *Journal of Memory and Language, 29*, 86-102.
Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
Sexton, N. J., & Cooper, R. (2017). Task inhibition, conflict, and the n-2 repetition cost: A combined computational and empirical approach. *Cognitive Psychology*.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127.
Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
Smith, M. C. (1997). How do bilinguals access lexical information? In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 145-168). Mahwah, NJ: Lawrence Erlbaum Associates.
Smolentseva, I. G., Sozinova, E. V., Vasenina, E. E., & Levin, O. S. (2014). Characteristics of cognitive and behavioral impairments in patients with semantic dementia with predominantly right- and left-sided cerebral atrophy. *Neuroscience and Behavioral Physiology, 44*(1), 69-75.
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, and product in language learning* (pp. 157-169). Stockholm: AFinLa.
Sökmen, A. J. (1993). Word association results: a window to the lexicons of ESL students. *JALT Journal, 15*(2), 135-150.
Son, J.-S., Do, V. B., Kim, K.-O., Cho, M. S., Suwonsichon, T., & Valentin, D. (2014). Understanding the effect of culture on food representations using word associations: The case of “rice” and “good rice”. *Food Quality and Preference, 31*, 38-48. doi: 10.1016/j.foodqual.2013.07.001

Stolz, W. S., & Tiffany, J. (1972). The production of "child-like" word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Verbal Behavior, 11*, 38-46.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28*, 387-422.

Swan, A., Aboshiha, P., & Holliday, A. (Eds.). (2015). *(En)countering native-speakerism: Global perspectives*. Basingstoke: Palgrave Macmillan.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2*(1), 45-58.

Tannenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557-580.

Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*(30), 12530-12535. doi: 10.1073/pnas.0609927104

Trofimovich, P., & McDonough, K. (Eds.). (2011). *Applying priming methods to L2 learning, teaching and research*. Amsterdam: John Benjamins.

Truscott, J. (1998). Noticing in second language acquisition: a critical review. *Second Language Research, 14*(2), 103-135.

Uemura, T., & Ishikawa, S. (2004). JACET 8000 and Asia TEFL vocabulary initiative. *Journal of Asia TEFL, 1*(1), 333-347.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in blingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193-211.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*(4), 780-789.

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458-483. doi: 10.1006/jmla.1998.2584

Vezzoli, C., Bressi, C., Tricarico, G., Boato, P., Cattaneo, C., Visentin, U., & Invernizzi, G. (2007). Methodological evolution and clinical application of C. G. Jung's word association experiment: a follow-up study. *Journal of Analytical Psychology, 52*, 89-108.

Weinreich, U. (1953/1968). *Languages in contact: Findings and problems*. The Hague: Mouton.

Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Edward Arnold.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: a depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*, 41-69.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics, 32*(4), 430-449.

Woodrow, H., & Lowell, F. (1916). Children's association frequency tables. *Psychological Monographs: General and Applied, XXII*(5), 1-110.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

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. doi: 10.1177/0267658307076543

Zareva, A. (2010). Multicompetence and L2 users' associative links: being unlike nativelike. *International Journal of Applied Linguistics, 20*(1), 2-22.

Appendix I

Word association response categories and subcategories from Fitzpatrick (2009; *x* = stimulus, *y* = response).

Category	Subcategory	Definition
Meaning-based associations	Defining synonym	x means the same as y
	Specific synonym	x can mean y in some specific contexts
	Lexical set / context relationship	x and y are in the same lexical set, are coordinates, meronyms, superordinates, or provide context
	Conceptual association	x and y have some other conceptual link
Position-based associations	Consecutive xy collocation	y follows x directly (including compounds)
	Consecutive yx collocation	x follows y directly (including compounds)
	Other collocational association	y follows or precedes x within a phrase with word(s) between them
Form-based associations	Change of affix	y is x plus or minus affix
	Similar form only	y looks or sounds similar to x but has no clear meaning link
Others	Erratic associations	y has no decipherable link to x
	Blank	no response given

Appendix II

Word association response categories based on Fitzpatrick (2006) and refined in light of the findings from the study described in Chapter 3. This classification forms the basis for response categorizations implemented in the research described in subsequent chapters.

Category	Subcategory	Definition
Meaning-based associations	1. Synonym	*x* means the same as y, has a similar meaning, or means the same thing in certain contexts
	2. Lexical set / context relationship	x and y are in the same lexical set, are coordinates, meronyms, sub/superordinates, or provide context
	3. Conceptual association	x and y have some other conceptual link
Position-based associations	4. Consecutive xy collocation	y follows x directly (or with function words between them; includes compounds)
	5. Consecutive yx collocation	x follows y directly (or with functional words between them; includes compounds)
	6. Phrasal collocation	y follows or precedes x in a phrase with word(s) between them
Form-based associations	7. Change of affix	y is x plus or minus affix
	8. Similar form only	y looks or sounds similar to x but has no clear meaning link
Others	9. Foreign associates	y is not a word in the target language, or is based on a foreign cognate
	10. Mediated responses	y is a result of a chain of associations
	11. Meta-associations	About the respondent or the situation
	12. Erratic associations	y has no decipherable link to x
		No Response

Appendix III

Stimulus words for the L2 association task (Chapter 4).

Word	Word Class	AWL Sublist	JACET Level	Word	Word Class	AWL Sublist	JACET Level
access	v/n	4	2	link	n/v	3	2
achieve	v	2	1	location	n	3	2
alternative	a/n	3	2	majority	n	1	2
annual	a/n	4	2	mechanical	a	4	3
approach	v/n	1	1	method	n	1	1
area	n	1	1	minor	a/n	3	3
aspect	n	2	2	negative	a	3	2
assist	v	2	2	normally	adv	2	2
assume	v	1	2	obviously	adv	4	2
authority	n	1	2	occupation	n	4	3
circumstance	n	3	2	occur	v	1	1
civil	a	4	2	percentage	n	1	3
communicate	v	4	2	perception	n	2	2
community	n	2	1	period	n	1	1
computer	n	2	1	phase	n/v	4	3
concentrate	v/n	4	2	philosopher	n	3	3
conclude	v	2	3	physics	n	3	3
consequence	n	2	3	policy	n	1	2
considerably	adv	3	3	positive	a	2	2
consist	v	1	2	potential	a/n	2	2
constant	a	3	2	predict	v	4	2
constitution	n	1	3	previous	a	2	2
construct	v/n	2	3	principal	n	4	2
contract	n/v	1	2	process	n/v	1	1
contribute	v	3	2	promotion	n	4	3
create	v	1	1	range	n	2	2
culture	n	2	1	react	v	3	3
cycle	v/n	4	3	register	v/n	3	3
data	n	1	2	reliable	a	3	3
define	v	1	2	remove	v	3	2
design	n/v	2	1	research	n/v	1	1
document	n/v	3	3	resource	n	2	1
economic	a	1	1	response	n	1	2
element	n	2	2	section	n	1	2
emphasize	v	3	3	security	n	2	2
error	n	4	3	seek	v	2	2
export	v/n	1	2	select	v	2	2
finally	adv	2	1	significant	a	1	3
focus	n/v	2	2	similar	a	1	1

fund	n/v	3	2	site	n	2	2
identify	v	1	2	status	n	4	2
immigrant	n	3	3	stress	n/v	4	1
income	n	1	2	survey	n/v	2	2
individual	n/a	1	1	task	n	3	2
injury	n	2	2	technology	n	3	1
institution	n	2	2	text	n/v	2	2
interpretation	n	1	3	traditional	a	2	1
involve	v	1	1	transfer	v/n	2	3
issue	n/v	1	1	vary	v	1	2
job	n	4	1	volume	n	3	2

Appendix IV

Stimulus words for the L1 association task (Chapter 4). These were translated into Japanese for the Japanese participants.

Word	Word Class	AWL Sublist	JACET Level	Word	Word Class	AWL Sublist	JACET Level
achievement	n	2	2	interpret	v	1	2
acquire	v	2	2	item	n	2	2
affect	v/n	2	1	journalist	n	2	3
analyze	v	1	3	labor	n/v	1	2
apparently	adv	4	2	layer	n/v	3	2
assistant	n	2	2	legal	a	1	2
assumption	n	1	3	locate	v	3	2
attitude	n	4	1	maintain	v	1	1
benefit	n/v	1	2	major	a/n	1	1
civilization	n	4	3	normal	a	2	1
code	n/v	4	2	obtain	v	2	2
comment	n/v	3	2	obvious	a	4	2
communication	n	4	1	occupy	v	4	3
complex	n/a	2	2	parallel	a/n	4	3
concentration	n	4	2	participate	v	2	3
concept	n	1	2	partner	n	3	2
conclusion	n	2	2	percent	n	1	1
conduct	v/n	2	2	perceive	v	2	3
conference	n	4	2	philosophy	n	3	2
constantly	adv	3	3	physical	a	3	2
consumer	n	2	2	previously	adv	2	3
context	n	1	2	principle	n/a	1	2
contrast	v/n	4	2	proceed	v	1	3
contribution	n	3	2	project	n/v	4	1
creative	a	1	2	promote	v	4	2
credit	n/v	2	2	publish	v	3	1
cultural	a	2	2	purchase	v/n	2	2
debate	v/n	4	2	reaction	n	3	2
demonstration	n	3	3	region	n	2	2
designer	n	2	3	rely	v	3	2
dominant	a	3	3	require	v	1	1
economy	n	1	2	researcher	n	1	2
emphasis	n	3	3	respond	v	1	2

environment	n	1	1	restriction	n	2	3
establish	v	1	2	role	n	1	1
estimate	v/n	1	2	selection	n	2	3
ethnic	a	4	3	series	n	4	1
evidence	n	1	2	sex	n	3	2
factor	n/v	1	2	shift	v/n	3	2
feature	n/v	2	2	significance	n	1	3
final	a	2	1	similarly	adv	1	3
financial	a	1	2	source	n	1	1
function	n/v	1	2	specific	a	1	2
goal	n	4	1	strategy	n	2	2
grant	n/v	4	2	structure	n/v	1	2
illustrate	v	3	3	sum	n/v	3	3
impact	n/v	2	2	technique	n	3	2
indicate	v	1	2	textbook	n	2	3
initial	a/n	3	3	theory	n	1	1
injure	v	2	3	tradition	n	2	2

Appendix V

Example of an English Task 1 form (Chapters 5 and 6).

Write the first word that comes to mind for each of the following words. Don't think about them too much as there are no right or wrong answers. If nothing comes to mind immediately, you can move on to the next word and come back to it later, but don't change the answers you have already written.

all	
head	
bike	
basket	
bear	
leap	
throw	
symbol	
chocolate	
wolf	
order	
trick	
permit	
birth	
trouble	
fashion	
gold	
attack	
calm	
blame	
origin	
pig	
choose	
express	
loose	
report	
eye	
voice	
shoulder	
bomb	

growth	
fish	
spider	
amount	
explain	
school	
death	
age	
sweet	
agree	
limit	
talent	
bag	
ladder	
gentle	
stomach	
house	
method	
cupboard	
equal	
string	
routine	
dance	
boot	
fence	
source	
tour	
whistle	
song	
leave	
bone	
appeal	
accept	
afraid	
moon	
extra	
store	
bond	
pride	

cancer	
secret	
heart	
horn	
pot	
ability	
whale	
force	
copy	
cope	
canal	
cloud	
ideal	
able	
window	
market	
rock	
mood	
peak	
labour	
orange	
policy	
country	
baby	
concentrate	
cheese	
lead	
fly	
soldier	
cottage	
blow	

Name: _____

Circle one of the following:

1. I am a native English speaker.
2. English is not my first language, but I do not have any problems using English in my daily life.
3. I have some difficulties using English.

Appendix VI

Example of a "blank" English Task 2 form utilized in the study appearing in Chapter 5 (with cues numbered and individualized cues marked in colour).

Name: Jane Doe

Write the first word that comes to mind for each of the following words. Don't think about them too much as there are no right or wrong answers. If nothing comes to mind immediately, you can move on to the next word and come back to it later, but don't change the answers you have already written.

1	cool	
2	finger	
3	alone	
4	plain	
5	fruit	
6	*filler 1*	
7	*filler 2*	
8	*filler 3*	
9	*filler 4*	
10	*filler 5*	
11	*filler 6*	
12	*filler 7*	
13	*filler 8*	
14	*filler 9*	
15	*filler 10*	
16	*control 1*	
17	*control 2*	
18	*control 3*	
19	offer	
20	*filler 11*	
21	touch	
22	health	
23	devote	

24	solid	
25	weather	
26	spill	
27	needle	
28	bean	
29	*filler 12*	
30	water	
31	much	
32	thick	
33	heaven	
34	another	
35	diet	
36	carry	
37	*filler 13*	
38	band	
39	deep	
40	miracle	
41	*filler 14*	
42	imagine	
43	tax	
44	loss	
45	seek	
46	lower	
47	business	
48	*prime 1*	
49	*prime 2*	
50	*prime 3*	
51	*prime 4*	
52	*prime 5*	
53	*prime 6*	
54	*prime 7*	
55	*prime 8*	
56	*prime 9*	
57	*prime 10*	
58	*target 1*	
59	*target 2*	
60	*target 3*	

Appendix VII

The English version of a Task 2 form for the experimental-first, meaning-based priming condition utilized in Chapter 6 (with cues numbered and individualized cues marked in colour).

Name: Jane Doe

Part 1

For each word below, write a related word. For example, if the word is *cat*, you might write *pet, animal*, or *lion*. Any answer is okay as long as it is related to the meaning of the word printed here. Write only one word for each.

1	admit	
2	carry	
3	business	
4	knowledge	
5	thick	
6	deep	
7	offer	
8	tour	
9	spill	
10	salt	
11	cool	
12	beautiful	
13	water	
14	street	
15	narrow	
16	band	
17	wander	
18	poison	
19	joy	
20	punch	

Part 2

Write the first word that comes to mind for each of the following words. Don't think about them too much as there are no right or wrong answers. If nothing comes to mind immediately, you can move on to the next word and come back to it later, but don't change the answers you have already written.

21	*target 1*	
22	*target 2*	
23	*target 3*	
24	finger	
25	solid	
26	double	
27	health	
28	magic	
29	tax	
30	heaven	
31	tape	
32	future	
33	devote	
34	bath	
35	plain	
36	native	
37	pencil	
38	computer	
39	path	
40	diet	
41	fruit	
42	gas	
43	miracle	
44	*control 1*	
45	*control 2*	
46	*control 3*	

Finally, please circle one of the following:

1. I am a native English speaker.
2. English is not my first language, but I do not have any problems using English in my daily life.
3. I have some difficulties using English.

Appendix VIII

Example of an English Task 2 form for the control-first, meaning-based priming condition utilized in Chapter 6 (with cues numbered and individualized cues marked in colour).

Name: Jane Doe

Part 1

Write the first word that comes to mind for each of the following words. Don't think about them too much as there are no right or wrong answers. If nothing comes to mind immediately, you can move on to the next word and come back to it later, but don't change the answers you have already written.

1	finger	
2	solid	
3	double	
4	health	
5	magic	
6	tax	
7	heaven	
8	tape	
9	future	
10	devote	
11	bath	
12	plain	
13	native	
14	pencil	
15	computer	
16	path	
17	diet	
18	fruit	
19	gas	
20	miracle	
21	*control 1*	
22	*control 2*	
23	*control 3*	

Part 2

For each word below, write a related word. For example, if the word is *cat*, you might write *pet, animal*, or *lion*. Any answer is okay as long as it is related to the meaning of the word printed here. Write only one word for each.

24	admit	
25	carry	
26	business	
27	knowledge	
28	thick	
29	deep	
30	offer	
31	tour	
32	spill	
33	salt	
34	cool	
35	beautiful	
36	water	
37	street	
38	narrow	
39	band	
40	wander	
41	poison	
42	joy	
43	punch	

Part 3

Write the first word that comes to mind for each of the following words. Don't think about them too much as there are no right or wrong answers. If nothing comes to mind immediately, you can move on to the next word and come back to it later, but don't change the answers you have already written.

(This is the same task as Part 1.)

44	*target 1*	
45	*target 2*	
46	*target 3*	
47	lower	
48	imagine	
49	ghost	
50	rescue	
51	bean	
52	alone	
53	needle	
54	seek	
55	loss	
56	another	
57	health	
58	much	
59	*filler*	
60	*filler*	
61	*filler*	
62	*filler*	
63	*filler*	

Finally, please circle one of the following:

1. I am a native English speaker.
2. English is not my first language, but I do not have any problems using English in my daily life.
3. I have some difficulties using English.