Learning the meaning of new words:

Behavioural and neuroimaging evidence

Roberto Andrés Ferreira Campos

Thesis submitted for the degree of Doctor of Philosophy

University of York

Department of Psychology and York Neuroimaging Centre

September 2011

Abstract

This thesis aimed to shed light on the process of word learning and the consequences of storing, retrieving, and using new lexical representations. A number of behavioural experiments and one final fMRI study were conducted. Experiments 1-2 investigated effects of context variability (number of different contexts) and semantic richness (number and type of semantic features) on word learning in a second language. Experiment 1 suggested that context variability benefits word naming and semantic decision. Experiment 2 showed that semantic richness leads to better performance in semantic decision and cued recall, but does not affect word naming or recognition memory. Experiment 3 investigated semantic richness effects across speakers of English L1 and English L2. Results showed that participants did not differ regarding recognition memory and semantic decision; however, L1 speakers outperformed L2 speakers in naming and cued recall. Experiments 4-5 investigated the time course of word learning and examined effects of semantic richness at two different time points. The findings suggested that semantics affects recognition memory, but only a week after training. Effects of semantic richness on categorization and cued recall were found a day and a week after training, with participants showing improvement over time in all conditions. Experiments 6-7 assessed whether improvement over time in Experiments 4-5 was simply due to the passing of time or due to the effect of previous test instances. Results showed that performance improved over time in all tasks when participants were tested on the same novel words, but declined when they were tested on a different set of novel words, suggesting that performance only improved if mediated by a test instance. Experiment 8 was aimed at collecting semantic features from British speakers for 100 familiar words. Finally, Experiment 9 explored the neural correlates of familiarity and semantic richness. Two distinctive brain networks were observed during the categorization of familiar and novel words, consistent with previous findings. Rich semantics was associated with increased activation in conceptual representation areas, whereas poor semantics was reflected in heightened response in semantic control areas. The findings of this thesis have important implications for theories of word learning and semantic memory.

Tables of Contents

Abstract 2
Table of contents 3
List of tables 7
List of figures 9
Acknowledgements 12
Declaration 14

Chapter 1 – Word learning and semantics 15

1.1 Introduction 15
1.2 General aspects of the word learning paradigm 18
1.3 Learning new words in a native language (L1) 20
1.3.1 Phonological word-form learning and the process of lexicalization 20
1.3.1.1 Sleep effects in the consolidation of novel words 22
1.3.2 Learning and integration of novel words with meaning 25
1.3.3 Learning new words from context 29
1.3.3.1 The emergence of the learning-from-context approach 29
1.3.3.2 The encoding variability hypothesis 31
1.3.3.3 The context variability hypothesis 33
1.4 Learning new words in a second language (L2) 35
1.4.1 Learning new words from context in L2. 36
1.4.2 Cognitive aspects of semantic development in L2 37
1.5 The semantic representation of words 40
1.5.1 Concepts and word meanings 40
1.5.2 Representation of word meaning 41
1.5.2.1 Holistic approach 41
1.5.2.2 Featural approach 42
1.5.3 Semantics and word processing 43
1.5.3.1 Number of semantic features (NSF) 45
1.6 Conclusion and thesis overview 47

Chapter 2 – Learning words in a second language: effects of contact and feature variability 49

2.1 Introduction 49
2.2 Experiment 1 50
2.2.1 Method 52
2.2.2 Results 57
2.2.3 Discussion 61
2.3 Experiment 2 64
2.3.1 Method 66
2.3.2 Results 73
2.3.3 Discussion 80
2.4 General discussion 84

Chapter 3 – L1 and L2 novel word processing 89

3.1 Introduction 89
3.2 Experiment 3 91
 3.2.1 Method 92
 3.2.2 Results 93
 3.2.3 Discussion 99
3.3 Experiment 2 and Experiment 3 (combined analysis) 101
 3.3.1 Method. 103
 3.3.2 Results 103
 3.3.3 Discussion 111
3.4 Summary and conclusion 114

Chapter 4 – Learning and consolidation of new words 116

4.1 Introduction 116
4.2 Experiment 4 119
 4.2.1 Method 120
 4.2.2 Results 127
 4.2.3 Discussion 135
  4.2.3.1 Recognition memory 135
  4.2.3.2 Word naming 137
  4.2.3.3 Semantic categorization and word production 138
  4.2.3.4 Summary and conclusion 139
4.3 Experiment 5 140
 4.3.1 Method 141
 4.3.2 Results 143
 4.3.3 Discussion 148
4.4 General discussion 149

Chapter 5 – Effect of testing on the consolidation of newly learned words 152

5.1 Introduction 152
5.2 Experiment 6 153
 5.2.1 Method 154
 5.2.2 Results 155
 5.2.3 Discussion 160
5.3 Experiment 7 161
 5.3.1 Method 162
 5.3.2 Results 162
 5.3.3 Discussion 166
5.4 General discussion 168

Chapter 6 – Neural correlates of semantic richness for familiar and novel words 171

6.1 Introduction 171
6.2 Review of relevant studies 173
6.2.1 Neuroimaging studies of word learning 173
6.2.2 Studies of semantic memory 177
6.3 Theoretical motivation and chapter outline 183
6.4 Experiment 8 186
6.4.1 Method. 187
6.4.2 Analysis and selection of features 190
6.4.3 Conclusion 191
6.5 Experiment 9 191
6.5.1 Method 195
6.5.2 Results 208
6.5.3 Discussion 228
6.5.3.1 Behavioural findings. 229
6.5.3.2 fMRI findings 231
6.5.3.3 Summary and conclusion 240

Chapter 7 – Thesis summary and conclusions 242

7.1 Summary of main findings 242
7.2 Implications of the current findings 247
7.3 Strengths 249
7.4 Weaknesses 252
7.5 Future directions 254

Appendices 257

Appendix 2.1 257
Appendix 2.2 258
Appendix 2.3 259
Appendix 2.4 262
Appendix 2.5 263
Appendix 2.6 266
Appendix 2.7 267
Appendix 4.1 268
Appendix 4.2 268
Appendix 4.3 269
Appendix 4.4 271
Appendix 4.5 272
Appendix 4.6 272
Appendix 4.7 273
Appendix 4.8 275
Appendix 4.9 276
Appendix 5.1 277
Appendix 6.1 278

Appendix 6.2 280
Appendix 6.3 282
Appendix 6.4 284
Appendix 6.5 285
Appendix 6.6 286
Appendix 6.7 287
Appendix 6.8 291

References 292

List of Tables

Table 2.1. Mean naming latencies and percent error rates in the naming and semantic decision task……….......…………………………………..…………58
Table 2.2. Sample set of sentences corresponding to rich consistent semantics (RCS), rich inconsistent semantics (PCS), and rich inconsistent semantics (RIS)……………………………………………….....……68
Table 2.3. Mean latencies and percent error rates for word naming, recognition memory, and semantic decision. Percent words recalled in cued recall………………………………………………....……..74
Table 3.1. Mean latencies and percent error rates for word naming, recognition memory, and semantic decision. Percent accuracy for cued recall……………………………………………………………....…..94
Table 4.1. The novel word *adertmon* presented in 4 sample sentences in the trained conditions: rich semantics, poor semantics, and no semantics…………………………………………………………..122
Table 4.2. RTs for recognition memory, word naming, and semantic categorization; percent accuracy of response for word production on day 3 and day 8……………………………………………………...…127
Table 4.3. RTs corresponding to recognition memory and semantic categorization; percent accuracy of response for word production on day 2 and day 8………………………………………….........143
Table 5.1. RTs corresponding to recognition memory and semantic categorization; percent accuracy of response for word production on day 2 and day 8...155
Table 5.2. RTs for recognition memory and semantic categorization; percent accuracy of response for word production on day 2 and day 8...163
Table 6.1. Mean number of semantic features (NSF) for the words selected in Experiment 8...190
Table 6.2. Sample sentences in the rich semantics and the poor semantics conditions…………………………..……………………………..196
Table 6.3. Characteristics of familiar and novel words used in Experiment 9…………………………………………………………………202
Table 6.4. Semantic features listed for the word *cheese* in McRae et al. (2005)……………………………………………………..……….203
Table 6.5. Behavioural results in the semantic categorization task.………..…..….208
Table 6.6. Mean number and proportion (%) of features recalled for novel words……………………………………………………….……...213
Table 6.7. MNI coordinates for peak voxels showing increased

activation for familiar words versus novel words…………………………….…..216
Table 6.8. MNI coordinates for peak voxels showing higher
activation for novel words versus familiar words……………………………...217
Table 6.9. MNI coordinates for corrected peak voxels showing
increased brain activity for familiar poor versus familiar rich........……………219
Table 6.10. MNI coordinates for corrected peak voxels showing
higher brain activity for novel rich versus novel poor…………………..……...220
Table 6.11. MNI coordinates for corrected peak voxels showing
increased brain response for novel poor versus novel rich……………………..221

List of Figures

Figure 2.1. Structure of training procedure in Experiment 1………………………..56
Figure 2.2. Percent errors and reaction times (ms) for high context variability (HCV), low context variability (LCV), and no training (NT) in the word naming task...59
Figure 2.3. Percent errors and reaction times (ms) for high context variability (HCV) and low context variability (LCV) in the semantic decision task……………..60
Figure 2.4. Structure of training procedure in Experiment 2…………….………….71
Figure 2.5. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS) and minimal training (MT) in the word naming task………..75
Figure 2.6. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics (RIS) in the recognition memory task…………….……..77
Figure 2.7. Semantic decision. Percent errors and RTs for semantically related (R) and unrelated (U) words in rich consistent semantics (RCS) and poor consistent semantics (PCS)………………….…......78
Figure 2.8. Percent words recalled in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars show standard error (SD) of the mean...80
Figure 3.1. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS) and minimal training (MT) in the word naming task…..………95
Figure 3.2. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics (RIS) in the recognition memory task..............................96
Figure 3.3. Semantic decision. Percent errors and RTs for semantically related (R) and unrelated (U) words in rich consistent semantics (RCS) and poor consistent semantics (PCS)……………………………………..97
Figure 3.4. Percent words recalled in rich consistent semantics (RCS) and poor consistent semantics (PCS)…………………………………………….99
Figure 3.5. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS), and minimal training (MT) in the word naming task in L1 and L2 speakers...104
Figure 3.6. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), and rich

inconsistent semantics (RIS) in the recognition memory task for L1 and L2 speakers…………………………………………………………………...106
Figure 3.7. Semantic decision. Percent errors and RTs in L1 and L2 speakers for semantically related and unrelated words in rich consistent semantics (RCS) and poor consistent semantics (PCS)…......……....108
Figure 3.8. Cued recall. Percent words recalled by L1 and L2 speakers in rich consistent semantics (RCS) and poor consistent semantics (PCS)…………...110
Figure 4.1. Structure of training procedure in Experiment 4 (day 1).......................124
Figure 4.2. Percent errors (PE) and reaction times (RT) in rich semantics (RichSem), Poor semantics (PoorSem), and No semantics (NoSem) on day 3 (D3) and day 8 (D8)………………………......128
Figure 4.3. Percent errors (PE) and reaction times (RT) on day 3 (D3) and day 8 (D8) in rich semantics (RichSem), poor semantics (PoorSem), no semantics (NoSem), and untrained……………………………..130
Figure 4.4. Percent errors (PE) and reaction times (RT) in the semantic categorization task performed on day 3 (D3) and day8 (D8)…………………...132
Figure 4.5. Percent accuracy of response for rich semantics (RichSem) and poor semantics (PoorSem) on day 3 and day 8 in the production task……...134
Figure 4.6. Percent errors and RTs in the recognition memory task on day 2 and day 8...144
Figure 4.7. Percent errors and RTs in the semantic categorization task on day 2 and day 8…………………………………………………………...146
Figure 4.8. Percent accuracy of response in the production task on day 2 and day 8.... 147
Figure 5.1. Percent errors and RTs in the recognition memory task on day 2 and day 8……………………………………………………………156
Figure 5.2. Percent errors and RTs in the semantic categorization task on day 2 and day 8…………………………………………………………...158
Figure 5.3. Percent accuracy of response in the production task on day 2 and day 8...159
Figure 5.4. Percent errors and RTs in the recognition memory task on day 2 and day 8…………………………………………………………...163
Figure 5.5. Percent errors and RTs in the semantic categorization task on day 2 and day 8…………………………………………………………...165
Figure 5.6. Percent accuracy of response in the production task on day 2 and day 8...166
Figure 6.1. Brain regions showing response differences between novel and existing words…………………………………………………………...175
Figure 6.2. Brain regions showing increased activation for new words

compared to real words as participants learn meaning from context…………...176
Figure 6.3. Signal level in the three subdivisions vAG (in yellow),
mAG (in blue), and dAG (in red) of the AG..180
Figure 6.4. Semantic foci in Binder et al. (2009) for general semantic
contrasts ....………………………………………………….……182
Figure 6.5. Correlation between number of features (North America)
and number of features (United Kingdom)…………………………………..191
Figure 6.6. Sample images used in rich semantics and poor semantics…...............196
Figure 6.7. Structure of training procedure in Experiment 9 (day 1)……………...197
Figure 6.8. Sample presentation of the stimuli during the semantic
categorization task in the scanner………………………………………………201
Figure 6.9. Mask showing location of ventrolateral angular gyrus (vAG)...............205
Figure 6.10. Mask showing location of left and right precuneus.............................206
Figure 6.11. Mask showing location of left and right pars opercularis...................207
Figure 6.12. Semantic categorization RTs and errors for familiar and
novel words with rich and poor semantics across two presentations………......209
Figure 6.13. Mean number and proportion of features (%) recalled in
the feature recall task for novel words…………...…………………………......213
Figure 6.14. Thresholded (Z=2.3) brain images of the contrasts
familiar words versus novel words (in blue) and novel words
versus familiar words (in red)..218
Figure 6.15. Thresholded (Z=2.3) brain images of the contrasts
novel rich versus novel poor (in blue) and novel poor versus
novel rich (in red)…..……..222
Figure 6.16. Results of ANOVAs conducted on percent signal
change in three regions of interest: ventrolateral angular gyrus
(vAG), precuneus, and pars opercularis………………………………………...224

Acknowledgements

Above all, I am especially grateful to my supervisor, Andy Ellis, for his excellent guidance and constant support throughout the years. His patience and kindness, as well as his academic experience, have been invaluable to me.

I would also like to express my gratitude to the members of my research committee, Gareth Gaskell and Piers Cornelissen, for their sound advice and suggestions as I was progressing in my research. I am also immensely grateful to Silke Göbel for her significant contribution to the development of the fMRI project of this thesis. Her support, teaching and guidance were extremely important in achieving the project aims. I would also like to thank Silvia Gennari for her assertive comments and suggestions at moments when help was much appreciated.

I would like to acknowledge the contribution of the Gaskell Lab group to my formation as a researcher. I am very grateful to all of its members for giving me the opportunity to develop essential research skills throughout the reading, analysis, and discussion sessions.

I would also like to thank the members of the York Neuroimaging Centre (YNiC) for allowing me to use their research facilities and for providing me with all the support I needed. I am grateful to Mark Hymers for his collaboration in the fMRI project and to Andre Gouws for his help and advice on neuroimaging data analysis. I would also like to thank Anastasia Roshchupkina for her excellent assistance in running participants and helping in data analysis.

I am also indebted to my many student colleagues in the Department of Psychology, for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to my office mates of C224 Vicky, Jakke, Mark, Claire, Chris, Gitte, Helen, Elaine, Gina, and Alex for making my stay both interesting and pleasant. I also wish to thank the informal support and encouragement of my friends back home in Chile and here in the United Kingdom. I would particularly like to thank Juan, Isabel, Maritza, Francisco, Andrea, Rana, Victor, Elizabeth, Migue, Alexandru, Alejandro, Debborah, Salomé, Silvana, Irene, Bere, Chris, José, and Carolina.

I will always be immensely grateful to my 'polola', María, who proofread a big part of this thesis. She also sincerely deserves my greatest recognition and admiration for her encouragement, and unconditional emotional support throughout the writing-up period.

Lastly, I wish to thank my parents, Ana Campos and Manuel Ferreira, and my siblings Cecilia, Ana, Carlos, David, Mónica, and Sergio, who have been a constant source of inspiration and moral support throughout my life. To them, I dedicate my thesis.

A mis padres…

Quiero aprovechar esta oportunidad para agradecer a mis padres todo el cariño y el amor que me han entregado. Me siento afortunado de poder ser receptor de sus enseñanzas y valores. Al final de este largo camino, solo me resta decirles que me siento tremendamente agradecido y orgulloso de ser su hijo.

Esta tesis está dedicada a ustedes.

Declaration

Hereby I declare that this work has been completed by the candidate under the supervision of Prof Andrew W. Ellis.

This research was supported by a studentship from the Comisión Nacional de Investigación Científica y Tecnológica, Gobierno de Chile (CONICYT).

Part of the work in this thesis has been presented at the following conferences:

Ferreira, R. A. & Ellis A. W. (2010). *Semantic information influences the recognition of newly learned words over time*. Paper presented at The First joined conference of the Experimental Psychology Society (EPS) and the Sociedad Española de Psicología Experimental (SEPEX), Granada, Spain.

Ferreira, R. A. & Ellis A. W. (2009). *Exploring semantic involvement in the recognition and production of newly learned words: Evidence from L1 and L2 English speakers*. Paper presented at The British Psychological Society, University of Hertfordshire, Hatfield, UK.

Ferreira, R. A. & Ellis A. W. (2009). *The provision of semantic information can influence novel word learning: Evidence from L1 and L2 English speakers*. Paper presented at Psycholinguistics in Flanders, Ambers, Belgium.

Chapter 1 – Word learning and semantics

1.1 Introduction

Human beings have the outstanding capacity to learn and use a language to communicate in a variety of forms. Children are able to quickly become fluent in the language spoken by their community and without any apparent effort. This is not surprising since children need to learn new words every day in order to communicate their thoughts and emotions as they grow up. Adults, on the contrary, do not seem to be in need of learning new vocabulary since they are able to use their mother tongue fluently to communicate even very abstract ideas without too much difficulty. This might be the reason why, within the scientific community, there has generally been more interest in studying word learning in children than in adults. Another reason might be that children, compared to adults, are expert language learners since they seem to learn new words simply by associating them with concepts they find in the environment without any explicit instruction (Bloom, 2001).

It is commonly thought that children start to learn their first words at around twelve months of age. They seem to start off learning words relatively slowly but somewhere between the acquisition of 20 and 100 words, there is a vocabulary 'spurt' that propels children to learn words at an incredibly fast rate of 10 or 20 new words per day (Bloom & Markson, 1998). Theories based on computational models have postulated the development of mechanisms that build on the first words in order to allow faster learning for the words that follow (McMurray, 2007). This means that words acquire information at a constant rate but the outcome becomes quicker as more linguistic, psychological, and statistical factors integrate during learning. It has also been proposed that the 'word explosion' phenomenon in children might be simply due to maturation of memory and attentional capacities (Bloom et al., 1998). Whatever the reasons for children's impressive ability to learn new vocabulary, this process does differ from that of adults, but it also has many similarities.

Adult word learning is most commonly associated with the learning of a second language (L2). Many of us have found ourselves trying to acquire new words in a foreign language and the process seems much harder than when we 'joyfully' learned our mother tongue. It is precisely this complexity that has brought about widespread interest in the area in the last decades. Learning new words in L2 involves processes similar to those experienced in L1 such as the learning of phonology, orthography, syntax, and semantics. However, the development of the mechanisms underlying word learning seems to differ much more extensively. Unlike L1 learners, when L2 learners acquire new words they go through the process of restructuring previously existing lexical and semantic representations in order to integrate new lexical entries. Thus, the new language is always dependent on the first language for the development of its own lexical and semantic representations. Particularly illustrative is the case of semantic development, which is a slow and very often incomplete process (Jiang, 2004). Even though L2 speakers already have a semantic system in their native language, the development of semantics in the second language introduces new lexical form-meaning mappings which differ from those in L1, especially regarding peripheral, figurative, and connotational meaning. Some researchers have argued that complete L2 semantic system is never achieved and even very proficient L2 speakers are unable to learn finer distinctions of some concepts in the second language (e.g., Jiang, 2000; Selinker & Lakshmanan, 1992).

Despite the fact that most behavioural evidence has repeatedly shown that L2 learning is a very laborious process for adults, recent neurophysiological evidence seems to put these ideas into question. Osterhout et al. (2008) conducted a study that tested American university students after one year of classroom in French (L2). They measured the intensity of two ERP components; the N400, which responds to semantic analysis, and the P600, which is sensitive to syntactic violations. They found that semantically and syntactically anomalous words elicited N400 and P600 effects for both English (L1) and French (L2). Interestingly, the P600 effect was much bigger for L1 than L2, whereas the N400 effect did not differ greatly across languages. Osterhout et al. concluded that L2 learners might quickly incorporate L2 meanings into the online processing system in a manner that resembles that of the L1. They added that syntactic information might follow the same pattern but its integration takes longer than that of semantics. Importantly, the N400 effect had been demonstrated in a previous study in which participants only received classroom instruction on L2 vocabulary for about 14 hours (McLaughlin, Osterhout, & Kim, 2004). Put together, these findings have demonstrated that adults might be better learners than initially thought since brain responses to recently acquired L2 words closely resemble those to L1 words.

While child word learning and adult second language learning have accumulated a sizeable number of studies over a few decades of investigation, much less time has been devoted to the research of word learning in adults' first language. Many of the words we learn as adults in our mother tongue seem to go unnoticed since they quickly become popular and start being used just like previously existing words. For instance, the emergence of new technologies and their sub-products have incorporated new words in our mental lexicon such as *ipad*, *blog*, or *facebook*. All these new lexical entries were once unknown to all of us and only recently became part of our everyday vocabulary. This phenomenon is a good example to illustrate that we do learn new words in our first language, and that we seem to perform fairly well.

Although the overall number of studies in adult (L1) word learning is less numerous than in child or adult (L2) word learning, in the last few years, this number has been on the increase. It is worth noting that in the L1 literature, the term 'word learning' is rather ambiguous and it does not always correspond to the same thing. This might be explained by the fact that researchers have different interests and they have focused on very specific aspects of word learning. Some have mainly studied spoken word learning giving emphasis to the learning of phonology (e.g., Dumay, Gaskell, & Feng, 2004), while others have put the accent on meaning and/or the mechanisms involved in word learning (e.g., Breitenstein, Kamping, Jansen, Schomacher, & Knecht, 2004; Mestres-Missé, Münte, & Rodriguez-Fornells, 2008).

Those whose interest is primarily phonology often use the term word learning for a type of phonological word-form learning (e.g., Gaskell & Dumay, 2003; Tamminen & Gaskell, 2008). In these studies, participants are generally trained on an artificial spoken lexicon (no meaning involved) and are tested immediately after training or a few days later using a variety of tasks such as cued recall, phoneme monitoring, lexical decision, etc. These studies have mainly shown that new phonological word-forms can acquire word-like characteristics after just a few exposures if followed by a period of consolidation that requires sleep. Other word learning studies have used the term word learning to refer to a type of associative learning. This type of learning does include semantics and involves simple frequent repetitions of stimulus configurations without the necessity for explicit feedback (Breitenstein et al., 2004). It is a methodology that has been used successfully in healthy participants and patients suffering from aphasia and it has demonstrated that extensive training of new vocabulary for less than a week shows excellent retention of pair associations over time (Breitenstein & Knecht, 2002).

Probably the earliest method used for word learning in L1 is contextual learning or incidental learning. This method has traditionally been applied in the context of a classroom and has targeted young children or teenagers who need to acquire new vocabulary as part of their courses. Investigation using this approach has produced numerous studies at different time periods. These studies have mainly demonstrated that both children and adults are capable of learning new words simply by extracting their meaning from the context in which they are presented (e.g., Jenkins, Stein, & Wysocki, 1984; Nagy & Herman, 1987; Nagy & Scott, 2000; Cain, 2007; Bolger, Balass, Landen, & Perfetti, 2008).

As shown above, adult word learning in L1 and L2 has taken different perspectives depending on researchers' interests. In the following sections, a review of the literature on L1 and L2 adult word learning and the factors that intervene in the process of acquiring new words will be presented. The review will also include relevant studies of semantics and its role in word processing.

1.2 General aspects of the word learning paradigm

In the last two decades, there has been an exponential increase in the number of studies that have used a word learning paradigm. These studies have provided evidence for a variety of factors that can influence the acquisition of new vocabulary such as semantic context (Bolger et al., 2008), sleep (Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010), previous lexical knowledge (Dahan & Brent, 1999), and statistical information (Saffran, Aslin, & Newport, 1996), among others. Additionally, word learning studies have gathered converging evidence that language users can very quickly learn new words, and that retention can last long even without additional exposures (e.g., Salasoo, Shiffrin, & Feustel, 1985; Tamminen & Gaskell, 2008). Beside the contribution to the understanding of word learning, these findings have provided a validation for the use of the word learning paradigm to test a variety of hypotheses in different language areas.

Word learning, as a paradigm, has many advantages over other methods of investigation since it allows the researcher in different domains of language acquisition and processing to overcome stimulus control problems by examining only the factors of interest. For instance, in order to achieve precise control over word frequency and phonological similarity, Magnuson, Tanenhaus, Aslin, and Dahan (2003) used artificial lexicons to examine the time course of lexical activation and competition in three eye tracking experiments. They argued that using a novel lexicon has proven very useful, especially to evaluate the microstructure of spoken language comprehension. Overall, Magnuson et al. recommended the use of artificial lexicons to complement traditional studies because they allow researchers to create stimuli to test specific hypotheses reducing the number of uncontrolled variables. They also point out that artificial lexicons show very little intrusion effects from participants' native lexicon. These effects occur very early during training and disappear as training progresses and novel words become more integrated in the mental lexicon.

The word learning paradigm has also been useful in testing the effects of some neurotransmitters during the encoding of new material. For instance, in a study conducted by Knecht et al. (2004), participants were given either 100mg of the dopamine precursor Levodopa or a daily placebo dose over the course of 5 days. Ninety minutes after taking their corresponding dose, participants received training on a novel vocabulary using simple repetition tasks. Results demonstrated that participants learned the novel words faster and with better retention over time in the Levodopa condition. The importance of the paradigm here is that it allowed scientists to track the effects of the neurotransmitter as participants were encoding new linguistic material, which would have been impossible with a different method.

The artificial lexicon approach has also been applied to study the development of early language mechanisms in infants aiming at assessing the contributions of innate and acquired knowledge (see Gómez & Gerken, 2000, for review). This approach has permitted the identification of fine-grained characterizations in infant word learning mechanisms by testing aspects such as word order, recognition of word units in speech, and generalization of grammatical relations.

To sum up, the flexibility of artificial lexicons allows researchers to control for variables, otherwise impossible to control. The approach has proved very reliable and it can be applied to different domains of language learning and language processing in general.

1.3 Learning new words in a native language (L1)

Studies in this section are classified according to the focus of research. Thus, the first section covers studies which have primarily focused on phonology and the process of lexicalization, while the next section includes studies that have emphasised the learning of semantics in lexical integration and have primarily used an associative learning paradigm. The last section presents contextual learning studies which comprise research into how the meaning of unknown words can be derived from linguistic contexts. Even though some studies might correspond to more than one category, they were only included in the category which seemed most relevant.

1.3.1 Phonological word-form[1] learning and the process of lexicalization

In these studies, participants are normally exposed to new phonological forms a certain number of times over the course of, for instance, a phoneme-monitoring task session. They are then tested immediately after training or on different consecutive days depending on the purpose of the research. Language tasks such as word naming, lexical decision, and pause detection, among others, are of common use as they provide relevant measures of word recognition. The focus of many of these studies is not the immediate effects of factors such as speaker gaze and syntactic context on word acquisition, but rather the time-course of novel phonological-word forms entering the mental lexicon and then competing for lexical selection with real words. This area of research is particularly important since it aims at disentangling the mechanisms involved in word learning, which have been neglected in the past.

The notion that words compete for selection, a process called *lexical competition*, comes from theories of word recognition which suggest that people recognize words by first activating multiple matching lexical candidates, but as more speech information is encountered, the number of candidates is gradually reduced until there is only the target item left (e.g. Marslen-Wilson, 1987; McClelland & Elman, 1986). This idea is in line with the fact that acoustic information unfolds over time and so can activate multiple competitors in the listener's mental lexicon

[1] In this section, the term *phonological word-form* is sometimes referred to simply as *word*.

(Marslen-Wilson, 1993). According to this view, all words have a uniqueness point (the point at which no other words are competing for selection). This uniqueness point can affect a word's recognition time so that, for example, if they have an early uniqueness point they can be recognized faster. This effect has been demonstrated in a number of studies (see McQueen, 2007, for review).

In a pioneering word learning study looking at lexical competition, Gaskell and Dumay (2003) proposed that in order for novel words to become part of the mental lexicon they have to be fully engaged in lexical competition with similar-sounding existing words. They added that only by examining lexical competition effects of newly learned words on existing words, it is possible to hypothesise about the nature of the memory trace created when a new vocabulary item is stored. Gaskell and Dumay showed that newly learned words can be easily recognized immediately after training even though they do not become engaged in lexical competition until a week later. This was demonstrated in three experiments. In Experiment 1, participants performed a lexical decision task immediately after hearing 12 oral presentations of the novel items. They found good recognition rates but no competition effects of novel words on the recognition of similar-sounding existing words. In Experiment 2, participants were tested over the course of five days, so they were allowed to see at which point in time new words were to become engaged in lexical competition. Lexical competition effects were still not present on day 2, but decision latencies for phonologically similar words with offset, divergent from the novel words gradually increased throughout the days; so on day 5 there was a clear inhibition effect on existing words. Finally, almost the same effects were replicated in a pause detection task used in Experiment 3; the only difference was that comparisons were made between day 1 and day 8. The lexical competition effect has also been found for written words, particularly in a study conducted by Bowers, Davis, and Hanley (2005). They had participants learn new words that corresponded to orthographic neighbours (e.g., *banara*) for existing words that did not have any previous written neighbours (e.g., *banana*). Bowers and colleagues showed that the novel words were able to compete for activation with the existing words one day after training with the effect becoming stronger on the following day.

Given the above evidence, it appears that *lexicalization*, the stage of learning at which words can be engaged in lexical competition, occurs some time after training and it is distinct from simple storage. However, the exact moment in time when this process takes place seems less clear. For this reason, it would be interesting to examine which factors determine the appearance of a new lexical entry. Certainly the role of time and level of exposure can make a difference in the lexicalization process since it is not as simple as phonological storage, whose effects are seen immediately after training. It is possible that lexicalization is not automatic and it may require a considerable amount of time to allow the consolidation of episodic traces (O'Reilly & Norman, 2002).

One of the few studies that has investigated adult word learning over long periods of time was conducted by Salasoo et al. (1985). They presented participants with words and nonwords and after only 6 visual exposures to the items, recognition accuracy for words and nonwords showed no significant difference. Most importantly, when the same participants were tested a year later, they still showed good recognition but with a decline in performance, which affected both words and nonwords equally. These findings suggest that novel words can be recognized very accurately shortly after training and can become interleaved with previous knowledge for long periods of time.

In a more recent study, Tamminen and Gaskell (2008) investigated the long-term effects of word learning using a similar methodology to that of Gaskell and Dumay (2003). They included phoneme monitoring and repetition tasks in order to measure lexical competition. The results of their study showed that newly learned words were still engaged in lexical competition with existing words even 8 months after initial exposure.

Given the above evidence, it seems that words need some time for consolidation so they can acquire the same characteristics of pre-existing items in the mental lexicon. Additionally, it appears that the lexical competition effects, which generally occur at least a day after training, are long lasting since novel words can interfere with existing words months or even a year after training. It is important to note, however, that in all the studies reviewed above, participants were only exposed to the phonological or orthographic forms of the novel words.

1.3.1.1 Sleep effects in the consolidation of novel words

Dumay and Gaskell (2007) proposed that words do not engage in lexical competition immediately after training because they need to undergo an incubation-like period of consolidation overnight in order to form new representations in the lexical memory. More specifically, they suggest that new words are first stored in the hippocampus and are only transferred to neocortical areas for long-term storage during a night's sleep. The idea of sleep playing a key role in memory consolidation has been around for a considerable amount of time, but only recently has been applied to word learning. Memory consolidation over sleep has previously been reported in the procedural domain (e.g. Maquet, 2001; Stickgold & Walker, 2005), and there is consistent evidence showing that perceptual and motor skills learned during a period of wakefulness can be improved after sleep (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994); Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002).

In one of the first word learning studies looking at memory consolidation during sleep, Dumay and Gaskell (2007) assessed whether nocturnal sleep was involved in the lexicalization of newly learned phonological-word forms. Sixty-four participants were divided into 2 groups and were assigned identical experimental sessions except for the time in which they took place. Using a phoneme monitoring task, participants in both groups were presented each word 36 times and in random order. One group learned the words at 8 p.m. and the other at 8 a.m. Both groups were tested three times following training (immediately after, 12 and 24 hours later) in order to investigate whether lexicalization only takes place after a night's sleep or is simply due to the passing of time. While recognition reached ceiling in all sessions and in both groups, a pause detection task revealed a clear association between sleep and the emergence of a change in lexical activity after the acquisition of a new competitor. Exposures to novel words had no effects on cohort existing words when measured immediately after training. However, latencies increased for existing words with novel word competitors after 12 hours but only for the group that had a night's sleep. As for the a.m. group, no competition effects were observed after 12 hours. Nevertheless, similar patterns to those found in the p.m. group were found in the a.m. group 24 hours later when they also had a period of sleep. These results suggest that the lexicalization of newly learned phonological word-forms takes place during nocturnal sleep and it is not simply due to the passing of time.

In order to explore the neural basis of spoken word learning and consolidation overnight, Davis, Di Betta, Macdonald, and Gaskell (2008) conducted an fMRI experiment that compared activation for existing and novel words learned 1 day before scanning and on the same day. They found increased activation for novel words compared with familiar words in a number of brain regions including the left superior temporal gyrus (STG), inferior frontal and premotor regions, and right cerebellum. Interestingly, activity for untrained and trained items did not differ when the trained items were learned on the same day of scanning. However, activation was reduced for items that were learned 1 day prior to scanning. Elevated hippocampal activation was only found for untrained novel words, but consolidated and unconsolidated novel words did not differ from each other. These findings suggest that overnight sleep produces neural changes in regions involved in phonological processing and speech motor integration, which allows consolidated novel words to acquire word-like characteristics.

A more recent study by Tamminen et al. (2010) examined the sleep-related electrophysiological basis of lexical integration and novel word recall. They looked at sleep spindle[2] activity after participants learned new spoken words. Sleep spindle activity has been previously associated with improvements in procedural and declarative memory (e.g., Schmidt et al., (2006). Tamminen et al. used a similar methodology to that used in the study by Dumay and Gaskell (2007). Consistent with previous studies, they found that recall of novel items improved overnight but not during the awake period, and that lexical competition only emerged after a night of sleep. More importantly, they also found that higher number of sleep spindles predicted a larger increase in the magnitude of the lexical competition effect. This finding suggests that sleep spindles play an important role in the integration of newly learned words overnight.

Phonological word-form learning studies have been of great importance in the word learning literature since they have investigated the time-course of novel words entering the mental lexicon and becoming integrated with existing novel words. They have combined behavioural, electrophysiological, and neuroimaging data to assess lexical competition effects before, during, and after consolidation. They have also demonstrated that sleep plays a fundamental role in the integration of novel words. Despite the fact that the studies reviewed so far have provided striking new information on the process of learning and consolidation of novel linguistic material, they have completely neglected the role of semantics. This is surprising

[2] Sleep spindles are 11-15 Hz oscillations that can last up to 3 seconds.

given that the purpose of learning and using a language is to communicate meaning and not only phonological word-forms. Thus, the studies reviewed above can only account for phonological learning, which might differ in many aspects from word learning that includes semantics.

1.3.2 Learning and integration of novel words with meaning

One way to address the learning of new words and their meanings is by using associative learning, which consists of simple frequent repetitions of stimulus configurations without the necessity for explicit feedback (Breitenstein et al., 2004). This paradigm has proved successful for teaching new words to healthy participants and patients (e.g., suffering from aphasia). The main aim in associative learning studies is the mapping of novel words to their meanings.

In a study conducted by Breitenstein and Knecht (2002), participants were presented with drawings and spoken novel words. They were required to indicate whether the pairings were correct or incorrect. Each novel word had a high frequency of occurrence with one drawing and a very low frequency of occurrence with other drawings. The aim of the experiment was to find out whether participants were capable of deriving the statistical occurrence of the pairs and so learning to associate the correct drawings with the correct words. During 5 days of training, participants improved from chance level, in the first session, to 90% in the last session. Participants were tested again on two occasions, 1 week and 1 month later, and performance remained steady overtime. These results showed that individuals can successfully learn new words using a paired-association paradigm and, more importantly, with good retention over time.

In spite of the successful results showed by associative learning, there are doubts whether the newly acquired word forms become integrated in the mental lexicon and behave like existing words.

> Learning novel words by simple associative principles may result in mere superficial couplings of highly specific acoustic and visual information, with no connection between the novel word forms and semantic/conceptual information in long-term memory (Breitenstein, et al., 2007).

Breitenstein et al. (2007) emphasized that the pairing of sound and meaning should provide a clear link between the visual stimuli and the conceptual level of the language system making generalization possible. Another factor that should be taken into account is timing; successfully learned words should behave like familiar words in both speech production and comprehension. Since language users can fluently speak and effortlessly understand up to three words per second (Cutler & Clifton, 1999; Levelt, 1989), newly acquired words should be processed in a similar fashion to be considered fully integrated words in the language system.

One way to find out whether associative learning allows newly learned words to become fully integrated in existing lexical/semantic networks would be to use semantic priming. In a follow-up study conducted by Breitenstein et al. (2007), with a similar methodology to that of Breintenstein and Knecht (2002), participants underwent training in a novel vocabulary (45 words) for 5 consecutive days. As in previous associative learning studies, vocabulary training involved higher co-occurrence of “correct” arbitrary object and novel word pairings as compared to “incorrect” pairings. The results of this study replicated previous findings showing that participants could successfully learn the novel vocabulary. More importantly, learning went beyond isolated stimulus-stimulus associations since trained novel words were able to prime target pictures in a semantic decision task. This was taken as evidence that the newly learned words were successfully linked to the concepts and so were behaving like real words.

In an ERP study that included paired-association with words and sentences, McCandliss, Posner, and Givon (1997) looked at changes in brain circuitry after participants learned a miniature vocabulary, which they called ‘Keki’, during the course of 5 weeks. They spent a total of 50 hours learning the language with lessons consisting of pictures depicting simple objects in isolation or in small groups with words or sentences presented at the bottom. The procedure gradually increased in complexity as participants learned. Eventually, sentences progressed to the level of short descriptions of situations that presented characters interacting with objects and other characters. At the end of the training session, participants were tested using passive viewing, semantic judgment and feature search tasks while ERPs were recorded. Analyses were conducted on an early N1 window (170-230ms), and a later P2 window (280-360ms). An orthographic effect was found in the first window. English words produced the least negativity followed by Keki words and Keki control, which did not differ from each other. Consonant strings elicited the most negative N1. Training had no effects on the N1 and orthographic effects were purely explained by regularity. Hence, the greater the orthographic regularity of the letter string the less negativity occurred in the N1 window, which explained the difference between Keki words and English words. In the P2 window, results varied according to the task. In the semantic task, English words elicited the least positivity and Keki words showed reduced positivity with respect to consonant string and untrained Keki words. This difference was not found in the viewing task or the feature search task. These data indicated that participants learned the meaning of the Keki words and that their meaning modulated the amplitude of the P2.

In an eye-tracking study conducted by Magnuson, Tanenhaus, Aslin, Dahan (2003), participants were required to learn 16 bisyllabic new words by associating them with new shapes. Novel words were organized in four 4-word sets such as /pibo/, /pibu/, /dibo/, and /dibu/, so some words had onset-matched competitors (e.g., *pibo* – *pibu*) or rhyming competitors (e.g., *pibo* – *dibo*). They manipulated the frequency of novel words during training and the frequency of their onset competitors and rhyme competitors. The aim of this study was to examine the time-course of lexical activation and competition. The results showed that the novel vocabulary behaved similarly to real words since onset-matching words competed for activation up to the point of diversion from the target word, and rhyme competitors showed a delayed and much weaker effect. Interestingly, high-frequency targets received more fixations than low-frequency targets, and the same was true for high- and low-frequency competitors. Lexical competition effects were also observed since targets presented with low-frequency competitors recorded more fixations than targets presented with high-frequency competitors.

In another study that involved learning novel words with meaning, Leach and Samuel (2007) investigated the role of semantics in perceptual learning (implicit knowledge). The study comprised 5 experiments that required participants to learn a novel vocabulary for 5 consecutive days. Participants were exposed to a variety of tasks that included the presentation of words with meaning or without meaning. They found that perceptual learning was achieved when novel words were associated with semantics during training, either using word-picture association or a short passage in which the novel words were presented. However, when only a phoneme monitoring task was used during training, participants did not show any reliable effects of perceptual learning. Importantly, when participants were given a test of explicit knowledge, performance was good independent of the training received (with or without meaning). These results were inconsistent with a previous study in which semantics did not improve lexicalization (Dumay et al., 2004). An important difference between the two studies is that in Leach and Samuel (2007) words trained with meaning where given a much richer meaning (a picture or a story context) than in Dumay et al. (one feature and a sentence context), and received 5 exposures compared to 2 in Dumay et al. Other factors that could have caused discrepancy in the results of the two studies are the number of words participants learned (24 in Dumay et al. compared to 6 or 12 in Leach and Samuel), and overall learning success in the recall task, which was less than 50% in Dumay et al.

All the studies reviewed in this section included the learning of new words with meaning. They used paired-association (word or sentence + picture) or definitions as part of the training session. Two of these studies (Breitenstein & Knecht, 2002; and McCandliss et al., 1997) showed that participants could successfully learn the meaning of new words. However, they only indicated that a link between a novel word and a concept had been created, without providing evidence on whether the words were successfully integrated with existing words in the mental lexicon. In the other 3 studies (Breitenstein et al., 2007; Magnuson et al., 2003; and Leach & Samuel, 2007), words were also successfully mapped to their meanings, but additionally, they could interact with previously existing words in the mental lexicon by showing evidence of priming (Breintenstein et al.), lexical competition (Magnuson et al.), and perceptual learning effects (Leach & Samuel, 2007). This distinction in word learning was first established by Leach and Samuel (2007) who proposed a theoretical differentiation between lexical *configuration* and lexical *engagement.* They described lexical configuration as the knowledge associated with a word such as its phonology, or semantics (as in Breitenstein & Knecht, 2002; and McCandliss et al., 1997) whereas lexical engagement is the process of dynamic interactions between a word and other words in the mental lexicon, as shown in studies of lexical competition (e.g., Gaskell & Dumay, 2003; Bowers et al., 2005) or priming (e.g., Breitenstein et al., 2007).

1.3.3 Learning new words from context

As shown in preceding sections, a large number of studies looking at word learning have emerged recently. These studies have provided very valuable evidence on different cognitive processes such as the learning of phonological word-forms, the mapping of word-forms to meaning, the process of lexicalization, integration of new material, and sleep-dependent memory consolidation for words. However, an aspect that has not yet been put into consideration is whether the methodologies used in teaching new words reflect real word learning. Phonological word-form learning studies have primarily used phoneme monitoring tasks to train participants (e.g., Gaskell & Dumay, 2003). Phoneme monitoring involves the conscious act of detecting target phonemes appearing anywhere in a novel word. Certainly, this is not a task that we would naturally perform to learn new vocabulary in a language, so the effectiveness of this methodology in simulating real word learning situations can be questionable. In studies in which participants have been required to learn the meaning of novel words (e.g., Magnuson et al., 2003), most researchers have used paired-association. As explained earlier, this methodology has produced excellent results and participants show long-term retention. However, the methodology might only be effective for teaching superficial features of concrete objects (e.g., shape, colour, size, etc), but might fail to target more elaborate features (e.g., functionality, significance, etc).

Phoneme monitoring and paired-association are both a rather poor representation of how we actually learn new words in real life. In order to acquire a more complete phonological and semantic representation of novel words, it might be necessary to experience words in a more natural environment. A methodology that has long been used primarily in educational research is *contextual learning*. This approach seems to offer many advantages over other approaches because it reflects more closely how words are learned in real life.

1.3.3.1 The emergence of the learning-from-context approach

The main purpose of language is to allow individuals to communicate in the best possible way. As we engage in conversations with other people, we exchange ideas and thoughts that contain an incredibly elaborate combination of words aimed at conveying a precise meaning of what it is on our minds at a particular moment in time. Apart from exchanging ideas and thoughts, from time to time we accidentally come across words or phrases produced by our interlocutor which sound unfamiliar to us. However, as the conversation unfolds, the meaning of those initially unfamiliar words might become more familiar. This occurs because relevant meaningful information is likely to become available during the conversation so we are able to infer the meaning of the previously presented unknown words. A similar process occurs while reading a newspaper, a scientific paper, or a textbook. Surprisingly enough, we do not normally reflect on this remarkable phenomenon, which does not seem to occur so often and with such spectacular results. In fact, most words that enter our mental lexicon are acquired incidentally during conversations with other individuals or by reading texts of all sorts. It is then very relevant to consider this type of learning of key importance for the study of cognitive processes of word learning.

Systematic investigation into how people, especially young children, extract the meaning of unknown words from linguistic context has been progressing for at least 30 years. One of the first serious attempts to validate the contextual learning approach corresponds to an early publication entitled *Teaching vocabulary-building skills: A contextual approach* by Sternberg, Powell, and Kaye (1983). In their book, they emphasised the benefits of the approach in teaching vocabulary and outlined the development of a number of learning skills to enhance vocabulary acquisition from context.

In a classroom setting, the contextual learning approach consists of having students attempt to infer the meaning of a new word from the contextual cues provided by sentences in which the word is used (Dempster, 1987). Discourse context has been considered a fundamental source for learning the meaning of novel words. In fact, there is substantial evidence that primary-school students learn most new words via incidental learning while reading texts (Jenkins et al., 1984; Nagy & Anderson, 1984; Nagy, Anderson & Herman, 1987). A good example of this method is the study conducted by Nagy et al. (1987), which presented children of third, fifth, and seventh grade with 2 narrative passages in which unknown words were introduced. Children were only told to read the passages but were not informed about the purpose of the investigation. One week later, they were given a surprise multiple-choice test in order to assess the knowledge of the target words embedded in the passages. Researchers found that children of all grades showed small but reliable gains in vocabulary knowledge. It is worth noting that this was one of the first studies to demonstrate that learning from context takes place during reading in a completely incidental way (without the reader being aware of the process). Previous studies had shown that children were able to derive meaning from context, but when words were presented highlighted in the text, so they were aware of which words they had to learn (e.g., Sternberg & Powell, 1983). Evidence from Nagy et al.'s work also supports the idea that the mechanisms involved in language are highly specialised allowing for rapid acquisition of words' meaning from context. This is consistent with previous research which found that approximately 5% to 12% of the 3,000 words students learn every year are learned after only a single exposure (Nagy & Anderson, 1984).

It is important to note that although individuals can quickly learn new words encountered in a linguistic context, the representations of these words can only be strengthened with additional exposures to the novel word in different contexts (Jenkins et al., 1984; van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001). Following the same argument, Bridge (1986) further suggested that word learning is facilitated if words are encountered in different contexts and if the contexts provide additional meaning. This idea finds support from another study conducted by Durkin (1990) which assessed the effects of frequency of exposure on the incidental learning of unknown words in context. Participants were a group of fifth grade students who were asked to read passages that contained 60 unknown target words. Half of the words were presented four times during training whereas the other half only once. Results showed that children performed better in multiple-choice and definition tasks when words were provided with four contextual exposures than just one exposure.

1.3.3.2 The encoding variability hypothesis

Ever since contextual learning started to be investigated, researchers have tried to elucidate the nature of context itself. Certainly, contexts can vary quite extensively depending on factors such as length, meaning, or syntax. For instance, it is not clear how extensive linguistic contexts can be since they might involve a single sentence, a paragraph or an entire text. As a response to some of the questions that gathered interest over the years, some scientists put forward an idea that was known as the *encoding variability hypothesis* (e.g., Bower, 1972; Hintzman, 1974). This proposal states that the probability of recall varies directly with the number of retrieval routes that are provided when the stimuli are learned. The theory received much interest from different researchers. For instance, Martin (1972) listed a variety of phenomena that he argued could only be explained in principles of encoding variability. The theory also quickly gained interest from those looking for a more practical application such as Bjork (1979) who recommended the application of the encoding variability hypothesis to a variety of aspects within instruction and education. He argued that there was substantial evidence that college students benefited enormously when given the opportunity to encode an instructional objective in multiple contexts rather than in a fixed context.

In educational terms, contextual learning can vary depending on the methodology used to get students to derive the meaning of words from context. If participants in an investigation are not aware of the fact that they are learning the meaning of new words, they are then performing a type of contextual learning that it is *incidental* and differs from the idea of consciously deriving word meaning from context. Bolger et al. (2008) suggested that the latter involves an explicit task aiming to learn the meaning of a target word from context, whereas incidental learning occurs naturally without any explicit learning goal.

Other studies have looked at how people learn new words from a combination of both dictionary definitions and context (e.g., Nist & Olejnik, 1995; Fischer, 1994). Nist and Olejnik (1995) claim that students can learn more words and more profoundly if they are provided with dictionary definitions apart from the context - weak or strong - in which the words are found. This is in line with the idea promoted by Curtis (1987), which suggests that both definitional and contextual knowledge may be needed for complete understanding of a word. Given this account, it is of interest to ask whether direct instruction through dictionary definitions alone could be sufficient to acquire new words. McKeown, Beck, Omanson, and Pople (1985) compared the gains in vocabulary acquisition during training that involved only definitions and definitions plus sentence contexts. They demonstrated that training including only definitions of words resulted in poorer performance in comparison with training in which context and definitions were provided.

Even though most evidence has supported the validity of the encoding variability hypothesis, some studies have challenged this idea. For instance, Dempster (1987) found no differences in learning when participants were presented with novel words either in three sentence contexts or in a single definition. Hence, these results provided no evidence that the opportunity to establish multiple retrieval routes by means of contextual information is helpful to vocabulary learning. This is surprising since previous studies seemed to support the role of different contexts in the acquisition of new vocabulary. It is important to note that sentence contexts can vary quite extensively from one study to another, so it is difficult to generalise from one particular experiment. It is also possible that one single semantically rich sentence can activate different semantic routes while three semantically poor sentences activate only one route. Furthermore, definitions might provide an instance for inferring new information, which in turn might activate other semantic routes. In order to carefully assess the role of the encoding variability hypothesis, it might be necessary to control for the amount of semantic information conveyed in sentences and/or definitions. This can be done, for example, by counting the number of semantic features provided in each linguistic context. The semantic features approach will be addressed in following sections of this Chapter.

1.3.3.3 The context variability hypothesis

In previous studies of contextual learning, researchers have explored the idea that many contexts are better than one context (encoding variability hypothesis) because of the number of retrieval routes associated with each context. The theory seems to have gathered support throughout the years despite some inconsistencies, particularly outlined in the study by Dempster (1987). One aspect that is worth noting is that the studies above did not control for the number of times participants were presented with the target words in each condition. For instance, a word presented in a three-sentence context plus a definition received 4 exposures whereas a word presented in a one-sentence context plus a definition received only two exposures.

In a recent study looking at the influence of context and definitions on the learning of new word meanings, Bolger et al. (2008) provided new insights into the aspects surrounding sentence-context learning. The main difference between this study and previous studies is that Bolger et al. controlled for the number of exposures words received during training by manipulating the number of different contexts. In two experiments, these researchers presented participants with words embedded in either four different sentences or four identical sentences. Experiment 1 was a 2 X 2 within subjects design with sentence conditions having either definitions or no definitions. In Experiment 2, no definitions were included; instead a no-sentence condition was added. The main purpose of this study was to examine whether people were able to acquire more meaning when words were presented in a variety of contexts as opposed to one single context. Furthermore, Bolger et al. predicted that deeper meaning would arise with the summation of each unique context the words appeared in. On the contrary, words experienced in a single context – even if repeated the same number of times - would only acquire an incomplete superficial meaning. Additionally, if context variability could produce full abstract meaning, then participants would be more accurate at generating definitions for words they experienced in a variety of contexts. The first study showed that when definitions were included to the one-context and the four-context conditions, no significant differences were observed between the two in a meaning generation task. However, when definitions were taken out, there was a clear advantage for the four-context condition. Bolger et al. concluded that context variability affected accuracy of response and interacted with the presence of definitions. A second experiment replicated the results found in Experiment 1 in the meaning generation task, with the four-context condition showing an advantage over the one-context condition. Additionally, in a forced-choice sentence completion task, participants did not differ regarding accuracy, but were significantly faster to respond in the four-context condition.

Overall, the results of these two studies supported the predictions of the researchers. Bolger et al. (2008) hypothesised that seeing a word in a variety of contexts results in a more decontextualized or abstract knowledge of a word's meaning. These results can be explained by the model of Reichle and Perfetti (2003), which assumes instance-based word memories with resonance processes that activate these memories when a word is experienced in a new context. This model keeps track of incremental learning of word meaning from context. Most importantly, it assumes a single mechanism to explain both abstract and context-specific word knowledge. According to this model, each instance with a word will leave specific memory traces of the event, so when the same word is encountered again in a different discourse, previous memories will become activated and will affect the representation of the new memories. The process continues until words acquire abstract representations that can be retrieved without the help of context.

In summary, contextual learning has the advantage of presenting students or participants in an experiment with a form of learning that closely resembles real life situations. However, most studies have failed to adopt more sophisticated methodologies to assess the gain obtained by participants during training. Except for Bolger et al.'s (2008) study, almost all other studies have only used offline measures to assess performance (e.g., definition accuracy, or multiple choice), instead of using online measures such as reaction times (RTs) for lexical or semantic decision.

1.4 Learning new words in a second language (L2)

Learning words in a second language involves similar processes to those experienced when learning a first language (L1). As reviewed in earlier sections of this Chapter, these processes are far from being simple and include the learning of words' components such as phonology, orthography, and semantics, which are necessary for words to become integrated in the language system. Additionally, the learning of a second language requires reshaping existing L1 representations in order to create new L2 representations. This makes the L2 highly dependent on the L1, particularly at an early stage of learning.

Jiang (2004) proposed that L2 vocabulary acquisition has two dimensions or categories. The first dimension includes processes developed throughout the time-course of words entering the mental lexicon (retention, consolidation, and automatization). At this stage, L2 word knowledge is still very dependent on the mother tongue and aspects such as pronunciation, use of grammar, and semantics will resemble very closely those of the L1. The other dimension could be defined as a stage of further development in vocabulary acquisition. At this point, L2 words are enriched and refined as the learner gets more knowledgeable through his/her experience with the language. This knowledge can be reflected in shifts of pronunciation with the learning of new phonemes or new variations of phonemes, which approximates the characteristics of L1 speech. The use of grammar would improve to the point of getting rid of common mistakes regarding word order or use of tenses, and semantic development would expand to the incorporation of new boundaries in concepts or completely new concepts. Of the three factors in the second stage of development, the latter seems to be the most important in L2 word learning since the ultimate purpose of learning words in a second language is to be able to use those words for successful communication (Jiang, 2004).

1.4.1 Learning new words from context in L2

Consistent with the L1 literature reported earlier, a large number of studies have reported evidence for contextual learning in L2 (e.g. Saragi, Nation, & Meisler, 1978; Nagy et al., 1985). A number of other studies have compared contextual learning with other methods used in teaching the meaning of words in a foreign language such as dictionary definitions or translations. In a study conducted by Hulstijn, Hollander, and Greidanus (1996), Dutch participants were randomly assigned to 3 groups: translation (participants were presented with a translation of the target words); dictionary (participants could use a dictionary if they did not know the meaning), context alone (no extra information was provided). Participants were first asked to read a short story in French L2 where 16 target words were presented. Then they were tested on the recognition of all target words found in the story. Finally they were required to provide a definition for each of the target words based on the meaning they derived from the story plus the translation or the dictionary definition. In the recognition task, they were asked to circle the options 'yes' (the word appeared in the story) or 'no' (it did not). In the recall task, they were required to provide a full definition of each target word. Results in the recognition task showed that participants recognized the words equally well in all three conditions. Additionally, an effect of frequency was found across all three conditions, with better recognition after three presentations than after only one. The recall task showed both a frequency effect and a group effect. Words with three presentations received more accurate definitions than words with one presentation. The translation group showed better meaning recall than the dictionary and context groups, but differences between the dictionary group and the context group were not significant.

The results of the above study are consistent with other studies using recall tasks. Nation (1982) failed to find an advantage of learning in context over translation, and Pickering (1982) found only very weak evidence that context was at an advantage over translation. It is worth noting that Hulstijn et al. (1996) only proves that the addition of translation helps meaning recall, but not that a translation is better than seeing a word in context. Hulstijn et al. further concluded that recalling words encountered earlier in text is extremely difficult and participants tend to ignore words they do not know. This might be the reason why participants benefited from the inclusion of the word's translation since they had the chance to momentarily access the meaning of a word they did not know. Another aspect that is important to consider is the learners' proficiency in the L2. This factor has been suggested to play a key role in the process of deriving meaning from context since beginners tend to use fewer contextual cues than more advanced learners. Cohen and Aphek (1980) found that advanced learners performed significantly better than beginners when the tasks involved recalling from L2 contexts. However, no differences in performance were observed when an L1 translation was used. This finding suggests that context is more useful when the L2 speakers have reached high proficiency in the language and that beginners benefit more from translation than contextual cues.

1.4.2 Cognitive aspects of semantic development in L2

Some might believe that learning the meaning of words in a second language does not involve a lot of effort because L2 learners have already acquired a semantic system in their mother tongue. For instance, they know familiar concepts such as *dog*, *fish*, or *ball*, which will not differ substantially in the second language – a dog is a dog in any language. However, some concepts do not have a one-to-one correspondence in other languages. For example, the concepts of *afternoon* and *evening* in English correspond to simply *tarde* in Spanish, so a Spanish speaker learning English as a second language will have to learn that the concept of *tarde* in Spanish does not have only one referent but two in English. Other more subtle examples are concepts that have a one-to-one mapping in the L2, but the meaning is slightly different. For instance, the concepts of *dinner* in English and *cena* (its closest translation in Spanish) do not correspond exactly to the same thing. Both are the last meal of the day but the times in which they are served tend to differ, *dinner* is usually served much earlier than *cena*. Also it is clearly the main meal of the day whereas *cena*, is the last meal of the day, but it is not necessarily the most important. Further distinctions can even be made regarding the type of food normally associated with *dinner* and *cena*. These are examples of how words in a foreign language do not always have an exact equivalent in the first language, so often L2 speakers need to develop slightly different or totally new concepts in the second language (Jiang, 2004). Jiang also argues that most new word learning studies in L2 have neglected semantic development and have only focused on word retention without carefully looking at how meaning is acquired. Thus, there is a need for more studies of semantics in L2 in order to understand the psycholinguistic processes and mechanisms involved in its acquisition.

Different approaches regarding semantic development have emerged throughout the years. One of the first hypotheses of semantic development suggested that speakers learning a second language only needed to learn new word labels for existing concepts without the need for semantic learning (e.g., Ausubel, 1964). However, more recent research has proposed that new L2 words are first mapped to existing concepts, but as speakers become more proficient in the second language a remapping process takes place (Giacobbe, 1992; Ringbom, 1983). This process involves the development of new concepts that, in the majority of cases, share many semantic properties with previously existing concepts. This idea fits in perfectly well with De Groot's distributed feature model (De Groot, 1992; Kroll & De Groot, 1997). The model suggests that concepts are made up of different conceptual features that are not unique to one particular concept but are shared by different concepts. According to this view, the conceptual level is only one and the configuration of new concepts in semantic memory is achieved as a function of how fluent speakers become in the second language. Second language learners initially access meaning for L2 words via the L1 and only at a later stage can access L2 meaning directly. Direct links between the L2 lexical level and the conceptual level will initially be very weak, but as speakers become more fluent in the L2 language, connections become stronger and might even approximate those of the first language.

Contrary to the belief that concepts are reshaped as speakers of a second language become more fluent, the alternative view suggests that the restructuring process is very slow and most of the time incomplete (Jiang, 2002). This idea finds support in Levelt's (1989) model of lexical representations, which assumes words have two separate components of linguistic information: lemma (including meaning and syntax), and lexeme (morphology and form). In this model, most L2 words are initially mapped to their L1 translation, and not to their meaning alone as it is the case in De Groot's (1992) model. As speakers encounter new L2 words, lemma information from the L1 translations starts to get literally copied on to the L2 word.

Since the lexeme information of the L1 word is not required, it is gradually deactivated and a direct connexion between the L1 concept and the new L2 word starts to develop. At this stage, the L1 and the L2 word share the same concept and the activation of the L2 word is mediated by its L1 translation. Nevertheless, new encounters with the new word will eventually weaken the L1 and L2 mapping and will allow a direct connexion between the L2 word and the pre-existing concept. Jiang's (2000) model took this view a step further and he suggested that when L1 words acquire the meaning from their L1 translation, it becomes extremely difficult to alter the existing meaning or acquire new meaning for the label. This view opposes heavily that of the De Groot's (1992) model because it gives no place for form-meaning remapping or recombination of conceptual features.

In a new study, Jiang (2002) gathered more data that further supports his view. He investigated errors that L2 speakers make when there is only one lexical entry in the L1 for two lexical entries in the L2. In an online semantic judgment task, participants were asked to decide whether two L2 words were related in meaning. The word pairs in one set had identical translation in the participants' mother tongue, while the words in the other set had different translations but were still semantically related. The results of the study revealed that participants were faster at judging whether 2 English words were related in meaning when they shared the same translation in the first language. Jiang interpreted the results as compelling evidence for L1 involvement in lexical processing of the L2, suggesting that even advanced L2 speakers do not develop new conceptual meaning; instead they simply transfer the semantic representation of the L1 words to the L2 words. It is worth noting that L2 speakers in Jiang's study were native speakers of Chinese, which is a language that differs greatly from English. Thus, it might be that these findings can only apply to languages that are very distant from each other, but not to languages that share common linguistic and cultural roots.

1.5 The semantic representation of words

In addition to the field of psychology, semantics has always been a topic of interest for researchers in many areas such as philosophy (e.g., Wittgenstein, 1922), logic (e.g., Frege, 1879), or linguistics (e.g., Chierchia & McConnell-Ginet, 2000; Dowty, 1979). This widespread interest has produced significant work resulting in different theories that offer a diversity of views regarding semantic representation. Given a full account of these theories is beyond the scope of this Chapter. Thus, only a review of relevant theories and studies of single-word semantic representation are presented here.

1.5.1 Concepts and word meanings

Researchers in the area of semantics persistently speculate among a variety of topics such as how word meanings are related to conceptual structures, how the meaning of each word is represented, or how the meaning of different words are related to each other. Whereas there seem to be some clarity about the representation of word meaning and its relation with other words, the relation between conceptual and semantic structures is less clear. Vigliocco and Vinson (2007) noted that most cognitive scientists and neuroscientists assume that conceptual knowledge (the non-linguistic mental representation of objects, events, etc) and word meanings (semantics) represent the same thing or, at least, there is a one-to-one mapping between the two (e.g., Humphreys, Price, & Riddoch, 1999). This is further supported by the fact that in most cases when we process word meaning we seem to activate equivalent conceptual information (e.g., reading the word *dog* activates the semantic representation of the animal we call *dog*). More striking examples can be observed in a neuroimaging study in which participants were required to read or hear motion words (e.g., *jump*, *run*, etc.). Beside the expected activation found for words of any type, results also showed increased activation in motor areas (e.g., Tettamanti et al., 2005; Vigliocco et al., 2006). This reflects that word meaning and conceptual knowledge are tightly related since the activation of a semantic representation produced subsequent conceptual activation even beyond the linguistic domain of meaning. Despite the example above, there are other cases in which the two domains do not seem to match well. Vigliocco et al. suggests that this is simply because language speakers have a lot more concepts than words; so many concepts do not have a lexical representation. For instance, we might be familiar with a certain type of object or animal because we have seen it many times, however, its lexical representation could be completely unknown to us if we have not learned its name. Further support for this distinction between concepts and words come from the neuropsychological literature where patients have shown semantic impairment only in the linguistic domain (e.g., while performing a naming task), but not during non-verbal tasks (Cappa & Gordon, 1992). Taken together, these findings suggest that there is a distinction between concepts and words and that we normally have more concepts than words since concepts do not necessarily need to be lexicalized.

1.5.2 Representation of word meaning

Theories of semantics can be divided into two greatly distinct approaches. One proposes that a word's meaning is holistic or unitary and cannot be decomposed (e.g., Anderson & Bower, 1973; Collins & Loftus, 1975) whereas the other view proposes that meaning is decomposable and can be represented as features (Smith, Shoben, & Rips, 1974; Norman & Rumelhart, 1975). Vigliocco and Vinson (2007) further explain that the non-decompositional view holds the idea that conceptual structures are holistic concepts which have a correspondent word in any given language. In contrast, in the decompositional approach word meanings are regarded as having combinations of conceptual features, which are mapped onto lexical representations and can interact with other linguistic information such as phonology (Damasio, et al., 2004; Vigliocco, Vinson, Lewis, & Garrett, 2004).

1.5.2.1 Holistic approach

As explained earlier, in a holistic approach lexical concepts are equivalent to word meanings (e.g., Fodor, Garrett, Walker, & Parkes, 1980; Level, 1989; Roelofs, 1997). Thus, the mental representations of concepts such as animals, objects, or actions, which can be lexicalized in a given language, are represented as unitary, abstract concepts in the conceptual system of a language user. The approach proposes that some representations are innate for concepts that do not require combination of different features during learning. The fact that concepts are learned by association of different features does not imply that they can be decomposed. The features can be individualised during learning but as the knowledge of concepts evolve, individuals acquire a single non-decompositional unity, which does not require the activation of separate features during retrieval.

A weak point of this view is perhaps the lack of clarity in establishing links between sub-components of conceptual knowledge. A contradiction that it is worth pointing out is the fact that holistic lexical concepts assume a strict one-to-one mapping between concepts and lexical representations. This goes against the idea of innateness or universality since conceptual structures in different language do not necessarily have the same lexical representations.

1.5.2.2 Featural approach

The first attempt to propose a model of feature meaning representation came about with the Feature Comparison Model (Smith et al., 1974). The model stated that word meaning was achieved based on combinations of features. Models like this received a lot of criticism because of their failure to define features for all possible meanings (Fodor et al., 1980). Another problem these models faced was the impossibility to explain the hyponym/hyperonym problem, which states that if the features of a word can be decomposed, then speakers would very often produce the category name instead of a member of the category (e.g, *animal* instead of *dog*). This would occur because *animal* contains all the features of *dog,* so it could well be selected during speech when attempting to produce *dog*. This issue was later solved using a computational implementation where lateral inhibitory connexions at the level of lexical units permitted perfect production of subordinate and superordinates.

More recently, different featural approaches have emerged that differ in whether features correspond to abstract or concrete units. A particularly influential one came from the field of neuropsychology (Allport, 1985; Warrington & Shallice, 1984), and computational neuroscience (Farah & McClelland, 1991). Unlike other approaches, features here are essentially concrete and grounded in perception (e.g., vision, hearing) and action (e.g., control, execution), or a combination of both. Thus, they are primarily embodied in a concrete experience with the environment. Accordingly, concepts in different semantic dimensions can vary as a function of their features, while some concepts rely mainly in sensory-related features (e.g., *is loud*, *is orange*), others rely essentially in motor-related features (e.g., *run*, *swim*) (e.g., Barsalou, Simmons, Barbey, & Wilson, 2003; Vigliocco et al., 2004).

Authors from different perspectives have tried to gain insight into the representation of features. While some have focused on identifying primitive features that give rise to other features and represent the lowest level of meaning decomposition (e.g., Jackendoff, 1983), others have shown interest in those features which are more salient and have a direct connection with the concrete world (e.g., McRae & Seidenberg, 1997; McRae, Cree, Seidenberg, & McNorgan, 2005; Vigliocco et al., 2004). The latter view has been implemented in several connectionist models of semantic representations. The training of the connectionist network usually involves input that corresponds to features either provided by speakers or directly selected by the researcher. These features typically contain specific characteristics to be tested (e.g., *perceptual* versus *functional*). An example of this is the model by Farah and McClelland (1991). They trained a network on the representation of living and nonliving things based on the proportion of perceptual and functional features associated with each category - living things contain mostly perceptual features while nonliving things have primarily functional features. When the model was lesioned, they found a differential pattern of results for living and nonliving that depended on whether the lesion affected visual-perceptual (impaired for living) or functional features (impaired for nonliving). In other similar implementations, researchers have focused on the simulation of intercorrelated features (features that are shared between concepts) versus distinguishing features (those which make the entity unique) (e.g., Devlin, Gonnerman, Anderson, & Seidenberg, 1998; Rogers & McClelland, 2004). More recent models of semantic representations have used semantic features directly provided by speakers in feature-norm studies (e.g., McRae et al., 2005; Vinson & Vigliocco, 2008). Further information on semantic features will be discussed in follow-up sections of this Chapter.

1.5.3 Semantics and word processing

Semantics can be classified as one of the main components of words together with phonology and orthography. As reviewed in previous sections, there are two main theories of semantics which include holistic and featural views. These two views differ widely since the first proposes that the meaning of words is non-decompositional and that there is a one-to-one mapping between concepts and lexical representations (e.g., Fodor et al., 1980; Levelt, 1989), whereas the latter assumes that meaning is made up of a combination of conceptual features (Vigliocco et al., 2004) and there is not necessarily a one-to-one mapping between concepts and word meanings. Since the featural approach offers more flexibility, it is more suitable than the holistic approach for studies investigating the relation between semantics and word processing. In the featural view, semantics can be easily quantified or assessed because the meaning of a word is mediated by the number or type of features, which have been combined through experience with a language. This leads to the conclusion that words can have different semantic representations depending on the experience of that word in a certain language.

One line of research has explored the concept of *semantic richness,* which emerged in response to the idea that words can have rich or poor semantic representations and that this difference seems to affect lexical processing. The way in which richness is determined can vary widely since there is not only one methodology to quantify semantics. Hence, researchers throughout the years have come up with different methods for this purpose and they have produced a sizeable number of variables [e.g., ambiguity, immageability, number of semantic neighbours (NSN), number of semantic associates (NSA), number of meanings, number of semantic features (NSF), etc.]. By manipulating these variables, scientists can get an insight into a word's semantic richness and, at the same time, assess whether or not semantics influences word processing. Behavioural tasks such as word naming, lexical decision, semantic decision, and semantic categorization have been used in order to look at the effects of semantics on word processing.

Variables that have produced relatively direct measures of semantic richness include the number of semantic features (e.g., Pexman, Holyk, & Monfils, 2003), the number of semantic associates (e.g., Buchanan, Westbury, & Burgess, 2001), and the number of semantic neighbours (e.g., Siakaluk, Buchanan, & Westbury, 2003). These variables have been widely used in studies with familiar words, where semantic richness effects have been reported for word naming (e.g., Pexman Lupker, & Hino, 2002), lexical decision (e.g., Pexman et al., 2002; Borowsky & Masson, 1996; Buchanan et al., 2001), and semantic categorization (Grondin, Lupker, & McRae, 2006; Pexman et al., 2003; Pexman et al., 2002).

Other variables that provide less direct measures of semantic richness include, for example, *semantic ambiguity*. Presumably, ambiguous words have richer semantic representations than unambiguous words because they have more than one meaning whereas unambiguous words have only one. Studies manipulating ambiguity have consistently found an advantage for polysemous over nonpolysemous words in lexical decision (e.g. Rubenstein, Garfield, & Millikan, 1970; Borowsky & Masson, 1996; Hino & Lupker, 1996); in naming for low-frequency words (e.g., Hino & Lupker, 1996) and irregular words (Rodd, 2004). It is worth noting that the ambiguity effect in naming is restricted to those situations in which the processing of a word is particularly difficult (e.g., Strain Patterson, & Seidenberg, 1995; Strain & Herdman, 1999). Another variable that has been widely studied is *imageability*. A number of studies have found imageability effects in naming irregular words (e.g. Cortese, Simpson, & Woolsey, 1997); and low-frequency words (Strain, et al., 1995), with high-imageability words producing faster responses than low-imageability words. Imageability effects have also been found using lexical decision (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Although many studies have shown that imageability effects are present in both lexical decision and reading aloud, there is evidence that suggests the effect is only consistent for lexical decision and might be confounded with other variables in naming. Morrison and Ellis (2000), in a regression analysis, found that imageability only had independent effects in lexical decision but not in naming, suggesting that the effects found in other studies might be due to other variables such as age of acquisition. In a bigger-scale study, Cortese and Khanna (2007) replicated the findings in the study by Morrison and Ellis (2000), and concluded that age of acquisition can predict naming and lexical decision above and beyond many other predictor variables, including imageability. Overall, effects of semantic variables such as ambiguity and imageability seem to be present in lexical decision. However, these effects are not very consistent for naming and are still part of a long-standing debate.

1.5.3.1 Number of semantic features (NSF)

As explained earlier, the featural approach has been very important in the understanding of semantic representations. One of the variables that emerged from this approach is the number of semantic features (NSF). This variable has been very useful in the understanding of semantic processing in numerous fields of research such as neuropsychology (e.g., Patterson, Nestor, & Rogers, 2007) computational modelling (Seidenberg, 2005), and behavioural psycholinguistics (Pexman et al., 2002), among others.

The number of semantic features for a given word is obtained by asking participants to list features or attributes that describe its meaning (see McRae et al., 2005; Vinson & Vigliocco, 2008). For instance, if participants are presented with the word *lion*, they will probably list features such as *has legs*, *is wild*, *lives in Africa*, *is dangerous*, etc. They normally list many features for some words (e.g., *eagle*), but few for others (e.g., *beetle*)[3]. This difference in the number of semantic features (NSF) provides an insight into words' semantic representations. For example, a word with 23 semantic features is assumed to have rich semantic representations whereas a word that has only 8 features is considered to have poor semantic representations. These differences in words' semantic representations have been used in a number of behavioural studies in order to understand the role of semantics in word processing. One of the first studies using NSF as a measure of semantic richness was conducted by Pexman et al. (2002). They wanted to test the notion of semantic feedback activation. This assumption is based on models of word recognition that assume bidirectional activation, so feedforward and feedback activation between orthographic, phonological, and semantic units is expected to occur during tasks such as reading and lexical decision. They predicted that high-NSF words would activate richer semantic representations than low-NSF words. Consequently, feedback activation from semantics to both phonological and orthographic representations would be greater, and consequently would produce better performance during naming and lexical decision. The results showed that the NSF effect was present in both tasks, which confirmed the researchers' predictions. Pexman et al. further argued that NSF effects were different from other semantic variables such as imageability and polysemy because all words used in the study were concrete and matched for imageability, and they only had one meaning. They concluded that NSF provided further support for the feedback activation account and semantic involvement in word recognition. Since Pexman et al.'s work, many other studies have provided concurrent evidence of the NSF effect on naming and lexical decision (e.g., Pexman et al., 2003; Grondin et al., 2009). Additionally, some of these studies

[3] A full account of the procedure to collect and record semantic features is provided in Chapter 6.

have extended the range of this effect including other tasks such as semantic categorization. A good example of the above is an investigation conducted by Pexman et al. (2003) whose main aim was to examine NSF effects to test the hypothesis that word meaning representations are distributed. This view comes from models of word recognition that assume a distributed representation of meaning as patterns of activation across meaning units (e.g, Borowsky & Manson, 1996; Seidenberg & McClelland, 1989). Unlike classical models, in which a word is represented by a single lexical unit and a single semantic unit (e.g., Fodor et al., 1980), in distributed models, meaning units do not correspond to specific concepts. Instead, they correspond to some kind of semantic features, as suggested in the model implemented by Kawamoto (1993), and in that of McRae, de Sa, and Seidenberg's (1997) where semantic features were collected empirically from participants. Going back to Pexman et al.'s research, they presented participants with target words embedded in sentences that had either congruent, moderately congruent, or incongruent contexts. They found a NSF effects in reading and semantic categorization when words were presented in incongruent contexts. They explained that the effects were not found for congruent contexts because the context activated semantic features of the target words before participants attempted to read them, which produced facilitation in both conditions (high and low NSF) washing out the NSF effect. The NSF effect here fits well with the distributed processing account since semantic representations varied in richness even in the case of concrete unambiguous words, which proves they are not holistic, but are likely to be made up of different features.

1.6 Conclusion and thesis overview

A few things can be concluded from the above sections of this Chapter. First, word learning as a paradigm seems to provide an excellent opportunity to investigate different aspects of language. Second, a few studies of adult word learning in L1 have provided substantial evidence for the existence of different processes and mechanisms involved in the acquisition of new vocabulary. These include the process of lexicalization and the effects of overnight sleep in the consolidation of newly learned words. However, most of these studies have neglected the role of semantics in the process of learning and integration. Third, L1 and L2 contextual word learning studies have focused primarily on the learning of word meaning, but most of them have failed to use appropriate methodology to carefully evaluate cognitive aspects of word learning. Fourth, semantics, as the main component of words, has been widely investigated and different theories of semantic representations have emerged. One that has been proved very useful in the understanding of semantic development is the featural approach because of the flexibility that offers to manipulate semantics experimentally.

Given the above, there is a lack of word learning studies investigating semantics and its effects on the process of learning and integration.

The following chapters of this thesis focus on word learning and semantic development and include eight behavioural experiments and one final combined behavioural and fMRI experiment. Chapter 2 presents two word learning experiments in English as a second language aiming at assessing the role of context variability and feature variability on the processing of newly learned words. A number of tasks will be used to assess the learning of meaning and its effects on reading aloud and visual word recognition. Chapter 3 aims at comparing L1 and L2 speakers regarding the acquisition and integration of new lexicon. Effects of semantics on visual word recognition and reading will be assessed. Chapter 4 explores effects of semantic richness on visual word recognition and looks at the time course of lexical integration. Chapter 5 also explores effects of semantic richness on visual word recognition but includes some changes in methodology to more carefully assess consolidation over time. Chapter 6 first presents an experiment aimed at collecting semantic features for 100 familiar words, followed by a combined behavioural and fMRI experiment looking at the neural correlates of familiarity and semantic richness. Finally, Chapter 7 presents a general discussion of the thesis and future directions in the investigation of word learning and semantics.

Chapter 2 – Learning words in a second language: effects of context and feature variability

2.1 Introduction

Chapter 2 explores the learning of new vocabulary in English as a second language. Experiment 1 examines how context variability (the number of different sentence contexts in which words are presented) affects word learning. As stated in Chapter 1, it is well-known that people can easily extract the meaning of unknown words while reading texts, either in a first or a second language (e.g., Jenkins et al., 1984; Nagy et al., 1987; Hirsh & Nation, 1992; Lawson & Hogben, 1996). It has also been reported that when contexts are semantically rich or meaningful, the process of acquiring novel words is largely facilitated (e.g., Bridge, 1986). Similarly, a number of experiments have shown that learning improves if novel vocabulary is presented in multiple contexts (e.g., Dempster 1987; Bjork, 1979; Bolger et al., 2008). Overall, the evidence suggesting an advantage for learning new words in multiple contexts seems to be very consistent. However, previous studies of word learning in L2 have mainly used offline measures (e.g., number of correct definitions, or number of words recalled), which do not take into account online processing times.

Experiment 1 analyzes accuracy and RTs for two new tasks: semantic decision and reading aloud. Experiment 2 investigates feature variability, which is defined here as the type and number of semantic features conveyed in sentence contexts. In studies of real word processing, semantic variables such as the number of semantic features or the number of semantic associates have been found to influence visual word recognition (e.g., Pexman et al., 2008; Pexman et al., 2003). More specifically, words that contain many semantic features or many semantic associates are generally processed faster in tasks such as reading aloud, lexical decision, and semantic categorization. Since effects of semantic variables have mainly been tested using familiar words, in Experiment 2 the aim was to look at the effect of semantic richness on newly learned words.

In summary, Experiment 1 examines context variability effects on semantic decision and reading aloud whereas Experiment 2 looks at feature variability effects on semantic decision, cued recall, reading aloud, and recognition memory.

2.2 Experiment 1

Broadly speaking, context can be defined as the linguistic environment in which a given word is presented. This linguistic environment can be very diverse including different semantic areas such as Arts, Science, History, Economics, Philosophy, etc. Thus, learning the meaning of a new word involves experiencing that word in the different contexts in which it is used. As presented earlier, a few studies have shown that the number of contexts in which a word appears - context variability - benefits the acquisition of new words (e.g., Jenkins et al., 1984; van Daalen-Kapteijns at al., 2001). As reviewed in Chapter 1, this effect can be explained in terms of instance-based models (e.g., Reichle & Perfetti, 2003), which assume that each experience with a novel word represents a specific memory trace and includes focal information (phonology and meaning) and contextual information (the physical setting in which the word is found).

An extension of the above view is found in a study conducted by Bolger et al. (2008), which proposed the context variability hypothesis. This hypothesis states that contexts, which vary enhance word learning more than contexts that do not vary. Bolger et al.'s framework combines the assumption of instance-based word memories with the idea of resonance mechanisms (Myers & O'Brien, 1998). The latter assumes that a nonselective memory for words can be reactivated when related words are read in a text. In contextual word learning, this means that reading words in a text produces the activation of related words in the speaker's mental lexicon, so when a new word is encountered its meaning is learned via association of familiar words in the text and activated words in the mental lexicon. Bolger et al. have argued that by combining the instance-based account and the resonance mechanism theory, it is possible to explain how both abstract and context-specific knowledge increase during learning. The new framework postulates that the encounter of a new word creates a specific memory of the event and when the word is encountered again in a new context, a more abstract meaning starts to develop as the previously stored representation(s) affect the processing of new instances.

The context variability hypothesis can be broken down in a series of assumptions: (i) Each instance with a new word in a different context represents the storage of a new episodic event. (ii) Each instance with a new word in the same context does not create a new memory trace but it is likely to strengthen an existing one. (iii) New episodic events are affected by previously stored events since they resonate or get reactivated by the context. (iv) Encountering a word in many contexts represents the combination of different instances and different episodic memory traces, which produce more decontextualised abstract learning. (v) Encountering a word many times but in a single context creates fewer but stronger memory traces, which produces a more rigid context-specific type of learning.

As reviewed in Chapter 1, Bolder et al. (2008) tested the context variability hypothesis in two experiments that involved learning new words in 1-sentence contexts and 4-sentence contexts. Across the two experiments, they found that words presented in 4 different sentences led to more accurate meaning generation and faster RTs in sentence completion. They concluded that exposure to variable contexts produced better learning of abstract meanings.

The following study investigates the effects of context variation on newly learned words in English L2 using two new tasks: semantic decision and reading aloud. Advanced L2 English speakers, whose native language was Spanish, learned new English words in 2 or 12 different sentential contexts over the course of 3 days. On day 1 and day 2, they received training on the novel lexicon, and on day 3 they performed the reading and the semantic decision tasks. Two main predictions were made based on the context variability hypothesis (Bolger et al., 2008). If the variation of contexts leads to better learning of meaning, higher accuracy and faster RTs should be expected in the semantic decision task. The second prediction comprises the role of semantics in reading aloud. As reviewed in Chapter 1, semantics has been found to affect reading aloud, but its effects are most commonly found for low-frequency words (e.g., Hino & Lupker, 1996) or irregular words (e.g., Rodd, 2004) because the reading of these words may involve semantic representations more than the reading of high-frequency or regular words (Harm & Seidenberg, 2004). If the above is true and context variation generates better learning of meanings, an advantage in reading times should be expected for L2 words learned in 12 contexts. Like low-frequency and irregular words, L2 words are particularly hard to read, so if there is an effect of semantics in reading, this might emerge during L2 word reading.

2.2.1 Method

Participants

Twenty-one L2 speakers (12 females; mean age 27.3 years, SD 5.5) whose native language was Spanish participated in the experiment. All individuals were postgraduate students at the University of York with no record of any language disabilities and with normal or corrected-to-normal vision. They were proficient in English as they had studied English as a second language for an average of eight years prior to arrival in the United Kingdom, and had spent at least a year in an English-speaking country before the experiment took place. In addition, they had all met the University of York minimum English language proficiency level for applicants whose native language is not English. In order to meet that requirement, they were asked to take any of the following tests and obtain a score above the minimum level, which is 6.0 in the IELTS Test (IELTS Partners, 2009-2010), 213 in the computer-based TOEFL Test (Educational Testing Service, 2011), A/B/C grades in the CPE Test (ESOL Examinations, 2011), and grade A in the CAE Test (ESOL Examinations, 2011).

Materials and design

Thirty-six obscure English nouns were used in the study. In order to select these items, a group of 15 English native speakers was asked to underline all unfamiliar words (either in meaning or orthographic form) from a list of 100 obscure English words preselected from different sources such as dictionaries, websites, and books. Participants in the selection of words were all university students or staff members and did not take part in the word learning experiment. Forty-four words were classified as unfamiliar by all 15 participants, of which 36 were finally used in the experiment. The words were divided into three sets of 12: set A (e.g., *abutment*, *barratry*), set B (e.g., *agterskot*, *baldachin*), and set C (e.g., *abaiser*, *bourdon*) and each set was assigned to a different condition (see Appendix 2.1 for full list). At the end of the study, participants were also asked to indicate whether they were familiar with any of the words prior to the experiment. If so, responses to that word were deleted from the analysis.

Throughout the selection of the items used in the experiment, words starting with voiceless fricative consonants (e.g., /f/, /s/) were avoided, as they are not always quickly detected by the microphone and voice key during reading aloud tasks. Furthermore, all words in each set had a different initial letter, with the only exception of /b/ which was used twice in each set. Items were between 6 and 12 letters long (mean length: 8.3) and sets were matched on average word length.

There were two tasks to be performed in the study: word naming and semantic decision. Because there were three conditions in the naming task (high context variability, low context variability, and no training), all three sets of words were used. However, in the semantic decision task, there were only two conditions, so only the two sets of words that received training were included in the test. Additionally, 72 filler words were used during the testing session (see Appendix 2.2). Each of these items was either related or unrelated in meaning with one of the obscure words presented in the training session. The filler words had Standard Frequency Indexes (SFI) between 41 and 67.6, according to The Educator's Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995). A paired samples t-test was run on the frequency scores of the filler items showing no significant differences in frequency between the related and unrelated word lists, $t_1(36) = .53$, $p = .28$. In addition, each pair of filler words was matched on initial letter and length.

During the training session, twelve sentences were used as linguistic contexts to present each of the 36 unfamiliar words. The sentences were mainly taken from a wide variety of documents available on the Internet and later adapted so that all of them fell between 10 and 20 words long. Below are some sample sentences used in Experiment 1. More sample sentences per condition can be found in Appendix 2.3.

The failure of the ABUTMENT may mean the collapse of the bridge.

Persons convicted of BARRATRY shall be barred from practice of law.

Before pouring the water over the king's hands, the EWERER was supposed to taste it.

As stated earlier, the three sets of items were distributed across three conditions: high context variability (HCV), low context variability (LCV) and no training (NT). Words in HCV were embedded in 12 different sentence contexts while in LCV, only 2 different sentence contexts were used. In the NT condition words did not receive any training and were only presented to the participants during the word naming task (test procedure). In order to rotate the sets of words across conditions, participants were divided into 3 groups and each group received a slightly different training session. Thus, group 1 was given set A in HCV, set B in LCV, and set C in NT, while group 2 received set A in LCV, set B in NCD, and set C in HCV. Finally participants in group 3 were presented with set A in NT, set B in HCV, and set C in LCV. In the LCV condition, where words were presented in only 2 linguistic contexts, the 12 sentences available were rotated across participants (e.g. participant 1 in group 1 got sentences 1 and 2 from set B, while participant 2 in group 1 was presented with sentences 3 and 4 from set B, and so on).

Procedure

The experiment took place over 3 consecutive days. On day 1, participants had a training session which lasted approximately 40 minutes. On day 2, they repeated the same training procedure from day 1. Finally, on day 3 they performed a word naming and a semantic decision task.

Training procedure

On day 1, prior to the first training session, participants were required to sign a consent form which contained a brief description of the study.

The training session started with some practice trials, so participants would familiarise with the task. As shown in Figure 1.1, each novel word was first presented in isolation for 2 seconds. Novel words were displayed in capital letters, and in 44 point black Arial font on white background. Participants were instructed to read each novel word aloud and as accurately as possible. Each novel word was then presented in 12 sentences. Sentences were displayed one by one and remained visible for 8 seconds each. The novel word embedded in the sentences was then displayed again in capital letters, but this time in 28 point black Arial font. The rest of the words in each sentence were presented in lower case, but with same size and font as the novel word. Participants were required to read the sentences silently, and try to infer the meaning of the novel word (in capital letters). The same procedure was repeated for each item. In order to avoid any fatigue that might have caused distraction or lack of attention, participants were allowed to take 3 breaks during the training session. The conditions in which novel words were presented differed in the number of different sentence contexts they were embedded in. Thus, in the HCV condition, novel words appeared in a different sentence every time they were displayed in context. However, in LCV, novel items occurred in only two different sentences, each of which was repeated six times, so the frequency was kept constant in both trained conditions. On day 2, the training procedure was exactly the same as that on Day 1. The order of the conditions was counterbalanced across participants and sessions, so half of the participants saw HCV before LCV on day 1 and a reverse order of the conditions on day 2. The other half saw LCV before HCV on the first day and LCV followed by HCV on the last day of training.

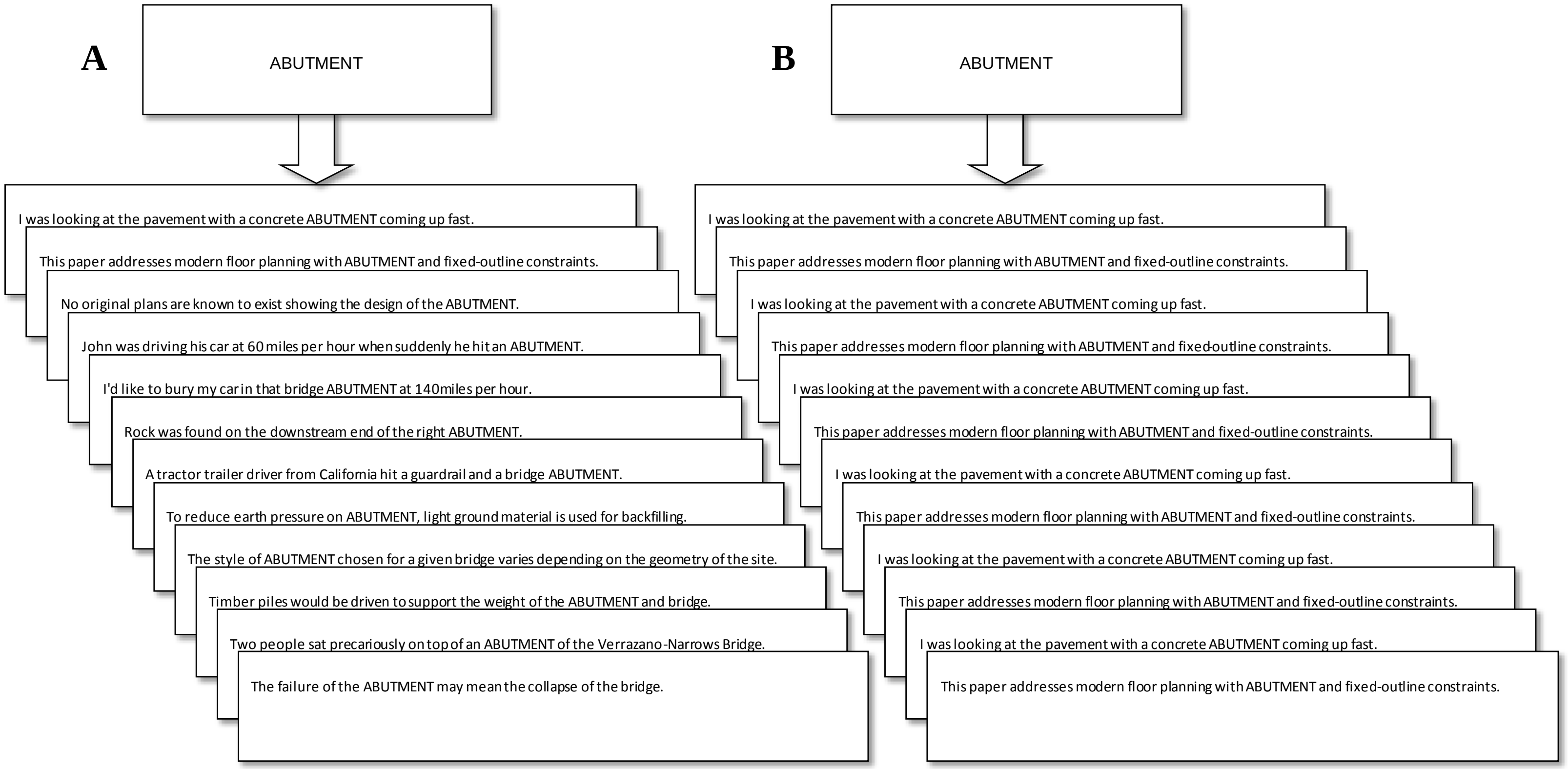


Figure 2.1. Structure of training procedure in Experiment 1. The target word *abutment* is presented first in isolation for 2 seconds and then in 12 sentence contexts every 8 seconds. Figure A represents training in high context variability and Figure B in low context variability.

Testing procedure

The experiment was run on a PC computer using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). The stimuli were presented in random order with lower case 18 point black Currier New font on white background. In the word naming task, a central fixation cross was displayed on the screen for 1 second, followed by the newly learned word for 3 seconds or until participants responded. The screen then remained blank for 2 seconds until the next trial began. Individuals were instructed to read the items aloud as quickly and accurately as possible. Twenty practice trials, which included low and high-frequency words, preceded the experiment, so participants could get used to the task. A microphone connected to a voice key was used to collect onset reading latencies. Additionally, a tape recorder was used to record each trial in order to assess participants' pronunciation and make sure that the voice key was not activated by something other than the word.

After completing the naming task, participants were instructed to perform the semantic decision task, which only included the two trained conditions. The task started with the presentation of a fixation cross for 2 seconds, followed by the newly learned word for 2 seconds. Then two other words were displayed - one to the right and the other to the left of the computer screen - and stayed on until participants made a response. They were required to press "1" if the item on the left was semantically related to the target or "2" if the item on the right was related to the target. A response box especially designed for the E-Prime software was used to collect response latencies to each word.

At the end of the experiment, participants were presented with the entire list of obscure words in written form, and they were asked to underline the words they were familiar with prior to the experiment.

2.2.2 Results

The data analysis was performed on all 21 participants by means of repeated-measures ANOVAs and Bonferroni-corrected paired samples t-tests. Analyses were run with participants (F_1) and items (F_2) as the random variables. When sphericity was not assumed, the reported p-values associated with the F statistics were adjusted via Greenhouse-Geisser. Effect sizes were obtained by means of Partial Eta Squared (η_p^2).

Three items were removed from the list (one from each set) due to pronunciation errors, which reached 30% or more in the word naming task. The final analyses were therefore carried out with 33 items in both tasks. Mean reaction times (RTs) and error rates in the naming and semantic decision tasks are presented in Table 2.1.

Table 2.1. Mean naming latencies and percent error rates in the naming and semantic decision tasks.

	Experimental conditions		
	HCV	LCV	NT
	Naming		
Mean RT	749	781	843
SD	154	170	206
% errors	8.6	9.5	15.4
	Semantic decision		
Mean RT	1763	2017	
SD	897	1148	
% errors	10.8	20.9	

Note: HCV, high context variability; LCV, low context variability; NT, no training.

Word naming

The total number of responses collected in this task was 693, of which 74 were treated as errors (10.7%). Six (0.9%) corresponded to voice key trigger errors, 20 (2.9%) were the result of naming latencies either below 300 milliseconds or 2.5 standard deviations above the mean (by-participants), 41 (5.6%) were removed due to mispronunciations, and 7 (1.0%) because they corresponded to items that participants reported to be familiar with prior to training on the novel vocabulary.

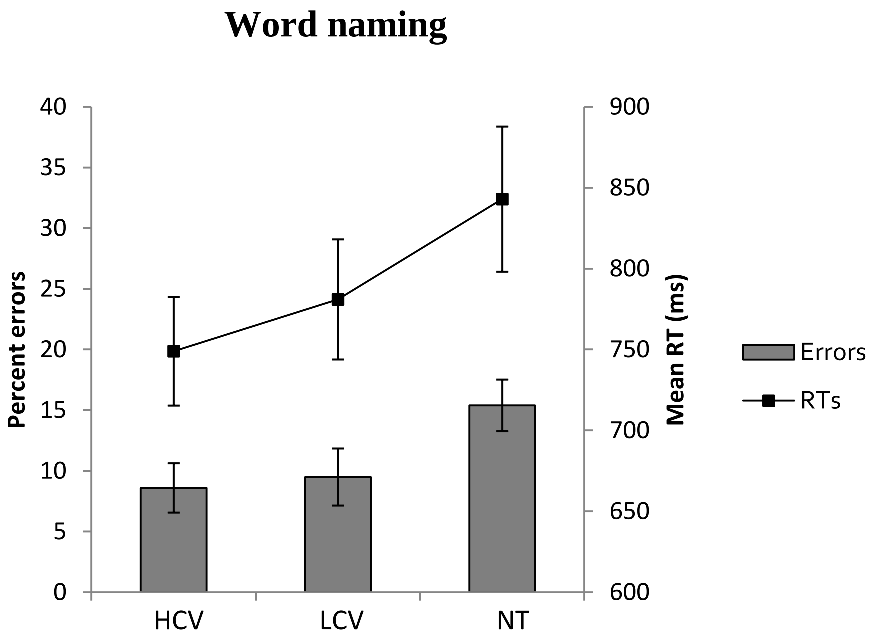


Figure 2.2. Percent errors and reaction times (ms) for high context variability (HCV), low context variability (LCV), and no training (NT) in the word naming task. Error bars represent standard error (SE) of the mean.

RTs

A one-way repeated-measures ANOVA was conducted on naming RTs for correct responses revealing a significant main effect of conditions, $F_1(1, 20) = 16.94$, MSE = 3769.98, $p < .001$, $\eta_p^2 = .46$; $F_2(1, 32) = 12.03$, MSE = 6388.87, $p < .001$, $\eta_p^2 = .27$. Bonferroni-corrected paired samples t-tests ($\alpha = .05$) showed a significant difference between all comparisons with faster RTs for HCV versus LCV; HCV versus NT, and LCV versus NT.

Errors

A one-way repeated-measures ANOVA was also conducted on error rates. Results showed no main effect of conditions in the by-participants analysis, $F_1(1, 20) = 2.74$, MSE = 73.61, $p = .08$, $\eta_p^2 = .12$. However, a main effect of conditions was found in the by-items analysis, $F_2(1, 32) = 3.22$, MSE = 102.71, $p = .05$, $\eta_p^2 = .09$. Inspection of Figure 2.2 suggests a difference between the trained conditions (HCV, LCV) and NT, but Bonferroni-corrected paired samples t-tests ($\alpha = .05$) yielded no significant differences in any of the comparisons: HCV versus LCV, HCV versus NT, LCV versus NT.

Semantic decision

Only the trained conditions (HCV and LCV) were included in this task. A total of 462 responses were collected, but 74 (16.0%) were excluded from the analysis. Fifty-six (75.7%) of the excluded responses corresponded to errors, 12 (16.2%) were the result of RTs above 2.5 standard deviations from the mean (by-participants), and 6 (8.1%) corresponded to words participants reported to be familiar with.

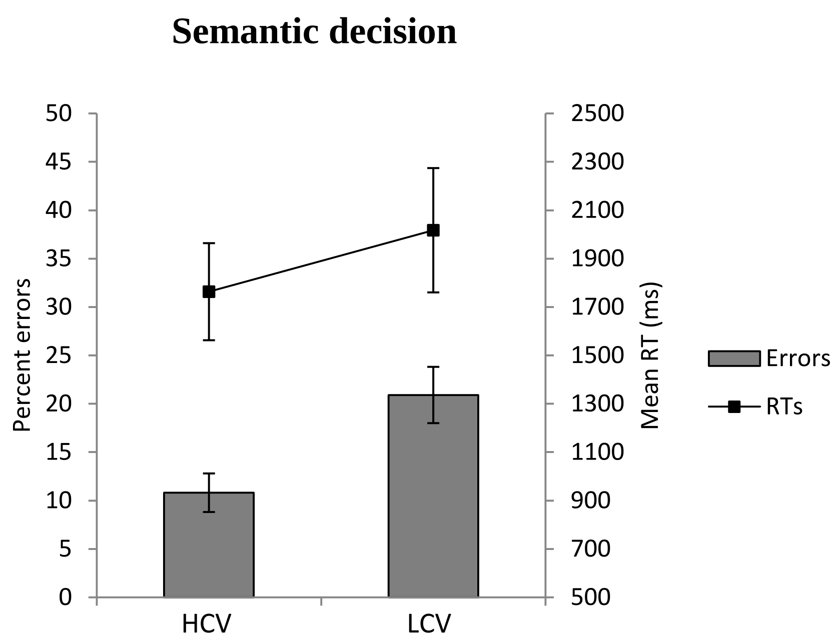


Figure 2.3. Percent errors and reaction times (ms) for high context variability (HCV) and low context variability (LCV) in the semantic decision task. Error bars represent standard error (SE) of the mean.

RTs

Paired-samples t-tests were run on semantic decision latencies to correct responses. Results showed a significant difference between HCV and LCV in the by-subjects analysis, $t_1(20) = 2.374$, $p = .03$, but no significant difference in the by-items analysis, $t_2(32) = 1.014$, $p = .32$.

Errors

T-tests were also run on error rates revealing a significant difference between HCV and LCV in both by-subjects and by-items analyses, $t_1(20) = 2.435$, $p = .02$; $t_2(32) = 2.401$, $p = .01$.

2.2.3 Discussion

Experiment 1 investigated to what degree context variability affects novel word learning in English as a second language. Context variability effects on reading aloud and semantic decision were assessed using error rates and RTs for words learned in high context variability (presented in 12 different sentences), and low context variability (encountered in 2 different sentences).

The first prediction stated that there would be higher accuracy and faster RTs in the semantic decision task for words in high context variability (HCV) compared to words in low context variability (LCV). In line with predictions, results showed the expected advantage for words learned in HCV, with clearer effects in the error analyses than in the RT analyses (not significant by-items). Taken together, these results support the idea that context variation can improve the acquisition of new word meaning (Cain, 2007; Bolger et al., 2008). This effect had previously been found using meaning generation and sentence completion tasks, particularly in the study conducted by Bolger et al. (2008). However, it had never been demonstrated using semantic decision. The results of this experiment fit well with the instance-based resonance framework of incremental word learning based on the Reichle and Perfetti (2003) model. The hypothesis states that experiencing a word in multiple contexts results in a more decontextualized and abstract representation of the word's meaning. This occurs because each encounter with a novel word in a different context creates a new episodic trace, so the more contexts a word appears in, the more abstract its meaning becomes. The results here showed that participants learned well in both conditions with around 80% accuracy in LCV and around 90% in HCV, which suggests that very few contexts are sufficient to acquire the meaning of novel words, which is in agreement with previous evidence (e.g., Jenkins et al., 1984). However, the fact that accuracy was higher in HCV and, particularly, that latencies were faster represents a more complete and decontextualized acquisition of word meaning, which supports the context variability hypothesis. This is also consistent with the results of the sentence completion task in Bolger et al.'s study, even though they only found the effect of context variability on RTs. They concluded that completing a sentence similar to those found during training, a single context might be sufficient to achieve good accuracy; however, more contexts provide stronger cues, which allow faster retrieval of word meaning. This in turn accelerates the comparison of the target word's meaning and the information provided by the test sentence.

The second prediction came as a consequence of the first one, so it stated that if the meaning of words is learned better in many sentence contexts, then it is possible that this would be reflected in reading speed given that semantics has been found to influence reading (e.g., Hino & Lupker, 1996; Pexman et al., 2002). The results of the naming task did confirm the prediction since faster RTs were found for HCV than for LCV. This is also in line with the results of an orthographic choice task in the study conducted by Bolger et al. (2008). They demonstrated that context variation produced gains in orthographic knowledge, as measured by the ability of participants to select the right spelling for each newly learned word. This finding suggests that orthographic learning improves as a consequence of better learning of meaning. This is not surprising since studies in children have also linked word recognition ability with vocabulary knowledge (e.g., Nation & Snowling, 2004: Ouellette, 2006).

The idea that semantics can or cannot speed up reading times is part of a long-standing debate. A study conducted by McKay, Davis, Savage, and Castles (2008) tested this hypothesis using a word learning paradigm. They found that novel words trained with meaning were read faster than novel words trained without any meaning. Importantly, this effect was only found for newly learned words with inconsistent spelling. They argued that the reason for the semantic effect to emerge only for newly learned words with inconsistent spelling was because semantics plays a more predominant role when the process of reading is particularly hard. Since the current study was conducted in English as a second language, the process of reading newly learned words was also challenging, and somehow comparable to the process of reading newly learned words with irregular pronunciation in a native language. This fits in well with models of reading aloud like the triangle model by Seidenberg and McClelland (1989), which specifies greater semantic input for more difficult words, such as those that have low frequency or inconsistent spelling. Even though the model does not include reading words in a second language, the idea can be generalised to L2 reading.

An alternative explanation for the results can be found in studies of corpora, which have reported that reading is enhanced for words occurring in many semantic contexts (McDonald & Shillcock, 2001; Adelman et al., 2006). The explanation they provide is that high context variability items are more likely to occur in a new context, so they are more available during recognition, which would speed up the process of recognition (e.g., Anderson & Milson, 1989; Anderson & Schooler, 1991). This explanation gives less importance to the effect of meaning and, instead, it emphasizes the simple statistical probability of a word to occur.

A possible confound to consider in this study is *attention* since novel words experienced in many contexts might demand more involvement from individuals than novel words presented in only 2 different contexts. Additionally, reading a different sentence each time requires deeper semantic analysis than reading the same sentence over and over again. Due to shallow processing in the latter situation, participants' attention is more likely to divert from the main task causing poor learning. This is supported by memory studies suggesting that deep semantic processing requires substantial attention (Craik & Byrd, 1982) and that diversion of attention to a secondary task can result in superficial encoding of information processed in the primary task (Naveh-Benjamin, Craik, Gavrilescu, & Anderson, 2000).

In summary, this experiment has shown that context variability, measured in number of different sentences, can affect both semantic decision and reading aloud for newly learned words in English as a second language. The study demonstrated that novel words experienced in many contexts acquire a more complete and abstract meaning, which supports the context variability hypothesis. Additionally, context variability reflects gains in the semantic representation of newly learned words, which seems to affect the reading speed of L2 vocabulary. It is important to note, however, that part of the context variability effect observed in this experiment might have been due to differential attention in each condition.

2.3 Experiment 2

Experiment 1 looked at context variability effects on word reading and semantic decision. An advantage for words learned in high context variability (HCV) over words in low context variability (LCV) was found in both semantic decision and word naming. Since conditions differed in the number of contexts (12 versus 2), it might be possible that the advantage for words learned in 12 different sentences was not only due to context variability but also due to attention. As explained earlier, the fact that participants were required to read the same sentences six times in LCV might have caused disengagement from the task and, as a consequence, lack of learning. On the contrary, when participants read 12 different sentences (HCV) they were probably more alert since every sentence conveyed new information, which made the task more interesting and demanding. In order to overcome this problem and make conditions equally demanding, a change in the methodology was required. Thus, Experiment 2 moved away from the idea of context variability and instead manipulated feature variability.

Here I propose the feature variability hypothesis, which states that people learn better if they are presented with core semantic features and contextual features, instead of only contextual features. This idea comes from studies of contextual word learning that have shown that people learn better when they are exposed to dictionary definitions and contexts than contexts or definitions alone (Beck, McKeown, & Omanson, 1987; Fischer, 1994; Nist & Olejnik, 1995; Bolger et al., 2008). The reason is that dictionary definitions convey core semantic features, which allow direct access to meaning (Bolger et al., 2008), but since they only represent one episode or instance, they lack contextual features, which are necessary to learn how to use words in a language (McKeown et al., 1985). As explained in Chapter 1, people can easily learn words from context but this knowledge has been suggested to be only partial and mainly reflects situational properties of word meaning (Durso & Shore, 1991; Shore & Durso, 1990) as opposed to more abstract, decontextualized knowledge of a word's core meaning (Goerss, Beck, & McKeown, 1999) that can be provided in dictionary-style definitions. As reviewed in Experiment 1, sentence contexts can also provide decontextualized abstract knowledge but this comes with experiencing a word in different contexts, which suggests that the process is much slower than if core features are directly provided via a definition (van Daalen-Kapteijns., 2001).

The evidence above suggests that a combination of dictionary definitions and contexts is best to learn new words. It is worth noting that according to the instance-based framework reviewed in Experiment 1 (Myers & O'Brien, 1998), dictionary definitions can also be considered contexts, but of a different nature due to the type of features they provide. A well-constructed definition has the potential to reactivate many contextual memory traces during learning because its features contain more abstract representations. This leads to thinking that the only difference between a single sentence context and a definition is that the latter conveys both contextual features and core features allowing more direct access to meaning. Thus, if we manipulate the type of features provided in sentence contexts, we can allow more or less direct access to meaning depending on whether or not we include dictionary-style core features.

The following experiment manipulated *feature variability*, the type and number of semantic features that participants were exposed to in sentence contexts during training on a new vocabulary. There were three semantic conditions: Rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics (RIS). In RCS, eight sentences were carefully constructed so they conveyed core features and together simulated a full definition of the novel words in addition to the contextual information (see Appendix 2.5). In PCS, sentences were also carefully created as to avoid or minimise the presence of core features, so they mainly conveyed contextual information. For instance, participants could infer that something had *a colour* but could not infer which colour, or that something was *useful* without specifying its use (see Appendix 2.5). The third condition (RIS) was used as a control condition and presented words in semantically inconsistent sentences so participants could not build a coherent meaning of the word.

Like Experiment 1, the current experiment used word naming and semantic decision. However, it included two new tasks: recognition memory and cued recall. Recognition memory has been widely used in previous word learning studies to assess explicit word knowledge such as phonology (e.g., Gaskell & Dumay, 2003; Davis et al., 2008) orthography (e.g., Nation, Angell, & Castles, 2007), and semantics (e.g., McKay et al., 2008). Cued recall and other similar recall tasks have primarily been used in semantic word learning studies where participants are asked to associate novel words with objects (e.g., Gupta 2003; Rueckl & Dror, 1994). The cued recall task in this experiment involved the presentation of orthographic cues along with the definition of the target word.

Predictions in Experiment 2 were made for each of the tasks above. First, participants are expected to learn words better when they are presented with core features in addition to contextual features (RCS) in comparison to contextual features alone (PCS). If direct exposure to core features reflects faster meaning acquisition, effects should be reflected in higher accuracy and faster RTs in semantic decision. A similar result is expected in the cued recall task because the acquisition of more meaning in RCS should produce a more accurate match between the orthographic cues, the definition and the target word. Second, predictions in the naming task are less clear since the role of semantics in reading aloud is debatable and might be confounded with other variables such as age of acquisition (see Monaghan & Ellis, 2002; Cortese & Khanna, 2007, for discussion). If, indeed, semantics affects reading as suggested in other studies (e.g., Pexman et al., 2002; Hino & Lupker), faster RTs should be expected in RCS compared to PCS and RIS. It is worth noting, however, that these studies did not control for age of acquisition. Third, if semantics plays a role in recognition memory as it has been demonstrated in previous word learning studies (e.g., McKay et al., 2008), and in lexical decision studies with real words (e.g., Pexman et al., 2003); better performance should be expected in RCS compared to PCS and RIS in the recognition memory task.

2.3.1 Method

Participants

Twenty-seven advanced L2 English speakers (18 females; mean age 22.8 years, *SD* 1.9) from the University of Concepcion, Chile participated in the study. All individuals had normal or corrected-to-normal vision, and had not been diagnosed with any language disorders.

They were all bilinguals who had Spanish as L1 and English as L2. They had all studied English in a scholastic setting for 8 years on average prior to enrolment at the university. All the students were in their fourth (last) undergraduate year, so they had completed at least 1,000 hours of instruction in English. ESL students at the University of Concepción are required to take several English language courses throughout the four years: Pre-intermediate, Intermediate, Upper-intermediate, and Advanced. These courses include the use of oral and written language, and a significant input of grammar and phonetics. Additionally, they attend English Literature, American and British History, Applied Linguistics, Translation (English-Spanish), apart from optional courses, such as Academic Writing, Short-story Writing, Drama, etc. In addition to the above academic courses, ESL students get a high input of formal and informal English through music, the internet, films, television, etc. Thus, by the end of their fourth year, all students have achieved an advanced level of English.

Materials and design

Thirty obscure English words from Experiment 1 were used (see Appendix 2.1). However, in order to have more control over linguistic variables such as word length and initial letter; the original orthographic forms for these concepts were replaced with nonwords (e.g., abutment → *abrutmon;* baldachin → *beelchan;* nankeen → *nanpheen*). An additional set of nonwords was added because of the extra condition in Experiment 2, so in total there were four sets of 10 items each (see Appendix 2.4). The sets were rotated across conditions to counterbalance the stimuli. As in Experiment 1, nonwords beginning with voiceless fricatives were avoided as onsets of these words are not always detected by the voice key in reading aloud tasks. All items in each set had a different initial letter and they were all eight letters long. Eight different sentences were used as linguistic contexts in each trained condition (e.g., *The ABRUTMON is shown in the original plans*). See Appendix 2.5 for full list. The majority of the sentences were based on real sentences found in a wide variety of documents available on the Internet and in which the source words were used. The sentences were later adapted so they conveyed contextual and core semantic features or only contextual features depending on the condition. The sentences were also limited to a minimum of 5 words and a maximum of 13.

Memory recognition filler items consisted of 40 nonwords that matched the nonwords learned in the training session in initial letter and length (e.g., *arneless, bireleny, neactlon*; see Appendix 2.6). Filler stimuli for the semantic decision task included 80 high-frequency real words. Forty of these words were related and 40 were unrelated in meaning with the corresponding novel words (e.g., *support* =

abrutmon; substance ≠ abrutmon; see Appendix 2.7). All the fillers had Standard Frequency Indexes (SFI) between 41 and 67.6, according to The Educator's Word Frequency Guide (Zeno et al., 1995). A paired-samples t-test found no differences in frequency between the related and unrelated lists of words, t_1 (39) = .46, p = .65. The four sets of items selected were distributed across four conditions: rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS), and minimal training (MT). Words in RCS were presented in 8 consistent sentences that conveyed core semantic features of the target concept. The idea was to show the novel words in sentences that convey features that are essential for full understanding of concepts. These included precise location (e.g., *over the altar*); specific function (e.g., *something you sit on*); exact material (e.g., *made of bronze*), etc. In PCS, items occurred in 8 consistent sentences that contained minimal semantic information, so participants could only infer unspecified information, which did not allow full access to meaning. This included unspecified location (e.g., *in most countries*); unknown material (e.g., *made of material*); general property (e.g., *has a colour*), etc. The RIS condition presented participants with inconsistent meaning with each sentence introducing attributes corresponding to a different concept. See Table 2.2 for sample sentences in each condition. Finally, MT words were only presented once in spoken and written modalities, but without any context.

Table 2.2. Sample set of sentences corresponding to rich consistent semantics (RCS), rich inconsistent semantics (PCS), and rich inconsistent semantics (RIS).

Rich consistent semantics (RCS)
An ABRUTMON anchors the cables of the bridge.
Two people sat precariously on top of an ABRUTMON.
John hit an ABRUTMON while driving his car.

Poor consistent semantics (PCS)
Peter walked passed the ABRUTMON.
They were not able to see the ABRUTMON.
The ABRUTMON is shown in the picture.

Rich inconsistent semantics (RIS)
I tried to portray the world in all its ABRUTMON and ugliness.
An ABRUTMON can have the water go from his land without obstruction.
Intravenous ABRUTMON caused an increase in salt gland secretion.

Participants were divided into 4 groups in order to rotate the sets of words and sentences across conditions. Then, a slightly different training session was designed for each group. Group 1 was given set A in RCS, set B in PCS, set C in

RIS, and set D in MT; group 2 was assigned set D in RCS, set A in PCS, set B in RIS, and set C in MT; group 3 was presented with set C in RCS, set D in PCS, set A in RIS, set B in MT; and finally group 4 was exposed to set B in RCS, set C in PCS, set D in RIS, and set A in MT.

Procedure

The experiment took place over 3 consecutive days with 1 session each day. On day 1, participants completed a training session that consisted of 2 parts and lasted approximately 30 minutes. On day 2, they were only presented with part 2 of the training again. Finally, on day 3 they had to perform 4 language tasks: word naming, recognition memory, semantic decision, and cued recall.

Training procedure

Prior to the first training session, participants were required to sign a consent form which contained a brief description of the study. Then they were instructed to sit in front of a computer to begin the first training session. In part 1, participants were presented with novel words in isolation in both written and spoken modalities. The words in written form were presented in upper-case, 32-point black Arial font on white background. Headphones were used to present the novel words in spoken modality. Participants were required to say each word aloud after hearing and simultaneously seeing it on the computer screen. In part 2, the same font and background used in part 1 were employed for the presentation of the words in isolation. However, when the words occurred in context, they were presented in upper-case, but in 22-point Arial font. The rest of the words in the sentences had lower-case letters, but were of the same size as the target word. See Figure 2.4 for structure of the training session.

One of the four sets of words was only presented once in part 1. This set corresponded to the minimal training condition (MT). Only the three remaining sets were included in part 2. They made up the three conditions with meaning and extensive training: Rich consistent semantics, poor consistent semantics, and rich inconsistent semantics (see Appendix 2.5). In each of these conditions, novel words appeared first in isolation for 2 seconds and then embedded in four different sentences that were displayed for 5 seconds one after the other. Then the novel words would appear in isolation again followed by four more sentences. The conditions were presented in separate blocks and the order of presentation for each block was counterbalanced, so that the same number of participants was exposed to each condition in first, second or third position. Participants were required to read the new words aloud when they appeared alone, but when they were shown as part of a sentence, participants were instructed to read the entire sentence silently and try to infer the meaning of the unfamiliar word. On day 2, participants repeated the same training session for part 2 in the previous day.

In all three sentence-context conditions, participants received 21 exposures to the novel words, so the frequency remained constant throughout the experiment. However, in the minimal training condition, participants were exposed to words only once in spoken and written form and without any context.

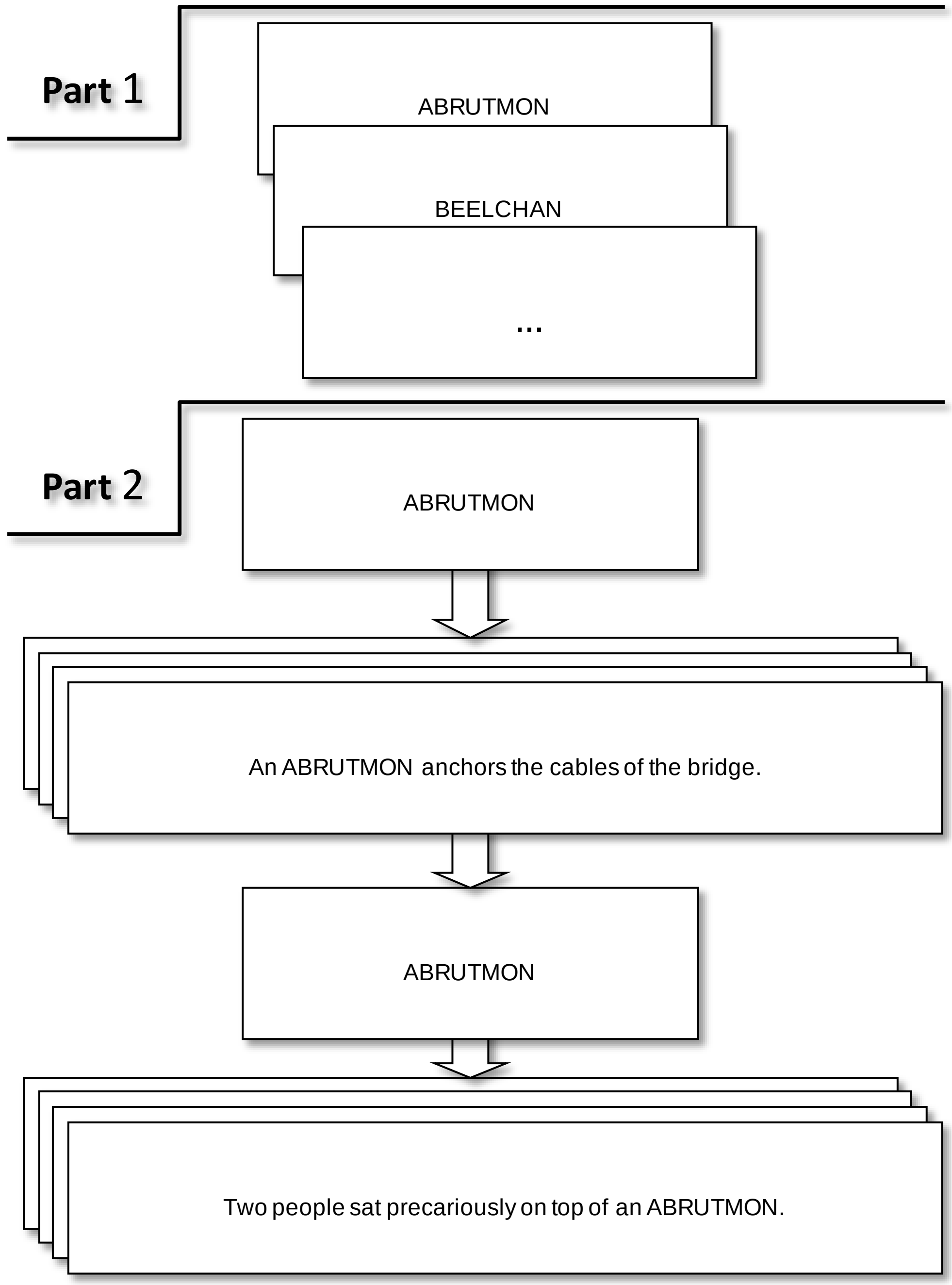


Figure 2.4. Structure of training procedure in Experiment 2. Each word in part 1 is presented for 2 seconds. In part 2, words in isolation are also presented for 2 seconds, but each sentence is on for 5 seconds.

Testing procedure

The experiment was run on a PC computer using E-Prime software (Schneider et al., 2002). Stimuli were presented visually in 18-point black Courier New font on white background. The naming and recognition memory tasks included four sets of words (four conditions), whereas the two semantic tasks (semantic decision and production) included only the three sets of words presented in context (three conditions). The order of the tasks was counterbalanced across participants, but the recognition tasks always occurred before the semantic tasks.

Word naming

Each word was preceded by a fixation cross (+) for 1 second, and then the newly learned word appeared and remained on the computer screen for 2 seconds or until participants made a response. The items were randomised and presented in different order to each participant. Individuals were instructed to read the items aloud as quickly and accurately as possible. Participants completed 10 practice trials before the actual experiment, including low-frequency words. A tape recorder was used to record all responses and naming latencies were collected by means of a microphone connected to a voice key.

Recognition Memory

Trials started with the presentation of a fixation cross (+) for 1 second and then a newly learned word was displayed on the screen for 3 seconds or until participants made a response. Trials ended with a blank screen presented for 2 seconds. Items were showed in a different random order to each participant, and responses were made by pressing either “1” or “2” on an E-Prime response box. Participants were told to press “1” if the word that appeared on the screen was a word they had seen during the training session, and to press “2” if the word on the screen was a word they had never seen before. Response latencies were measured from the onset of the stimulus presentation to the onset of the button press.

Semantic decision

The semantic decision task consisted of the presentation of a fixation cross for 1 second, followed by a blank screen for 500 milliseconds. A newly learned word was then presented for 2 seconds, followed by a familiar target word for 3 seconds or until participants made a response. The target was either semantically related or unrelated with the newly learned word presented earlier. Semantic decisions were made by pressing either the left button (for stimulus and target semantically related) or right button (for stimulus and target semantically unrelated). The rest of the procedure was exactly the same as the one used in the recognition memory task.

Cued recall

Participants were presented with the definition of a newly learned word for 5 seconds. They then saw a fixation cross for 1 second followed by the first 2 letters and the last letter of the target word (e.g., *ab_ _ _ _ _ n*). They were instructed to produce (orally) the complete word that matched the definition and the orthographic cues. Only accuracy of response was measured in this task.

2.3.2 Results

Three participants were removed from the analyses due to high error rates in at least one of the tasks. Two of them had less than 70% accuracy in reading aloud, and the other one had less than 65% accuracy in the recognition memory task. A summary of the results for word naming, recognition memory, semantic decision, and spoken word production for the remaining 24 participants is shown in Table 2.3. The data analyses were performed by means of repeated measures ANOVAs with participants (F_1) and items (F_2) as the random variables. Bonferroni-corrected paired samples t-tests were also used when specific comparisons between conditions were needed. When sphericity was violated, the reported *p*-values were adjusted using Greenhouse-Geisser correction. Partial Eta Squared (η_p^2) was used to report effect sizes.

Table 2.3. Mean latencies and percent error rates for word naming, recognition memory, and semantic decision. Percent words recalled in cued recall.

	RCS	PCS	RIS	MT
		Word naming		
Mean RT	701	687	716	756
SD	123	116	155	161
% error	12.5	11.3	10.0	19.2
		Recognition memory		
Mean RT	763	775	765	---
SD	176	199	186	---
% error	11.25	7.92	10.83	---
		Semantic decision ("Yes" responses)		
Mean RT	1045	1168	---	---
SD	227	239	---	---
% error	28.8	39.2	---	---
		Semantic decision ("No" responses)		
Mean RT	1231	1247	---	---
SD	284	262	---	---
% error	33.8	31.7	---	---
		Cued recall		
% words recalled	40.4	30.8	---	---
SD	18.8	16.1	---	---

Note. RCS, rich consistent semantics; PCS, poor consistent semantics; RIS, rich inconsistent semantics; MT, minimal training.

Word naming

The total number of responses collected in this task was 960, of which 126 (13.1%) were eliminated. Eighty-nine (9.3%) of the responses eliminated corresponded to mispronunciations, 20 (2.1%) corresponded to voice key errors, 10 (1%) to latencies over 2.5 standard deviations from the mean, and 5 (0.5%) were the result of RTs below 300 milliseconds. Finally, 2 (0.2%) other RTs were removed due to noise produced before the word onset, which resulted in the voice key being erroneously activated.

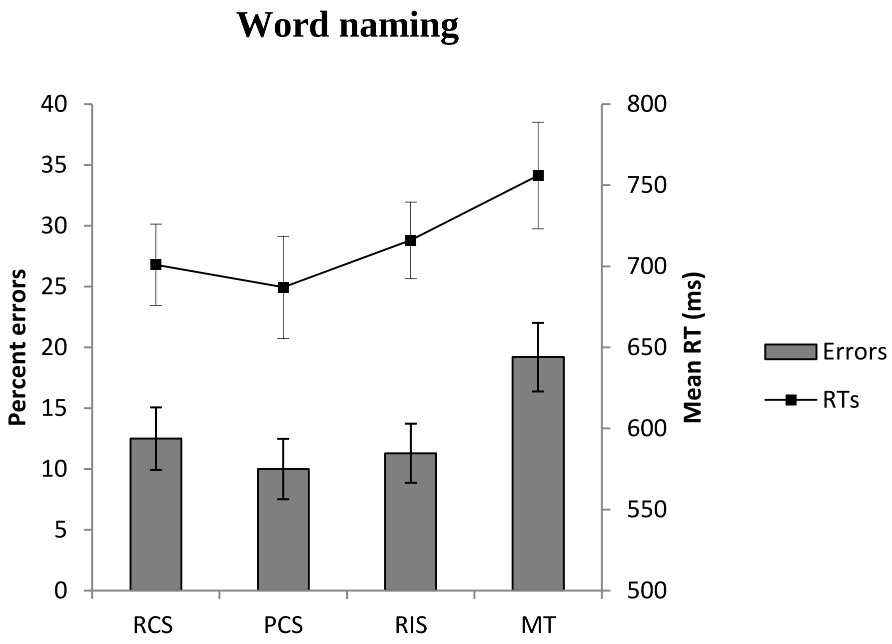


Figure 2.5. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS) and minimal training (MT) in the word naming task. Error bars represent standard error (SE) of the mean.

RTs

A one-way repeated-measures ANOVA was conducted on the data showing a main effect of conditions, $F_1(2, 23) = 9.50$, MSE = 2295.97, $p < .001$, $\eta_p^2 = .29$; $F_2(2, 39) = 8.56$, MSE = 5617.27, $p < .001$, $\eta_p^2 = .18$. Bonferroni corrected t-tests ($\alpha = .05$) showed no differences between RCS, RIS, and PCS, but all three conditions showed faster RTs than MT.

Errors

The one-way ANOVA on errors showed no main effect of conditions (by-participants), $F_1(2, 23) = 2.18$, MSE = 166.22, $p > .05$, $\eta_p^2 = .08$. However, the effect was present in the by-participants analysis, $F_2(2, 39) = 3.64$, MSE = 193.54, $p < .05$, $\eta_p^2 = .09$. Bonferroni corrected paired-samples t-tests ($\alpha = .05$) only showed a difference between PCS and MT with an advantage for PCS over MT. The other comparisons did not show any difference.

Recognition memory

Unlike the naming task, the analyses of the recognition memory task only included the results corresponding to the three conditions trained with meaning (RCS, RIS, and PCS). Words in MT were presented without meaning and only once during training, so participants were not expected to recognize them as newly learned words during the recognition memory test. As a result, the total number of responses collected was much lower than in the previous task reaching 720, of which 73 (10.1%) were removed from the analysis. These included 59 (8.2%) errors and 14 (1.9%) RTs situated 2.5 standard deviations above the mean.

RTs

A one-way within-subjects ANOVA was run on the data revealing no main effect of conditions, $F_1(2, 23) = .29$, MSE = 3460.16 $p = .75$, $\eta_p^2 = .01$; $F_2(2, 39) = .04$, MSE = 5764.13, $p = .97$, $\eta_p^2 = .00$.

Errors

The ANOVA conducted on errors also showed no main effect of conditions, $F_1(2, 23) = 1.20$, MSE = 66.12 $p = .31$, $\eta_p^2 = .05$; $F_2(2, 39) = 1.11$, MSE = 132.93, $p = .34$, $\eta_p^2 = .03$.

Recognition memory

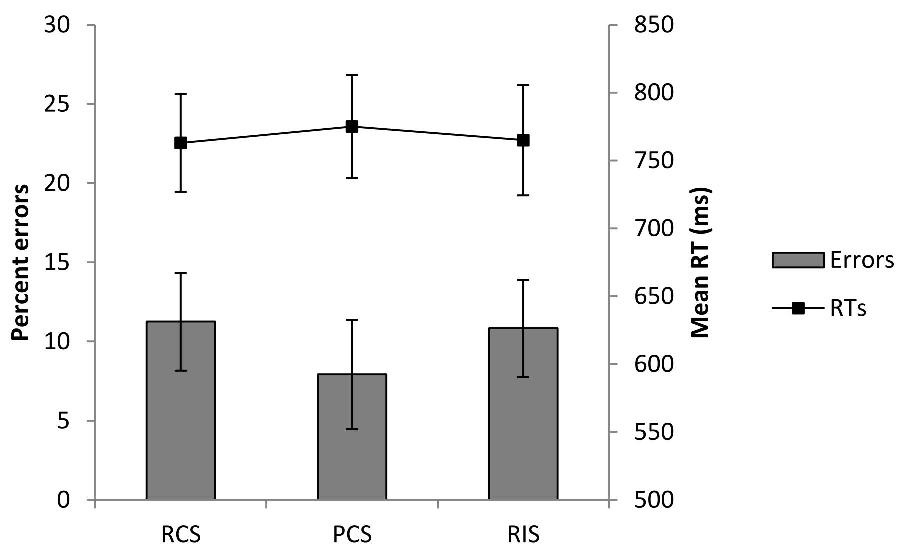


Figure 2.6. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics (RIS) in the recognition memory task. Error bars represent standard error (SE) of the mean.

Semantic decision

The semantic decision task produced two different sets of data. Set 1 contained responses for familiar words that were semantically related to the newly learned words (*YES* responses), and set 2 grouped responses to familiar words, which were unrelated in meaning to the newly learned words (*NO* responses). Hence, two separate analyses were conducted. See figure 2.7.

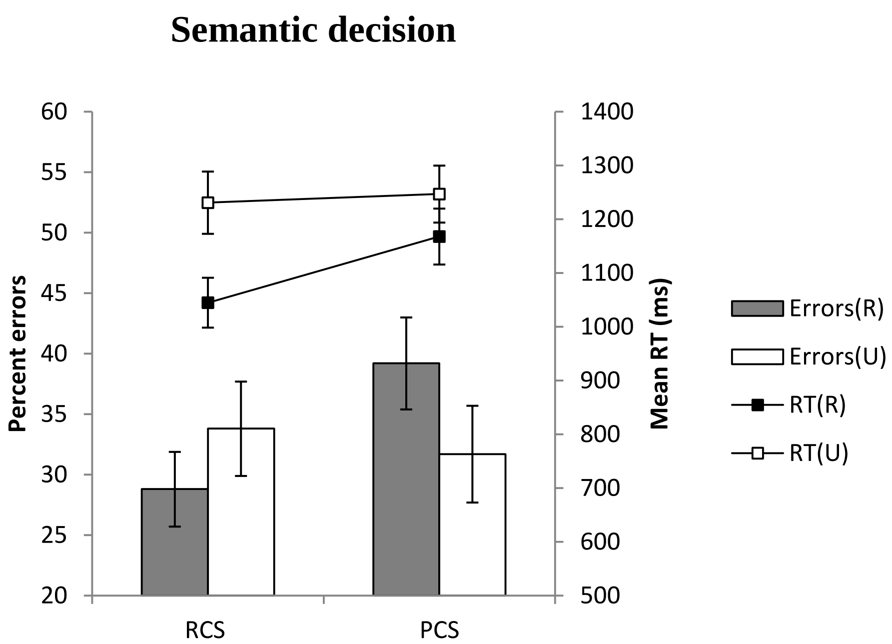


Figure 2.7. Semantic decision. Percent errors and RTs for semantically related (R) and unrelated (U) words in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars represent standard error (SE) of the mean.

Related (YES) responses

A total of 480 responses were collected, but 163 (34.0%) of these responses were removed from the analyses: 159 (33.1%) corresponded to errors and 4 (0.8%) to RTs over 2.5 standard deviations from the mean. Each condition was first compared to chance (50%) to find out whether participants had learned in both conditions. A Wilcoxon Signed Ranks Test (α = .05) was conducted on the data and showed that participants performed significantly better than chance in both conditions.

RTs

A paired-samples t-test was run on the data revealing a significant difference between RCS and PCS [$t_1(23) = 3.08$, $p < .01$; $t_2(39) = 2.00$, $p = .05$], with faster RTs in RCS.

Errors

The t-test conducted on errors also showed a significant difference between RCS and PCS, with increased number of errors in PCS, $t_1(23) = 2.57$, $p = .02$; $t_2(39) = 2.64$, $p = .01$.

Unrelated (NO) responses

The number of responses collected here was also 480. The total number of deleted responses reached 101 (24.8%), including 96 (20.0%) wrong answers, and 5 (1.0%) RTs over 2.5 standard deviations from the mean. The Wilcoxon Signed Ranks Test ($\alpha = .05$) here also showed that participants performed significantly better than chance in both conditions.

RTs

A paired samples t-test was first conducted on latencies showing no differences between the RCS and PCS regarding semantic decisions to unrelated word pairs, $t_1(23) = .662$, $p = .51$; $t_2(39) = 1.789$, $p = .08$.

Errors

The t-test conducted on errors showed the same results, $t_1(23) = 1.123$, $p = .27$; $t_2(39) = 1.031$, $p = .31$.

Cued recall

As in the semantic decision task, in the cued recall task there were only two conditions (RCS and PCS). Words were considered correct if they were fully elicited and deviated only one phoneme from their standard pronunciation. A group of 15 native speakers of English was previously asked to read the nonwords in order to establish their standard pronunciation. The proportion of correctly elicited newly learned words was analysed here.

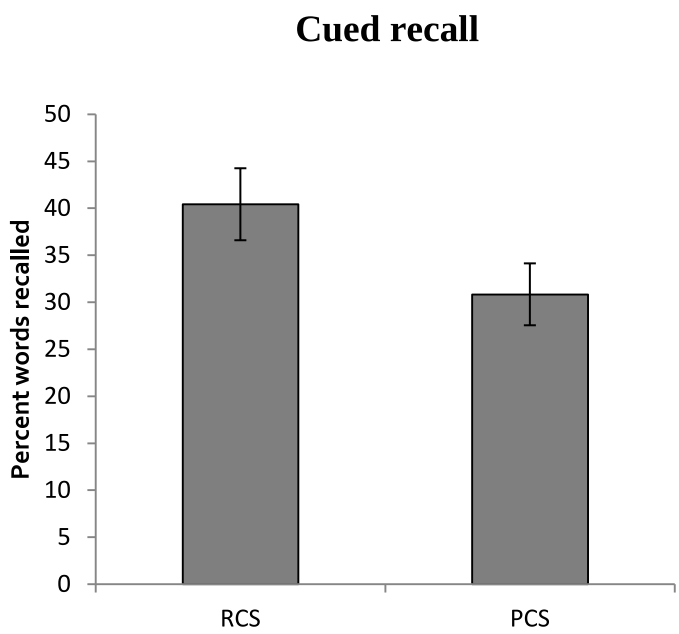


Figure 2.8. Percent words recalled in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars show standard error (SD) of the mean.

A paired-samples t-test was conducted on the data and showed a significant difference between the conditions [$t_1(23) = 2.673$, $p = .01$; $t_2(39) = 2.297$, $p = .03$], with higher accuracy in RCS than PCS.

2.3.3 Discussion

The first hypothesis predicted that participants would learn words better when presented with core semantic features during training and that improved performance would be reflected in semantic decision and cued recall.

First, the semantic decision task showed no difference between RCS and PCS conditions for negative semantic judgements (NO responses), which is not very informative and/or relevant because participants could have produced a "NO" response for any unknown pair of words as well as for those they knew were not semantically related. Regarding positive semantic judgements (YES responses), participants were more accurate and faster for words in RCS than for words learned in PCS, which is in line with the predictions. This suggests that the presentation of core features in sentence contexts can improve the learning of word meanings with respect to sentence contexts that do not explicitly present core semantic features. This supports studies which have shown that dictionary definitions plus sentence contexts favour the learning of words because participants can directly access meaning through definitions that provide core semantic features. Thus, the improvement in accuracy and RTs participants experienced in RCS must have been due to faster development of decontextualized meaning. If this is the case, then words in RCS acquired a richer meaning than words in PCS.

There are a number of previous studies using real words which have reported that semantic richness can affect semantic categorization, which is a similar task to the one used in this experiment. In both tasks participants have to activate the meaning of the target word in order to assess either its relatedness to another word (semantic decision), or the category to which the word belongs to (semantic categorization). For instance, a series of studies have reported number-of-features effects on semantic categorization (Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Grondin et al., 2006; Pexman et al., 2003; Pexman et al., 2002). In these studies, participants are normally presented with lists of words, which they have to classify into two categories (e.g., concrete or abstract). The semantic richness effect on semantic categorization measured in terms of number of semantic features seems very robust and it has been consistently found across different studies. Words that have many semantic features are processed faster than words that have fewer features, even when possible confounding variables are taken into account. The current experiment supports the above views and adds that semantic richness can also affect semantic decision for newly learned words in English L2.

Second, similar to the semantic decision task, the cued recall task showed better performance for the words that participants learned in RCS (associated with core features). In this task, participants were presented with the definition of a novel word and they had to elicit the novel word that corresponded to that definition. In cognitive terms, the definition represented a concept for which the participant had to find a phonological form in the mental lexicon. Since participants learned more detailed meanings for the novel words in RCS than in PCS, their recall was better for RCS words than for PCS words. The process of producing a novel word based on a definition is similar to naming a pictured object, so a parallel can be established between theories of speech production used in object naming and the results of the cued recall task in this study. In picture naming, the process first involves the recognition of the object; then semantics is activated and finally lexical access is achieved (e.g., Johnston, Dent, Humphreys, & Barry, 2010; Ellis, Kay, & Franklin, 1992; Snodgrass & McCullough, 1986). In the current study, participants were presented with a definition, which would be the equivalent of a picture in a picture naming task. By means of the definition, participants would activate the stored representation (concrete or abstract) of the newly learned concept, which would then generate the necessary semantic activation to access its phonological and orthographic form. Words in RCS acquired more meaning than in PCS, so their representation in the mental lexicon was stronger or more established compared with that of PCS words. A strong or rich semantic representation can produce an accurate match between the concept and its corresponding definition. On the contrary, if a semantic representation is weak, which was probably the case for PCS words, then the link between the concept's semantic representation and its definition is particularly hard to find.

Third, the last two predictions concerned the role of semantics in reading aloud and recognition memory. Regarding reading aloud, predictions were not completely straightforward since the role of semantics in reading is still a matter of debate. In the case of recognition memory, it was predicted that there would be an advantage for words in RCS in comparison with PCS and RIS.

In the word naming task, participants performed equally in all three trained conditions (RCS, PCS, and RIS). This would suggest that the learning of core semantic features is reflected in semantic tasks, but this semantic advantage does not have any impact on reading. It is worth mentioning that all the target words used in this study had highly regular pronunciation, so this might explain why semantics did not have any influence on reading times. This is in agreement with a study conducted by McKay et al. (2008) which also found no differences in performance between words with regular spelling learned with meaning and without meaning. However, when they trained participants on words with inconsistent spelling, an effect of semantics emerged, speeding up RTs for words with meaning. Overall, the results of Experiment 2 in the reading task support the view that semantic richness does not play a significant role in the process of reading aloud; in this case, newly learned words in English L2. These findings are also in line with previous word learning studies in L1, which have shown no effect of semantics on reading aloud (McKague, Pratt, and Johnston, 2001; Nation et al., 2007).

Results in the recognition memory task did not support the predictions of the experiment since participants performed equally well in all three trained conditions (RCS, PCS, and RIS), despite the fact that conditions did differ in semantic richness, which was confirmed by the results of the semantic decision and cued recall tasks described earlier. These findings are inconsistent with previous studies using real English words in which researchers have reported semantic effects in lexical decision manipulating different semantic variables: number of semantic features (e.g., Pexman et al., 2002; Pexman et al., 2003); imageability (Balota et al., 2004; De Groot, Borgwaldt, Bos, & Van den Eijnden, 2002); semantic neighbourhood (Buchanan et al., 2001; Siakaluk et al., 2003), and number of semantic associates (Buchanan et al., 2001). There are a few differences between the current study and the studies described above. This study used a word learning paradigm in English, so it is likely that newly learned words do not behave like familiar words regarding recognition. For instance, it has been found that newly learned words need time to consolidate before they become fully lexicalized and show the same properties of real words (e.g., Dumay & Gaskell, 2007; Tamminen & Gaskell, 2008). In the present study, participants were tested only one night after training, so it is possible that the novel vocabulary did not have enough time to consolidate in neocortical areas, which would allow more automatic processing. Possibly, more time is needed in order to see semantic richness effects on the recognition of newly learned words. Tamminen and Gaskell (2008) found that novel words were able to engage in lexical competition with similar sounding existing words only a week after training, suggesting that several nights of sleep are needed for words to consolidate in long-term memory. Another reason for the lack of an effect in this task would be the semantic manipulation itself. Even though participants did learn more meaning when words were presented with core semantic features (RCS), as shown by the results in semantic decision and cued recall, this difference might not be sufficient for semantic effects to emerge during the recognition memory task. Unfortunately, there seem to be no previous studies in which semantic richness has been manipulated and that have used a recognition task. Probably the closest work to the current experiment is the study conducted by McKay et al. (2008). These researchers found better recognition for novel words trained with meaning (definition + picture) in comparison with novel words that only underwent phonological and orthographic training. These results suggest that a difference in recognition memory might emerge, but only when comparing a semantic condition with a nonsemantic condition.

In summary, this study manipulated feature variability and found that adding explicit core features to sentence contexts can boost the acquisition of word meaning as confirmed by the results in semantic decision and cued recall. However, gains in meaning did not show any reliable advantage during reading aloud or recognition memory.

2.4 General discussion

The two experiments presented in this Chapter were conducted in English as a second language and used a contextual word learning paradigm to investigate context variability (Experiment 1) and feature variability (Experiment 2). A series of tasks were used to assess the performance of L2 speakers after learning novel words. The main motivation to investigate context variability in Experiment 1 was to test the variability encoding theory developed by Bolger et al. (2008), which states that words are better learned when presented in a variety of contexts than in a single context. Using two new tasks (semantic decision and word naming), results supported this hypothesis since words in high context variability (HCV) showed an advantage in semantic decision. Results also showed the advantage in the reading task which suggests a semantic effect on reading since words in HCV were associated with better encoding of meaning.

Experiment 2 moved away from context variability to manipulate another variable: feature variability. As explained in preceding paragraphs, the reason for this change of methodology was due to the possible effect of 'attention' in Experiment 1 where participants were exposed to the same sentences six times in LCV, which could have caused boredom and disengagement from the task. The feature variability hypothesis in this experiment stated that words presented in sentence contexts including explicit core semantic features would be better learned than if presented in sentence contexts that did not provide core features. This experiment did not manipulate the number of different sentences as in Experiment 1, but rather the type and number of features participants were exposed to in each condition. Predictions in Experiment 2 were partially confirmed since results showed an advantage in semantic decision and cued recall for rich consistent semantics

(RCS) over poor consistent semantics (PCS), but no differences between the three trained conditions (RCS, PCS, and RIS) in reading and recognition memory.

Even though the two experiments in this Chapter did not manipulate the same variables, both variables reflect direct or indirect measures of semantic richness. In Experiment 1, context variability refers to the number of different contexts words are experienced with, and states that more contexts produce a more abstract or decontextualised meaning of the words (Bolger et al., 2008). From this, it can be inferred that words experienced in many contexts acquire much richer semantic representations, so context variability can be understood as a semantic variable such as *the number of semantic features* (e.g., Pexman et al., 2002), or *the number of semantic associates* (e.g., Buchanan, Westbury, & Burgess, 2001). In Experiment 2, the variable was feature variability which assumes that the inclusion of core semantic features, which offer direct access to word meanings, produces better learning than contextual features alone, which take longer to build abstract knowledge.

Taken both experiments together and assuming that both variables represent measures of semantic richness, the results are consistent regarding the semantic decision task. Both experiments produced better performance for words associated with more meaning (HCV in Experiment 1 and RCS in Experiment 2). These results were expected on the basis that semantic decision tasks are sensitive to the speed of semantic coding (Pexman et al., 2003). Thus, when participants are required to determine whether a word they learned was related in meaning to an existing word, they are able to respond more accurately and faster when the word was acquired with richer meaning.

The two experiments predicted an advantage in reading times for words learned with richer meaning. In Experiment 1, the results in the word naming task confirmed the predictions, but Experiment 2 failed to find a difference between conditions with rich and poor meaning. It is worth noting that there were clear methodological differences between the two experiments. In Experiment 1, the variable manipulated was context variability and novel words were presented in either 12 or 2 different contexts, whereas in Experiment 2 the variable was feature variability and words were presented in equal number of different sentences in each condition. Another distinction is that in Experiment 1 words corresponded to real but obscure concepts, whose orthographic form was not necessarily regular (e.g., *rodomontade*, *epiphyte*, etc). In Experiment 2, concepts were similar to those of

Experiment 1, but their word form was replaced with regular nonwords (e.g., *abrutmon*, *birlette*, etc). This might explain why the condition associated with richer meaning in Experiment 1 showed an advantage in reading aloud and this effect was not found in Experiment 2. As explained earlier, a role of semantics in reading is expected when a direct orthography-phonology mapping cannot be achieved and the process becomes particularly hard (e.g., Rodd, 2004). This implies that the effects found in Experiment 1 might be due to some of the words having irregular spelling. On the contrary, since words were assigned a highly regular spelling in Experiment 2, this might have washed out any possible effect of semantics due to the easiness with which participants could read the words despite the fact that they were learned in a foreign language. All this suggests that the results of both experiments in the reading task are not necessarily inconsistent. They seem to confirm the idea that semantics plays an important role in reading but only when the process is particularly difficult.

This assumption is consistent with the most contemporary models of reading aloud, which accentuate the role of a direct route between orthography and phonology during reading. Parallel distributed processing (PDP) models (Plaut et al., 1996; Seidenberg & McClelland, 1989) contain the orthography–phonology (O–P) pathway, whereas the dual route model (Coltheart, 1978; Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) have a direct lexical and a nonlexical route. Despite the emphasis on a direct reading route, these models have also proposed another pathway which operates via semantics as the orthography-semantics-phonology pathway of PDP models and the lexical semantic route of the dual route model. This implies that both models assume a role of semantics in reading to a greater or lesser degree. The evidence accumulated throughout the years seems to suggest that semantics might only play a significant role in reading when a direct mapping between orthography and phonology cannot be automatically established, as when reading low-frequency irregular words (e.g., Strain & Herdman, 1999; Strain et al., 2002). In these cases, semantics plays a disambiguating role by providing the necessary feedback to the orthographic and phonological representations so that reading can be successfully achieved.

The experiments in the current chapter suggest that the processing of reading aloud newly learned words in a second language does not seem to differ substantially from what is known in L1 word reading. Taken together, the experiments suggest that semantics might play a role in reading aloud (newly learned words in a second language), but only when these words have been learned with irregular spelling (Experiment 1). However, if novel words are given consistent spelling, semantics does not seem to influence reading accuracy or speed (Experiment 2). Both of these findings are consistent with the study by McKay et al. (2008) discussed in previous sections.

As mentioned in the discussion of Experiment 2, results in the recognition memory task were against predictions and this was attributed to differences between the current study and previous studies reporting semantic effects on lexical decision. I specifically argued that words in a second language might require more time to consolidate in neocortical areas to show differences in RTs during visual recognition. Additionally, a more substantial difference between semantic conditions might be needed in order for semantic effects to emerge since previous word learning studies have found an effect when comparing semantic versus nonsemantic conditions (e.g., McKay et al. (2008).

The results of the semantic decision and cued recall task were in line with predictions and seemed to fit in well with the existing literature. As shown in the discussion of Experiment 2, a number of studies have reported semantic richness effects on semantic categorization (Pexman et al., 2008; Grondin et al., 2006; Pexman et al., 2003; Pexman et al., 2002). Here I have shown that semantic richness can also affect the acquisition of novel words in a second language and this is reflected in both accuracy and RTs for newly learned words during semantic decision. Finally, the cued recall task showed that when words are learned with substantial meaning, they tend to be produced more accurately than when learned with poor meaning. This is explained by the fact that words with rich meaning produce more rapid matching between the concept's semantic representation and its phonological and orthographic forms.

The experiments in this chapter have looked at two semantic variables that can affect the processing of newly learned words: Context variability in Experiment 1 and feature variability in Experiment 2. From these two experiments, it can be concluded that words that acquire more meaning during training can be processed faster during semantic decision and cued recall. These words also seem to show an advantage in reading when spelling is not consistent, but no advantage when consistent. Recognition memory (old/new judgement) does not seem to be affected by semantic richness, possibly due to lack of consolidation over time or insufficient semantic distance between the conditions.

Chapter 3 - L1 and L2 novel word processing

3.1 Introduction

Chapter 2 presented two word learning experiments in English as a second language. Experiment 1 manipulated context variability (the number of different sentences in which novel words were presented), and found that novel words learned in 12 different sentences were responded to more accurately and faster in a semantic decision task. This was taken as evidence that context variation affects the learning of word meaning. Another relevant finding in this experiment was the fact that context variation also influences reading speed with faster RTs for words in high context variability. This was interpreted as evidence for semantic involvement (more contexts provide better learning of meaning) in the recognition of newly learned words in a second language. Experiment 2 manipulated a different variable named *feature variability,* which is a measure of the number and type of semantic features. Participants in this experiment learned new words in rich consistent semantics (core features + contextual features), rich inconsistent semantics (core and contextual features but from different concepts), and poor consistent semantics (only contextual features). The results of Experiment 2 showed that participants learned the meaning of words better when presented in rich consistent semantics (RCS) than in poor consistent semantics (PCS), which was evident in both the semantic decision and the cued recall task. Despite the gains in meaning in RCS with respect to PCS, no differences between the two conditions were found regarding naming or recognition memory (old/new judgement), which led to the conclusion that semantics does not seem to influence the recognition of newly learned words in a second language when the spelling of the words is carefully controlled (unlike Experiment 1).

Since the two experiments above were conducted in English as a second language, their results cannot be automatically generalised to a first language due to underlying differences between monolingual and bilingual word processing. Researchers working on bilingualism seem to agree that bilinguals are not simply two monolinguals in one, as some people might think, but possess very specific characteristics as speakers and hearers (e.g., Grosjean, 1989; de Groot et al., 2002).

The first big distinction between monolinguals and bilinguals involves the number of lexical representations they know for each verbalised concept. Bilinguals have two lexical representations for virtually every concept whereas monolinguals have only one. Hence, it is worth wondering whether this could have an impact on language processing since more lexical representations might use extra resources or might also benefit performance in some circumstances. The question one could ask is: do bilinguals activate lexical representations only from the target language or from both languages? Most balanced bilinguals report that when they read a text or listen to some speech in L1 or L2, they never get interference from the nontarget language. Despite their beliefs, the evidence seems to suggest that they do, but they do not become aware of it (e.g., Dijkstra, 2005). The fact that bilinguals activate lexical representations from both languages during reading or recalling object names has led to the conclusion that bilingual language processing is nonselective. This means that there is parallel activation of the two languages during visual or spoken word recognition (e.g., Sunderman & Kroll, 2006), and during word production (e.g., Costa, 2005). Despite compelling evidence that processing is nonselective in bilinguals, there is very little research addressing the impact of cross-language activity. One question that can be asked is whether having two languages activated at the same time is beneficial or detrimental for the expected outcome. The evidence seems to suggest that it could be both depending on the circumstances. For instance, if bilinguals have to decide whether a cognate is a word or not, they are generally faster than when the decision is made for noncognate words (e.g., De Groot, Delmaar, & Lupker, 2000; Dijkstra, Jaarsveld, & Ten Brinke, 1998). This facilitation effect can be simply explained by differences in frequency with other words. Cognates are words that overlap at a phonological, orthographic, and semantic level in the L1 and the L2 languages, so the fact that bilinguals know the two overlapping representations makes cognates stand out from other words which only have a representation in one language.

Since most words in any given language are not cognates, it is likely that in most cases the activation of the nontarget language causes interference, which can delay and not facilitate processing. A number of studies have shown that the activation of L1 words can cause interference in the L2, leading speakers to elicit wrong lexical representations or the translation equivalent instead of the right word (e.g., Poulisse & Bongaerts; 1994; Poulisse, 1999). This can affect both beginner and proficient speakers of a second language and it has been taken as evidence for parallel activation during language production. Other studies have shown this interference using interlingual homographs or 'false friends', which are words that have the same orthography in the first and the second language but mean completely different things. Von Studnitz & Green (2002) conducted an experiment with proficient German-English bilinguals which compared recognition of interlingual homographs and matched control words. The task consisted of pressing a "yes" button if the letter string was a word in English and a "no" button if the letter string was not. They found that participants responded more slowly to interlingual homographs than to control words.

Overall, the evidence for nonselective language processing in bilinguals is quite substantial. The effects of the nontarget language words becoming activated seem to facilitate the processing of the target words in some exceptional cases (e.g., cognates), and slow down performance in most cases due to more competition of lexical representations from the nontarget language. Since there is strong evidence supporting the idea that monolinguals and bilinguals differ in word processing and that competition of lexical representations in the nontarget language seems to interfere in the processing of target words, it would be relevant to compare the performance of L1 and L2 speakers after learning new words. There are virtually no word learning studies of this kind since most previous studies have mainly focused on differences regarding the processing of familiar words.

In order to compare L1 and L2 speakers' performance regarding the learning of new words, the current chapter will first present Experiment 3, which is a homologue of Experiment 2, but was conducted in English L1. Then a combined analysis of Experiment 2 (from Chapter 2) and Experiment 3 will follow in order to run comparisons across groups of speakers.

3.2 Experiment 3

In Experiment 3, monolingual English native speakers participated in a new experiment, which was exactly the same as Experiment 2 from Chapter 2, except that it was conducted in English L1. As in Experiment 2, *feature variability* (a measure of semantic richness) was manipulated in three conditions including rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics

(RIS). See description of Experiment 2 for more details. Performance was assessed using the same four tasks: word naming, recognition memory, semantic decision, and cued recall.

Predictions regarding semantic richness in this experiment are similar to those of Experiment 2. First, participants are expected to show better performance for words learned in RCS in comparison with PCS in the semantic decision and cued recall tasks. Second, given that Experiment 2 did not show any difference in performance regarding the naming task, Experiment 3 is expected to show a similar pattern. It was concluded from Experiment 2 that semantic richness did not influence naming because the novel words used in the experiment had a highly regular pronunciation. Since Experiment 3 was conducted in English L1, even less impact of semantic richness should be expected on reading. This prediction is mainly supported by the word learning study conducted by McKay et al. (2008) which found no semantic effect on reading. Third, semantic richness did not show any reliable effect in Experiment 2 in the recognition memory task. The main reason for the lack of effect was that the semantic distance between the conditions was probably not big enough to show a difference in performance. Since Experiment 2 was conducted in English L2, it is not clear whether L1 participants will show the same pattern. The study by McKay et al. did find a difference in recognition in a group of L1 speakers, but it compared a meaningful condition with a meaningless condition, so it differed from the current experiment in which all three conditions contain meaning.

3.2.1 Method

Participants

The participants in this experiment were 27 monolingual English native speakers enrolled in undergraduate or postgraduate courses (17 females; mean age 22.8 years, SD 3.1) at the University of York community. All individuals had normal or corrected-to-normal vision.

Materials and design

The stimuli and tasks used were the same described in Experiment 2, Chapter 2.

Procedure

The procedure was exactly the same as in Experiment 2, so it took place over 3 consecutive days. On day 1, participants completed a training session that consisted of two parts and lasted approximately 30 minutes. On day 2, they repeated part 2 only, and on day 3 they were tested on word naming, recognition memory, semantic decision, and cued recall. See procedure section of Experiment 2, Chapter 2 for full details.

3.2.2 Results

As in Experiment 2, 3 participants were removed from the analyses due to low performance in at least one of the task. One participant read 55% of the newly learned words incorrectly and had less than 70% accuracy in the recognition memory task. The other two participants removed scored below 70% in the recognition memory task, and less than 65% in the semantic decision task. The results for the remaining 24 participants are shown in Table 3.1. By participants (F_1) and by items (F_2) analyses were conducted on the data. Pairwise comparisons included Bonferroni corrected paired-samples t-tests. When Mauchly's test of sphericity was significant (sphericity not assumed), Greenhouse-Geisser degrees of freedom was used to assess the significance of *F*.

Table 3.1. Mean latencies and percent error rates for word naming, recognition memory, and semantic decision. Percent accuracy for cued recall.

	RCS	PCS	RIS	MT
		Word naming		
Mean RT	640	663	650	737
SD	114	122	117	188
% error	5.4	7.5	4.6	16.7
		Recognition memory		
Mean RT	727	755	720	---
SD	161	180	139	---
% error	5.4	8.8	7.9	---
		Semantic decision ("Yes" responses)		
Mean RT	1050	1269	---	---
SD	257	284	---	---
% error	22.9	37.9	---	---
		Semantic decision ("No" responses)		
Mean RT	1189	1333	---	---
SD	270	269	---	---
% error	26.7	29.6	---	---
		Cued recall		
% words recalled	56.7	44.2	---	---
SD	21.0	22.8	---	---

Note. RCS, rich consistent semantics; PCS, poor consistent semantics; RIS, rich inconsistent semantics; MT, minimal training.

Word naming

The total number of responses collected was 960. Eighty-two (8.5%) of the responses were eliminated. Two (0.2%) corresponded to naming latencies below 300 milliseconds, 18 (1.9%) to responses of 2.5 SD above the mean, 45 (4.7%) to mispronunciations and 17 (1.8%) to voice key triggered by a noise other than the word onset.

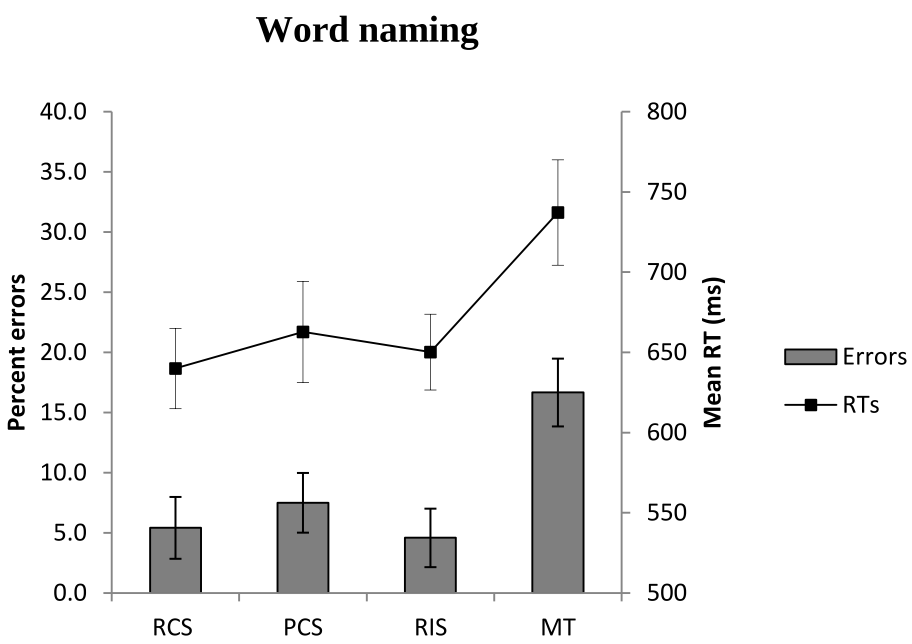


Figure 3.1. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS) and minimal training (MT) in the word naming task. Error bars represent standard error (SE) of the mean.

RTs

A one-way repeated-measures ANOVA was conducted on the data and revealed a highly significant main effect of conditions, $F_1(2, 23) = 17.59$, MSE = 5078.76, $p < .001$, $\eta_p^2 = .43$; $F_2(2, 39) = 13.73$, MSE = 92.93.45, $p < .001$, $\eta_p^2 = .26$. Bonferroni corrected t-tests ($\alpha = .05$) showed no differences between RCS, RIS, and PCS. However, all three conditions had faster RTs than MT.

Errors

The ANOVA conducted on errors also showed a main effect of conditions, $F_1(2, 23) = 9.50$, MSE = 77.96, $p < .001$, $\eta_p^2 = .29$, $F_2(2, 39) = 3.64$, MSE = 193.54, $p < .05$, $\eta_p^2 = .09$. Bonferroni adjusted pairwise comparisons ($\alpha = .05$) showed higher error rates in MT in comparison with RCS, PCS, and RIS. However, no difference was found for comparison between RCS > PCS, RCS > RIS, and PCS > RIS.

Recognition memory

In this task, participants were asked to visually recognize words for which they had received extensive training, so words in MT were not included. Hence, the number of responses collected was much lower than in the naming task reaching only 720 responses. Fifty-three (7.4%) were removed from the analysis: 27 (3.8%) corresponded to errors and 26 (3.6%) to latencies over 2.5 SD from the mean.

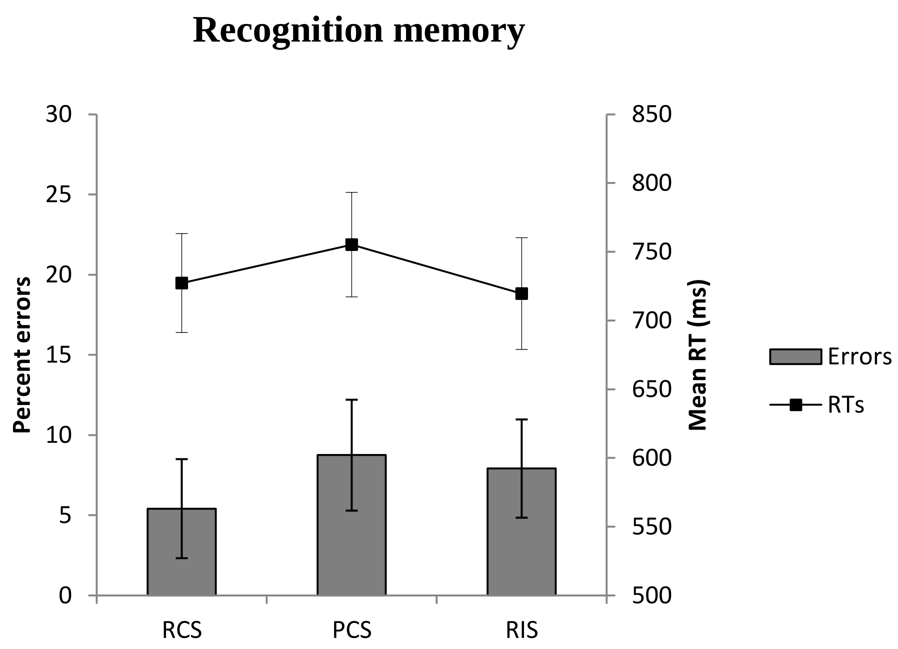


Figure 3.2. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics (RIS) in the recognition memory task. Error bars represent standard error (SE) of the mean.

RTs

A repeated-measures ANOVA showed a significant main effect of conditions, $F_1(2, 23) = 6.63$, MSE = 1281.29, $p = .01$, $\eta_p^2 = .22$; $F_2(2, 39) = 3.42$, MSE = 10689.36, $p = .05$, η_p^2 .08. Bonferroni corrected t-tests ($\alpha = .05$) revealed that recognition latencies in RCS and RIS did not differ significantly ($p = 1.00$). However, words in RCS and RIS showed significantly faster RTs than in PCS.

Errors

Unlike the RT data, the ANOVA conducted on error rates did not show an effect of conditions, $F_1(2, 23) = .75$, MSE = 96.86, $p = .48$, $\eta_p^2 = .03$; $F_2(2, 39) = 1.25$, MSE = 96.74, $p = .29$, η_p^2 .03.

Semantic decision

Two sets of data were collected in this task. The first set contained responses for familiar words that were semantically related to the newly learned words (*YES* responses). The second set grouped responses to familiar words which were unrelated in meaning to the newly learned words (*NO* responses). Hence, two separate analyses were conducted. Since participants only learned a consistent meaning for the novel words in two of the conditions (RCS and PCS), all analyses were limited to these two conditions.

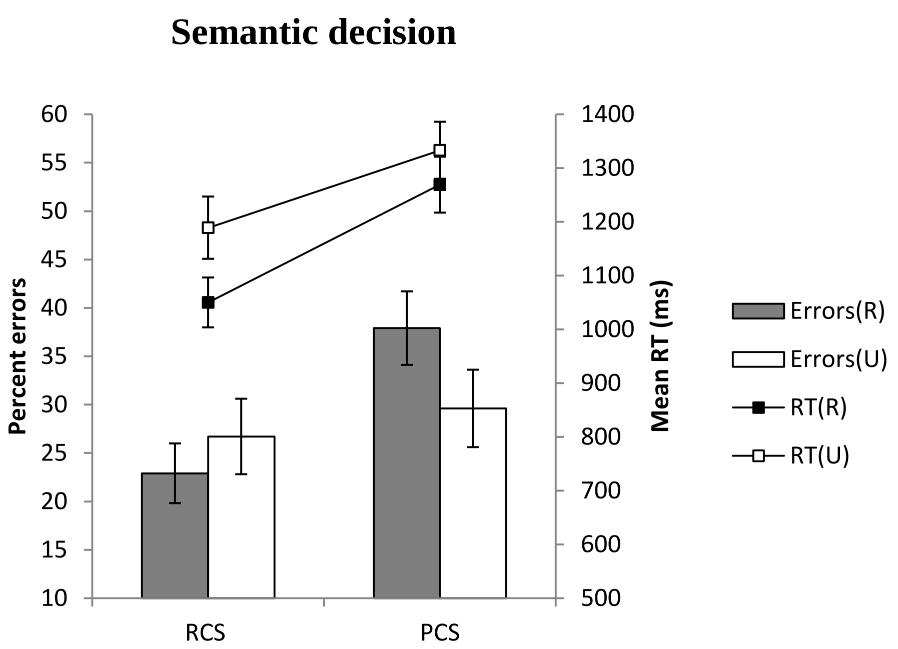


Figure 3.3. Semantic decision. Percent errors and RTs for semantically related (R) and unrelated (U) words in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars represent standard error (SE) of the mean.

Related (YES) responses

A total of 480 responses were collected. A hundred and forty-six (30.4%) responses were removed from the analysis. A hundred and thirty-four (27.9%) of these responses corresponded to errors, 8 (1.7%) corresponded to blank responses, and 4 (0.8%) to RTs over 2.5 SD from the mean.

RTs

A t-test was run on the data revealing faster responses in RCS than in PCS, $t_1(1, 23) = 5.28$, MSE = 41.44, $p < .001$; $t_2(1, 39) = 3.55$, MSE = 59.55, $p < .001$.

Errors

The t-test on errors showed significantly lower error rates in RCS than in PCS, $t_1(47) = 4.47$, $p < .001$; $t_2(79) = 3.98$, $p < .001$.

Unrelated (NO) responses

The number of responses collected for the unrelated trials was also 480. A total of 135 (28.1%) were removed from the analyses, of which 127 (26.5%) corresponded to errors, and 8 (1.7%) to failure in pressing any of the buttons.

RTs

A t-test was conducted on RTs showing shorter latencies for words in RCS than in PCS, $t_1(1, 23) = 2.40$, MSE = 59.85, $p = .03$; $t_2(1, 39) = 2.48$, MSE = 64.01, $p = .02$.

Errors

The t-test conducted on error rates did not show any significant difference between RCS and PCS, $t_1(1, 23) = .78$, MSE = 3.73, $p = .44$; $t_2(1, 39) = 1.05$, MSE = 3.75, $p = .30$.

Cued recall

In this task, participants were required to elicit the newly learned words based on a definition and some orthographic cues. The total number of correctly orally elicited newly learned words was 480. Participants produced 242 (50.4%)

correctly articulated words. Analyses in this task only included percent of words recalled.

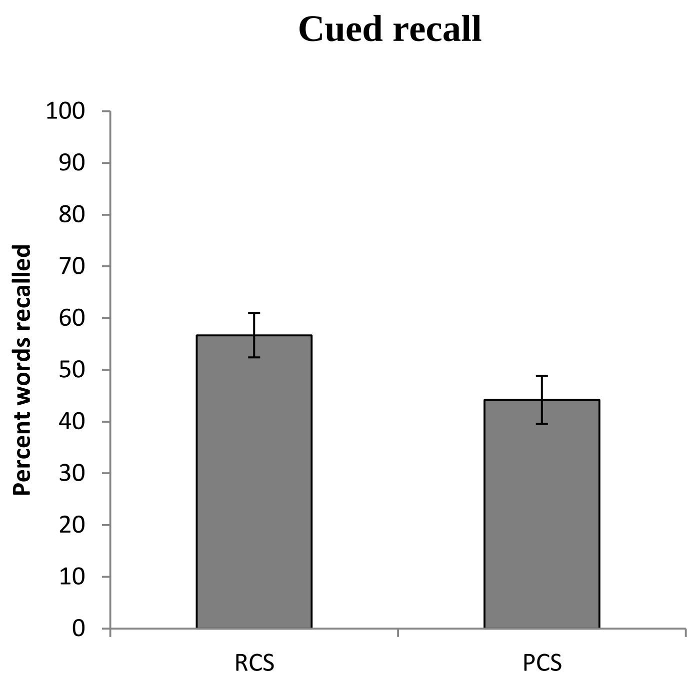


Figure 3.4. Percent words recalled in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars show standard error (SD) of the mean.

Percent accuracy

A t-test was conducted on the data revealing higher accuracy in RCS compared to PCS, $t_1(1, 23) = 2.12$, MSE = 5.91, $p = .05$; $t_2(1, 39) = 2.30$, MSE = 2.50, $p = .03$.

3.2.3 Discussion

As mentioned earlier, the predictions in Experiment 3 were similar to those in Experiment 2. The first prediction stated that participants were expected to learn words better when presented with words in RCS (core semantic features plus contextual features) than in PCS (contextual features alone). As in Experiment 2, this hypothesis was confirmed in both the semantic decision and the cued recall tasks. Regarding the semantic decision to semantically related word pairs (novel word and familiar word), participants showed a reliable effect in both RTs and error rates. In the decision to semantically unrelated pairs, this effect was only reliably significant for RTs. In cued recall participants recalled more items in RCS than PCS. The results of both tasks replicated the findings of Experiment 2, which suggests that words exposed to core semantic features during training acquired more complete and decontextualized meaning than words exposed to contextual features alone.

Predictions regarding reading aloud and recognition memory were less clear since Experiment 2 did not show any differences for trained conditions in any of these tasks. Thus, no difference in reading was expected in the current experiment whereas in recognition memory it was suggested that results might differ from those of L2 speakers even though no direction for the effect was predicted.

Experiment 3 replicated the findings of the naming task in Experiment 2 with participants performing equally well in all trained conditions (RCS, RIS, and PCS). However, in the recognition memory task, a different pattern of results emerged and participants showed faster RTs for words in RCS and RIS compared to PCS, while no differences between RCS and RIS were found. This finding was rather surprising since the L2 group did not show any effect. However, it is highly consistent with a previous word learning study conducted by McKay et al. (2008) which also found an effect of semantics on recognition memory, even though the manipulation was rather different since it included a semantic condition versus a nonsemantic condition.

The semantic effects on recognition memory are also consistent with previous studies with real words suggesting that semantic variables are more directly involved in lexical decision (a similar task to recognition memory) than reading (e.g. Rodd, 2004). This is supported by a number of studies that have found semantic richness effects on lexical decision, especially those that have manipulated the *number of semantic features* (e.g., Pexman et al., 2002; Pexman et al., 2003; Grondin et al., 2006). These studies normally explain this effect in terms of the feedback activation account, which assumes bidirectional semantic activation. According to this account, when a target word is presented, there is initial orthographic activation followed by semantic and phonological activation. Once semantic representations are activated, they can then increase orthographic activation via feedback connections. This implies that if words have rich semantic representations, feedback semantic activation to both phonological and orthographic representations is greater than when they have poor semantic representations. Hence, it can be hypothesised that when participants perform a word recognition task, words with rich meaning produce more semantic activation than words with poor meaning, which allows faster mapping between semantics, orthography and phonology, and consequently faster responses.

In accordance with the above, the advantage for words learned in RCS and RIS with respect to PCS might also reveal that meaning accelerates visual word recognition in L1 even when this meaning is inconsistent (RIS). The reason why this effect was only found for native speakers is somehow puzzling since all the other tasks showed the same pattern of results in both groups. It is possible, however, that due to more experience with the language, L1 speakers can benefit more from the exposure to core features in the training session and are able to extract and infer additional semantic features, which can boost their performance in RCS. This argument can be supported by the results of the semantic decision task since errors showed a trend in favour of the L1, which only affected RCS but not PCS.

Overall, the results of Experiment 3 did not vary quite substantially with respect to Experiment 2 regarding the learning of new words and the effects of semantic richness on naming, semantic decision, and cued recall. However, in the recognition memory task Experiment 3 showed a clear advantage for RCS and RIS in comparison with PCS. This suggests semantic richness can affect the recognition of newly learned words in English L1.

3.3 Experiment 2 and Experiment 3 (combined analysis)

Experiment 3 assessed the performance of L1 monolingual speakers in four different tasks: word naming, recognition memory, semantic decision, and cued recall. However, no direct comparison of results with the L2 group of speakers was done. The current analysis of the data includes a direct comparison between the results of Experiment 2 (L2 speakers) and Experiment 3 (L1 monolingual speakers). As explained in preceding sections of this chapter, both groups of participants underwent the same training and testing procedure under the same conditions. Predictions regarding differences between L1 and L2 speakers were made based on the experience of participants with the language and the differences regarding monolingual and bilingual memory.

L1 speakers were expected to outperform L2 speakers in all tasks due to the advantage of L1 speakers in terms of experience using the language. Even though the L2 speakers in this experiment were highly proficient, they did not learn English from birth, so factors such as Age of Acquisition (Morrison & Ellis, 2000), Familiarity (e.g., Connine, Mullennix, Shernoff, & Yelen, 1990), and frequency (e.g., Balota & Chumbley, 1984) might favour L1 speakers during learning, which can later be reflected in better performance in the tasks. If one also assumes that word processing is nonselective in bilinguals, as stated earlier, then more differences between L1 and L2 speakers should be found in the tasks that involve explicit word recognition (reading) and word production (cued recall). The activation of lexical representations in the nontarget language can interfere with the recognition and production of the target words in the L2 group of participants, which might accentuate differences in performance between L1 and L2 speakers.

The recognition memory task in this experiment might be less prone to interference from the nontarget language since it is a less direct measure of word recognition than reading aloud. This means that participants do not necessarily need to process a word completely to recognize it as a word they have learned. Previous studies have suggested that lexical judgment is not only a word-identification task but also a discrimination task which is sensitive to variables such as frequency, familiarity and the meaningfulness of the stimuli (e.g., Balota & Chumbley, 1984). Because participants can make a decision purely based on any of the variables above without completely processing the word, there is less chance that lexical representations from the nontarget language become activated and interfere with the word being recognized. This might result in more even performance between L1 and L2 speakers in this task.

Finally, in the semantic decision task, differences between the groups should be even less noticeable than in recognition memory because it directly targets meaning, reducing the number of competitors to only those which are semantically related.

In summary, differences between L1 and L2 speakers should be found in all tasks due to underlying differences in language experience, but effects are expected to be particularly large in naming and cued recall because these tasks would foster the activation of lexical representations from the nontarget language (Dijkskra, 2005), which can make the processing of the target words harder for L2 than for L1 speakers.

3.3.1 Method

Participants

The group of L2 English speakers (Experiment 2) and the group of monolingual native English speakers (Experiment 3).

3.3.2 Results

Results include a combined analysis of the data corresponding to both groups. Mixed-factorial ANOVAs were used to assessed overall performance of participants across the four tasks. By participants (F_1) and by items (F_2) analyses were used to identify main effects and interactions. Pairwise comparisons included Bonferroni corrected paired-samples t-tests. Results for word naming, recognition memory, semantic decision, and cued recall are presented.

Word naming

The combined analysis included four conditions: RCS, PCS, RIS, and MT. RTs and errors rates were analysed.

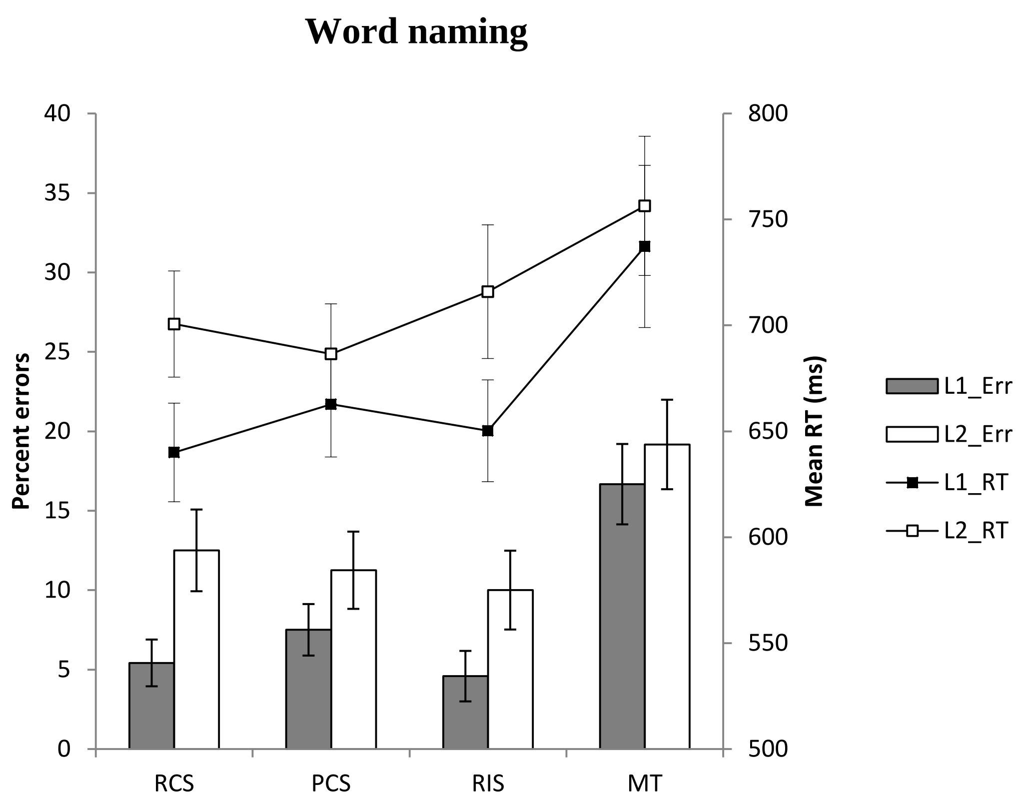


Figure 3.5. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), rich inconsistent semantics (RIS), and minimal training (MT) in the word naming task in L1 and L2 speakers. Error bars represent standard error (SE) of the mean.

RTs

A mixed factorial ANOVA was first conducted on RTs. A two-way mixed ANOVA was conducted on naming latencies. The mixed ANOVA showed no main effect of group by participants, $F_1(2, 46) = 1.23$, MSE = 17540.06, $p = .27$, $\eta_p^2 = .03$. However, this effect was significant in the by-items analysis with faster RTs for L1 speakers, $F_2(2, 78) = 12.42$, MSE = 2526.96, $p < .001$, $\eta_p^2 = .14$. There was a significant main effect of conditions, $F_1(2, 46) = 24.81$, MSE = 3527.06, $p < .001$, $\eta_p^2 = .35$; $F_2(2, 78) = 21.02$, MSE = 6173.05, $p < .001$, $\eta_p^2 = .21$. Bonferroni corrected paired-samples t-tests ($\alpha = .05$) showed slower response times for MT in comparison with all trained conditions (RCS, PCS, and RIS). However, no differences were found between the trained conditions. The interaction between group and conditions was marginally significant by subjects, $F_1(2, 46) = 2.87$, MSE = 3527.06, $p = .06$, $\eta_p^2 = .06$; but not significant by items, $F_2(2, 78) = 1.73$, MSE = 6587.22, $p = .16$, $\eta_p^2 = .02$. In order to break down the marginal interaction, Bonferroni corrected independent paired samples t-tests (α = .05) were also run on the data. None of the comparisons (L1 RCS > L2 RCS; L1 RIS > L2 RIS; L1 PCS > L2 PCS; L1 MT > L2 MT) showed a reliable effect by-participants. However, L1 RCS > L2 RCS [$t_2(1, 78) = 3.13$, MSE = 17.37, $p = .01$] and L1 PCS > L2 PCS [$t_2(1, 78) = 2.89$, MSE = 3.37, $p = .01$] showed a significant effect by-items with better performance for L1 speakers.

Errors

The mixed ANOVA conducted on errors showed a significant main effect of group with fewer errors for the L1 group, $F_1(2, 46) = 7.27$, MSE = 36.25, $p = .01$, $\eta_p^2 = .14$; $F_2(2, 78) = 5.19$, MSE = 81.23, $p = .03$, $\eta_p^2 = .06$. There was also a significant main effect of conditions, $F_1(2, 46) = 9.67$, MSE = 113.05, $p < .001$, $\eta_p^2 = .17$; $F_2(2, 78) = 9.78$, MSE = 251.79, $p < .001$, $\eta_p^2 = .11$. However, the interaction between group and conditions was not significant, $F_1(2, 46) = .42$, MSE = .113.05, $p = .74$, $\eta_p^2 = .01$; $F_2(1, 78) = .40$, MSE = 251.79, $p = .40$, $\eta_p^2 = .01$. Since there was an effect of group, but no interaction, the groups were pooled together and a one-way ANOVA was run on the data. Results showed a significant main effect of conditions, $F_1(1, 47) = 9.79$, MSE = 111.66, $p < .001$, $\eta_p^2 = .17$; $F_2(1, 79) = 9.85$, MSE = 249.38, p < .001, $\eta_p^2 = .11$. Bonferroni corrected t-tests (α = .05) showed no differences for contrasts between RCS, RIS, and PCS, but the difference between each of these conditions and MT was highly significant (p = 001).

Summary

Both RTs (by-items) and errors showed better performance for the L1 group. A main effect of conditions was also present in both measures. A marginal interaction between group and conditions was only observed for RTs. Further exploration revealed that the differences between the groups affected RCS and PCS more than RIS or MT. Bonferroni corrected paired-samples t-tests for both latencies and errors showed an advantage in performance for the trained conditions in each group with respect to MT. However, the trained conditions did not differ.

Recognition memory

Unlike the naming task, only three conditions were included here: RCS, PCS, and RIS. RTs and error analyses were conducted.

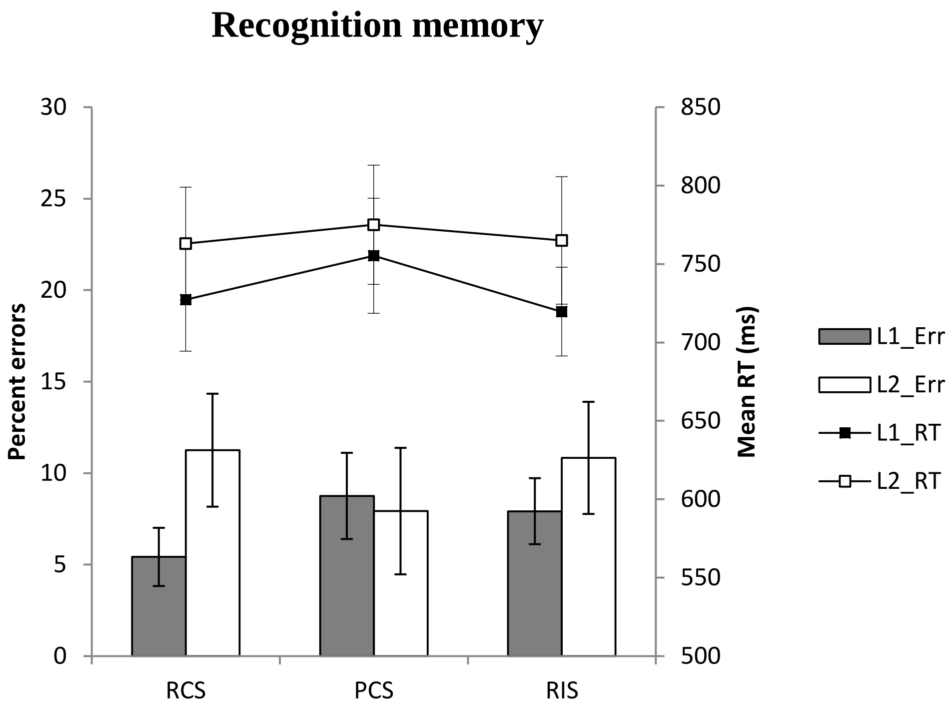


Figure 3.6. Percent errors and reaction times (ms) for rich consistent semantics (RCS), poor consistent semantics (PCS), and rich inconsistent semantics (RIS) in the recognition memory task for L1 and L2 speakers. Error bars represent standard error (SE) of the mean.

RTs

A mixed factorial ANOVA revealed no main effect of group, $F_1(2, 46) = .47$, MSE = 28910.86, $p = .50$, $\eta_p^2 = .01$; $F_2(2, 78) = .03$, MSE = 4742.88, $p = .87$, $\eta_p^2 = .00$. There was no significant effect of conditions, $F_1(2, 46) = 1.31$, MSE = 2370.73 $p = .28$, $\eta_p^2 = .03$; $F_2(2, 78) = 1.72$, MSE = 7660.43, $p = .19$, $\eta_p^2 = .02$. There was a marginal interaction between group and conditions (by-participants), $F_1(2, 46) = 2.70$, MSE = 2370.73, $p = .07$, $\eta_p^2 = .06$. The by-items analysis did not reach significance, $F_2(2, 78) = 2.33$, MSE = 7660.43, $p = .11$, $\eta_p^2 = .03$. Independent paired samples t-tests ($\alpha = .05$) were run to explore differences between the groups in each condition. However, none of the comparisons (L1 RCS > L2 RCS; L1 RIS > L2 RIS; L1 PCS > L2 PCS) showed a reliable effect.

Errors

The two-way mixed ANOVA on errors showed no main effect of group, $F_1(2, 46) = .73$, MSE = 114.48 $p = .40$; $F_2(2, 78) = 2.29$, MSE = 70.91, $p = .14$, $\eta_p^2 = .03$. There was no significant main effect of conditions, $F_1(2, 46) = .21$, MSE =87.49 $p = .81$, $\eta_p^2 = .01$; $F_2(2, 78) = .37$, MSE = 114.84, $p = .69$, $\eta_p^2 = .01$. The interaction between group and conditions was not significant either, $F_1(2, 46) = 1.65$, MSE = 81.49 $p = .20$, $\eta_p^2 = .04$; $F_2(2, 78) = 1.96$, MSE = 114.84, $p = .14$, $\eta_p^2 = .03$.

Summary

The mixed factorial analysis revealed no main effects of group and conditions regarding both latencies and errors. A marginal interaction between group and conditions was found in latencies, but not in errors.

Semantic decision

Analyses included the two conditions with consistent meaning: RCS and PCS. Latencies and errors for semantically related and unrelated word trials were analysed separately.

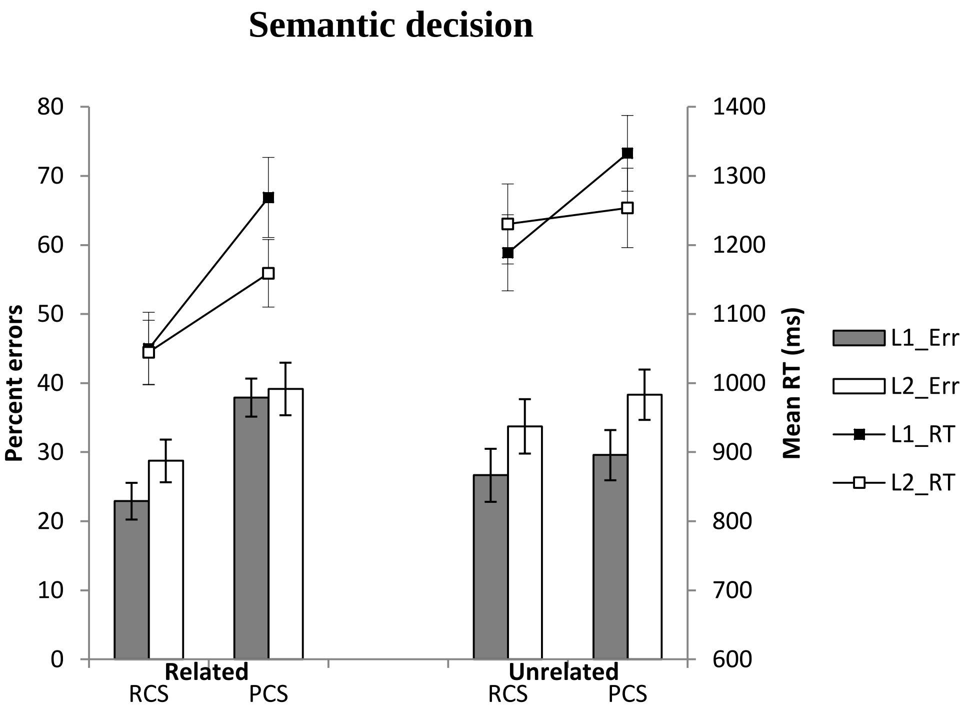


Figure 3.7. Semantic decision. Percent errors and RTs in L1 and L2 speakers for semantically related and unrelated words in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars represent standard error (SD) of the mean.

Related (YES) responses

RTs

A two-way mixed ANOVA was conducted on the data. There was no significant group effect, F_1 (2, 46) = .73, MSE = 54681.05, p = .40, η_p^2 = .02; F_2(2, 78) = 2.56, MSE = 40981.39, p = .11, η_p^2 = .03.There was a significant main effect of conditions, F_1(1, 46) = 35.84, MSE = 66678.59, p < .001, η_p^2 = .44; F_2(2, 78) = 15.45, MSE = 90582.81, p < .001, η_p^2 = .17. A marginal interaction between group and conditions was also found (by participants), F_1(1, 46) = 3.54, MSE = 65678.34, p = .07, η_p^2 = .44; F_2(2, 78) = 1.30, MSE = 90582.81, p = .26, η_p^2 = .02. Independent paired samples t-tests (α = .05) revealed no differences between groups in each condition (L1 RCS > L2 RCS; L1 PCS > L2 PCS).

Errors

The mixed factorial ANOVA conducted on errors showed no significant differences between the groups, F_1(1, 46) = 1.01, MSE = 148.80, p =.32, η_p^2 .02; F_2(2, 78) = .85, MSE = 297.17, p = .36, η_p^2 = .01. However, there was a significant main effect of conditions, $F_1(1, 46) = 23.31$, MSE = 166.26, $p < .001$, $\eta_p^2 = .34$; $F_2(2, 78) = 21.72$, MSE = 297.30, $p < .001$, $\eta_p^2 = .22$. The interaction between group and conditions was not significant, $F_1(1, 46) = .76$, MSE = 166.26, $p = .39$, $\eta_p^2 = .02$; $F_2(2, 78) = .71$, MSE = .297.30, $p = .40$, $\eta_p^2 = .01$. Since no differences between groups or interaction effect were found, the data from both groups were pooled together and a paired-samples t-test was conducted. The t-test showed a highly significant difference between the two conditions with better accuracy for RCS than PCS, $t_1(47) = 4.84$, $p < .001$; $t_2(79) = 4.67$, $p < .001$.

Summary

The groups of L1 and L2 speakers did not differ in either semantic decision latencies or error rates. The results also showed an advantage for RCS over PCS in both RTs and errors for both groups. There was a marginal interaction between group and conditions for semantic decision latencies, but no interaction was observed for errors.

Unrelated (NO) responses

RTs

The mixed factorial ANOVA showed no effect of group, $F_1(1, 46) = .07$, MSE = 61777.90, $p = .80$, $\eta_p^2 = .00$; $F_2(2, 78) = .13$, MSE = 26835.68, $p = .72$, $\eta_p^2 = .01$. However, there was a significant overall effect of conditions with faster RTs for RCS than PCS, $F_1(1, 46) = 5.80$, MSE = 28884.62, $p = .02$, $\eta_p^2 = .11$; $F_2(2, 78) = 9.33$, MSE = 61707.90, $p < .01$, $\eta_p^2 = .11$. No interaction was found between group and conditions, $F_1(1, 46) = 3.02$, MSE = 28884.62, $p = .09$, $\eta_p^2 = .06$; $F_2(2, 78) = .96$, MSE = 61707.90, $p = .33$, $\eta_p^2 = .01$.

Errors

The two-way mixed ANOVA carried out on the data revealed no differences between the groups in the by-participants analysis, $F_1(1, 46) = 3.02$, MSE = 248.69, $p = .09$, $\eta_p^2 = .06$. However, the by-items analysis showed a significant difference with the L1 group outperforming the L2 group, $F_2(2, 78) = 3.92$, MSE = 280.60, $p = .05$, $\eta_p^2 = .05$. There was a non-significant overall effect of conditions, $F_1(1, 46) =$

1.84, MSE = 183.61, p .18, η_p^2 = .04; F_2(2, 78) = 2.13, MSE = 339.99, p = .15, η_p^2 = .03. The interaction between group and conditions did not reach significance, F_1(1, 46) = .09, MSE = 183.61, p = .77, η_p^2 = .00; F_2(2, 78) = .01, MSE = 339.97, p = .91, η_p^2 = .00.

Summary

Native and non-native speakers performed evenly regarding semantic decision latencies to semantically unrelated word trials. Faster RTs for RCS than PCS, but no differences in error rates were found across the two groups. No interactions in the RT or error analyses were found. A significant difference (by items) was found in errors, with the L1 group showing lower error rates than the L2 group.

Cued recall

As in the semantic decision task, only RCS and PCS were included. Percent accuracy for words recalled is analysed.

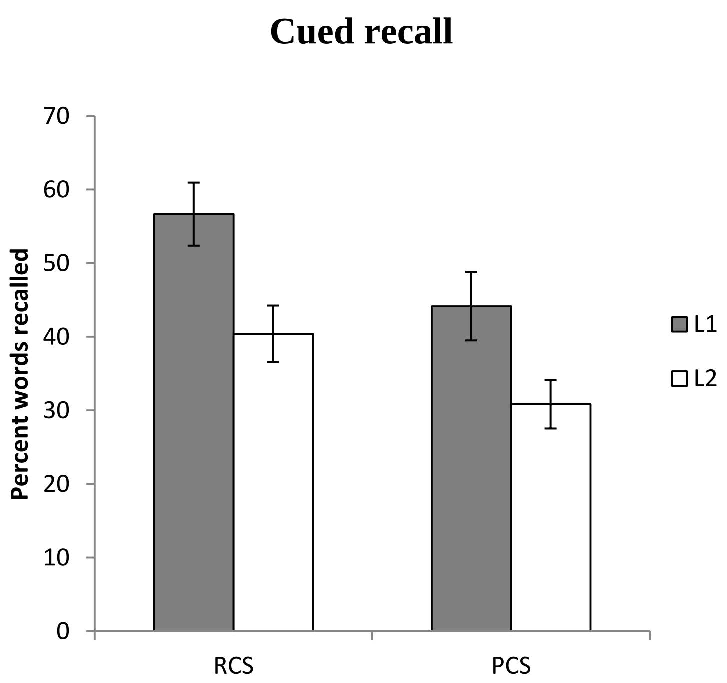


Figure 3.8. Cued recall. Percent words recalled by L1 and L2 speakers in rich consistent semantics (RCS) and poor consistent semantics (PCS). Error bars represent standard error (SD) of the mean.

Percent accuracy

Results showed a highly significant main effect of group with better performance for the L1 group, $F_1(1, 46) = 10.49$, MSE = 250.25, $p < .01$, $\eta_p^2 = .19$; $F_2(2, 78) = 10.64$, MSE = 296.07, $p < .01$, $\eta_p^2 = .12$. A significant main effect of conditions was also found revealing that performance was better in RCS than PCS, $F_1(1, 46) = 10.22$, MSE = 286.37, $p < .01$, $\eta_p^2 = .18$; $F_2(2, 78) = 13.88$, MSE = 126.46, $p < .001$, $\eta_p^2 = .15$. The interaction between group and conditions was not significant, F_1 (1, 46) = .18, MSE = 286.37, $p = .68$, $\eta_p^2 = .00$; $F_2(2, 78) = .24$, MSE = 126.46, $p = .62$, $\eta_p^2 = .00$. Corrected independent paired samples t-tests ($\alpha = .05$) showed reliable between group differences in each condition: L1 RCS > L2 RCS and L1 PCS > L2 PCS.

Summary

Results in the cued recall task showed a main effect of group with L1 speakers outperforming L2 speakers in each condition. There was also a main effect of conditions showing that participants in both groups performed better in RCS than PCS. No interaction between group and conditions was revealed.

3.3.3 Discussion

Overall, the results of the between group analysis confirmed the predictions. In the word naming task, the L1 group was more accurate than the L2 group, but this effect was less reliable in the RT data, which only became significant in the by-items analysis. The recognition memory task did not show any significant difference between the groups but overall there was a trend toward better performance of the L1 group. As expected, the semantic decision task showed the least difference between the groups with no clear trend in the data regarding related semantic responses and only an advantage in accuracy (by-items) for the L1 group in the unrelated judgements (‘No’ responses). The cued recall task showed the highest difference between the L1 and the L2 group with the first clearly outperforming the latter in both conditions.

As predicted, differences between L1 and L2 speakers were expected to be more substantial in naming and cued recall because these two tasks are more likely to be affected by cross-language interference and consequently producing a drop in performance in the L2 group. This idea is supported by models of bilingual word recognition (e.g., Dijkstra, 2005; de Groot et al., 2000) and production (e.g., Costa, Miozzo, & Caramazza, 1999; De Bot, 1992; Poulisse, 1999), which assume that second language activation is nonselective. This implies that the latent language also becomes activated during word recognition and word production. The fact that the semantic decision task showed virtually no differences between the two groups of participants suggests that participants did not differ substantially regarding the acquisition of word meaning. However, when performing tasks that require direct recognition (naming), and cued recall (production), L2 speakers seem to face more difficulties than monolingual speakers. The difference in these tasks, especially in cued recall might be due to the difference in the number of lexical representations that become available during word processing. Monolingual L1 speakers have fewer lexical representations than L2 speakers, so the searching of a newly learned lexical representations might be easier than for L2 speakers.

The group effect in the naming task can be explained using the Bilingual Interactive Activation Model (BIA) (Dijkstra et al., 1998). This is a very influencial model of bilingual memory and was based on McClelland and Rumelhart (1981)'s interactive model. The main difference between the BIA model and the monolingual model is that the first includes words from both languages in the 'integrated lexicon' and an extra node specifies language membership. The model represents different nodes that are arranged hierarchically beginning with the features node followed by letters, words, and finally languages. According to this model, when a bilingual is presented with a letter string, lexical candidates from both languages become activated and compete with each other during the recognition process. If, for example, a Spanish-English bilingual is presented with the letter string *notice*, it can initially activate neighbours in both languages, such as *noticia* (*news* in Spanish) and *notary* in English. Competition progresses until the top-down inhibitory mechanism suppresses the activation of the unintended language (Spanish in this case) and the correct word is recognized. Given the above, it can be proposed that a similar process occurred with bilinguals in the current study. Since they have lexical representations in both Spanish and English, more words were expected to compete for activation than in the monolingual group during word recognition. Thus, the cost of having more active lexical representations can affect the recognition speed of the target words as more candidates compete for selection. In the recognition memory (old/new judgement) task, a similar process probably occurred. However, the fact that participants do not need to fully identify the word in order to perform the old/new decision, interference from the nonresponse language is less severe than in naming, which is a more direct measure of word recognition (e.g., De Groot et al., 2002). Indeed, this was confirmed with results showing only a trend towards better performance for the L1 group in recognition memory.

It is true, of course, that L1 speakers were more proficient in the target language, and so they were expected to outperform L2 speakers. However, other tasks such as semantic decision did not show an advantage for the L1 group, which implies that the difference in reading might be mainly due to the number of lexical representations competing for activation during this process. Another important issue is the fact that newly learned words are not well-established representations and possibly not fully lexicalized after 2 days of training, so they might be more exposed to interference from competitors. As reviewed in Chapter 1, newly learned words need time to consolidate before they become fully integrated in the mental lexicon and can behave like real words (e.g., Dumay & Gaskell, 2003; Tamminen et al., 2010).

In the cued recall task, results showed the biggest difference between the groups. This was expected since the task involved the presentation of a definition and some orthographic cues upon which participants were required to elicit the correct word. In order to explain the results of this task, it is worth looking at models of speech production and their assumptions. Unlike bilingual word recognition, models of bilingual word production are much less developed. Costa (2005) proposed a model of bilingual word production based on the architecture of monolingual hierarchical models (e.g., Caramazza, 1997; Levelt, 1989). Both monolingual and bilingual models assume that activation flows from conceptual representations to the phonological representations. For example, upon the presentation of the picture of a dog (conceptual level), semantic representations start to be activated in the speaker's mental lexicon. These not only include the item denoted in the picture, but many other semantically related items. Then activation spreads to the lexical level or lexical nodes where lexical representations start to compete for selection. At this point, monolingual models propose the activation of several candidates within the target language, which have related semantic representations (e.g., *cat*, *dog*, etc). However, bilingual models also outline the lexical representations of the translations for *cat* and *dog* (in Spanish *gato* and *perro*). At this level, a bilingual person would probably have double the number of candidates than a monolingual, which can make the process of selection harder since more competition takes place due to the extra number of lexical candidates. Activation then flows to the phonological level but it is not clear whether all lexical representations remain activated or only the target word. In the final stage of the process (the phonological level) the target word is eventually produced.

Evidently, when more candidates become activated throughout the process of word production, more difficult the process becomes decreasing the chances of accurately eliciting a target word. In the current study, it can be proposed that higher number of lexical representations in the bilingual group might have produced a decline in performance in comparison with the monolingual group due to more interference from the nontarget language.

In summary, both word production and word recognition models assume a nonselective language mechanism during bilingual word processing. The results of the current analyses seem to support this view as differences in performance between bilinguals affected mainly naming and cued recall. These two tasks are likely to foster the activation of lexical candidates from both the response and nonresponse languages increasing the differences in performance between monolinguals and bilinguals.

3.4 Summary and conclusion

The analyses in the current chapter have shed light on the process of word learning in L1 and L2 speakers. It can be concluded that semantic richness affects semantic decision and cued recall in both groups of speakers. However, recognition memory seems to show semantic richness effects only in the L1 group. Regarding comparisons across groups, virtually no overall differences in performance were found between L1 and L2 speakers in recognition memory and semantic decision. However, naming and cued recall showed a clear advantage for the L1 group suggesting larger interference effects in these tasks from the nontarget language in the L2 group. This study is probably the first to compare the effects of semantic richness on word learning across monolingual and bilingual speakers. The current findings support the view that language processing is nonselective in bilinguals. Additionally, they suggest that the learning of word meaning in L1 and L2 does not differ substantially since differences in performance were mainly found for reading and cued recall, which are more prone to interference from the nonresponse language.

Chapter 4 – Learning and consolidation of new words

4.1 Introduction

Chapter 3 showed that both L1 and L2 speakers learned better when they were presented with novel words associated with core semantic features and contextual features (rich semantics) than contextual features alone (poor semantics). The Chapter also discussed differences in performance between L1 and L2 speakers in a series of tasks including naming, recognition memory (old/new judgement), semantic decision, and cued recall. Particularly relevant for the current Chapter was the fact that words learned with rich semantics were recognized faster in the recognition memory task, but only in the group of native speakers. This was interpreted as semantic involvement in the recognition of newly learned words, which was in line with previous studies using familiar words in which the same effects have been found in lexical decision tasks (word/nonword judgement) (e.g., Pexman et al., 2002; Borowsky & Masson, 1996). The results in the naming task did not produce any difference between the conditions with rich meaning and poor meaning in any of the groups, so this result was interpreted as lack of semantic involvement in naming newly learned words in L1 and L2. This finding supported previous studies that have not found a role of semantics in reading when novel words have been learned with regular spelling (e.g., McKay et al., 2008; McKague et al., 2001). In summary, the experiments in the previous chapter found that words learned with rich semantics produced better performance in semantic decision and cued recall in both groups of speakers. Regarding the recognition tasks, the effect of semantics on recognition memory was only found in the L1 group while no effect was found for naming in any of the groups. It is worth noting that participants in Experiment 3 were trained over 2 days and were tested on the third day, but no later test was given to assess retention over a longer period of time. It might be possible that the effects change over time as newly learned words get more consolidated.

As reported in Chapter 1, a number of studies have suggested that novel words need time to consolidate in order to become integrated into long-term memory (e.g., Gaskell & Dumay, 2003; Dumay et al., 2004; Dumay & Gaskell, 2007; Leach & Samuel, 2007; Davis et al., 2008). This process has been demonstrated using implicit measures of learning such as lexical competition (e.g., Dumay & Gaskell, 2007), perceptual learning (e.g., Leach & Samuel, 2007), and picture-word interference (e.g., Clay, Bowers, Davis, & Hanley, 2007). Even though the evidence is consistent regarding the fact that words need time to consolidate, it is not clear how much time is actually needed. It has been suggested that the integration of new lexical representations might take from 24 hours to a week (e.g., Gaskell & Dumay, 2007; Clay et al., 2007) and once established, lexical representations show long-term retention of up to 8 months (Tamminen & Gaskell, 2008). For instance, the pioneering study conducted by Gaskell and Dumay found that new lexical representations only engaged in lexical competition with similar-sounding existing words a week after training. They also found that performance in more explicit tasks such as recognition memory and cued recall showed good performance immediately after training and increased over time without additional training. In line with the findings above, a more recent study conducted by Davis et al. (2008) found that participants performed better in repetition, recognition memory, and meaning rating tests when tested on novel words learned the previous day rather than on words tested immediately after training. Likewise, they found a lexical competition effect when participants performed a lexical decision task on the real-word competitors of the novel words learned on the previous day. Taken together, these findings suggest that gains in performance on both explicit and implicit measures of learning seem to depend on the passing of time that takes place after training. Since this time involves at least 24 hours, it has been suggested that sleep might play a role in the consolidation of newly learned words (e.g., Dumay & Gaskell, 2007; Davis et al., 2008).

All the studies presented above trained participants on meaningless novel words, except for the works of Clay et al. (2007) and Leach and Samuel (2007). Particularly, Leach and Samuel argued that the addition of a new word in the mental lexicon should include its form (phonology and orthography), its meaning, and its syntactic role. They called this information *lexical configuration*, which they suggested develops over a long time course period, which can take weeks, months, or even years. Unlike, for instance, Dumay and Gaskell (2003), Leach and Samuel found that novel words were able to become integrated in the mental lexicon only if they were trained with meaning (a picture or a meaningful verbal context). When words were trained without meaning, they found no signs of integration reflected in lack of lexical engagement with existing words. This suggests that meaning is essential for words to become interleaved in neocortical areas and behave like real words, which is at odds with Dumay and Gaskell's findings. Leach and Samuel explained that the lexical competition effects found in Dumay and Gaskell's study were probably due to the fact that novel words derived from neighbour real words (e.g., *cathedruke* from *cathedral*), which would allow the new words to automatically acquire the meaning of their existing neighbours. In summary, there is substantial evidence to suggest that novel words need time to consolidate in order to become integrated in the mental lexicon, and that meaning seems to play a fundamental role in this process. Additionally, the integration of novel words in long-term memory seems to be accompanied by improved performance in both explicit (e.g., recognition memory) and implicit (lexical competition) tasks.

Even though studies such as that of Clay et al. (2007) and Leach and Samuel (2007) looked at semantic word learning, they did not assess whether performance on explicit tasks changes over time depending on whether novel words are learned with rich or poor meaning. There are reasons to believe that different patterns could emerge depending on a word's semantic richness. This idea is based on the *levels-of-processing* theory of long-term memory introduced by Craik and Lockhart (1972). The theory states that retention in long-term memory is determined by the depth of processing of the stimuli during encoding. Thus, when attention is diverted from an item that has just been learned, information about that item will decay at the rate appropriate to its level of processing. This means that if a stimulus is processed semantically (deep processing), it will be remembered better than if only processed in a perceptual fashion (shallow processing). This is in line with Leach and Samuel (2007)'s ideas, which suggest that in order for words to consolidate over time (not to be forgotten) is essential to link them with meaning during training.

A further step in the development of the levels-of-processing framework has established a distinction between shallow and deep semantic processing. For instance, it has been proposed that a high level of semantic elaboration increases the duration of memory traces in comparison with low semantic elaboration (e.g., Craik, 2002; Fliessbach, Buerger, Trautner, Elger, & Weber, 2010). This finer distinction in semantic knowledge was first introduced by Paivio (1975) who suggested that semantic knowledge is represented both verbally and nonverbally (e.g., an image). According to this view, processes that are represented in both verbal and nonverbal codes involved deeper semantic analysis and therefore contribute to better formation of memory traces than do processes which have only one code.

Since the evidence regarding levels of semantic processing suggests that deeper semantic analysis leads to better storage in long-term memory, it is reasonable to believe that novel words learned with meaning can be remembered better and for longer periods of time than novel words learned without meaning. Likewise, words acquired with rich meaning, should be stored more successfully than words learned with poor meaning. Additionally, according to the consolidation literature (e.g., Dumay et al., 2004), performance in both explicit and implicit measures should be enhanced over time and remain stable once words are integrated into the mental lexicon.

The experiments in this Chapter explore the assumption that integration of words in the mental lexicon is proportional to the acquisition of meaning during training. Thus, if rich meaning produces better formation of memories, performance for words with rich meaning should be enhanced over time or decay less than for words with poor meaning or no meaning.

4.2 Experiment 4

Experiment 4 used a similar methodology to that of Experiment 3. However, a few changes were incorporated in order to assess performance over a longer period of time (this is further explained in the Methods section). Four tasks were used in this experiment - recognition memory, word naming, semantic categorization, and word production. Unlike previous experiments, participants here were tested twice in order to assess long-term effects of word learning.

Predictions for the current experiment were made based on the findings of Experiment 3 and the literature discussed in the Introduction of this chapter.

First, Experiment 4 was expected to replicate the findings of Experiment 3 regarding the effect of semantic richness. Hence, in the recognition memory task, novel words learned with rich semantics should show an advantage in comparison with words learned with poor semantics or no semantics. Likewise, words with poor semantics should show better performance than words with no semantics. The word naming task showed no differences between trained conditions in Experiment 3, so a similar effect is expected in Experiment 4 regarding the data collected on day 3.

However, this might change on the second test due to possible differences between the conditions regarding consolidation over time. A semantic categorization task was also introduced here and if consistent with the findings in Experiment 3, it should show an advantage in performance for the rich semantics condition. A similar result should be found in the production task, which is a variation of the cued recall task in Experiment 3.

Second, predictions regarding the data collected on day 8 and the interaction between day and conditions are based on the evidence discussed in the Introduction. Given that performance is generally enhanced over time in explicit tasks, with no mediation of additional training (e.g., Dumay et al., 2004; Gaskell & Dumay, 2003), better results would be expected in all 4 tasks on day 8 in comparison with day 3. If formation of memory traces is proportional to the depth of semantic processing as in the levels-of-processing framework (e.g., Craik & Lockhart, 1972; Craik, 2002), words learned with rich meaning would show better consolidation over time than words with poor or no meaning.

4.2.1 Method

Participants

Twenty-one native English speakers (16 female; mean age 20.9, SD 2.8) from the University of York community participated in the study after given written consent. All individuals had normal or corrected-to-normal vision and had not been diagnosed with any language disorders.

Materials and design

A total of 40 nonwords were used in the experiment. All nonwords had a visual form and a spoken form, which was recorded by a male speaker with English as his native language. As in previous experiments, nonwords beginning with voiceless fricatives were avoided since they are not always detected by the voice key during reading aloud tasks. Four sets of items were created with nonwords matched on initial letter, length (8 letters long) and reading speed (e.g., Set A, *adertmon*; Set B, *apkander*; Set C, *almaisen*; and Set D, *ascarant*). See Appendix 4.1 for list of target nonwords. Two of the sets were assigned meaning, which corresponded to the meaning of real obscure concrete nouns such as names of animals (e.g., *axolotl*), plants (e.g., *epiphytes*), objects (e.g., *cestus*), etc. See Appendix 4.2 for list of obscure nouns and their corresponding definition. The original orthographic forms of these nouns were replaced with the nonwords in Appendix 4.1 in order to control for linguistic variables. Each of the 30 novel words used during the training session had versions of eight sentences in each condition (rich semantics, poor semantics, and no semantics), so 480 different sentences were created in total. In order to better control for semantic information in each condition, sentences were created based on the semantic features approach (McRae et al., 2005; Vigliocco et al., 2004) discussed in Chapter 1. The first two sets of sentences provided the same number of semantic features (1 feature per sentence), but differed in the type of features. In the rich semantics condition, participants were presented with the taxonomic or superordinate feature of the concept (e.g., *is an animal*) in the first sentence. Then, 2 or 3 other sentences conveyed core or specific semantic features (e.g., *has gills*, *lives in water*), while the rest of the sentences conveyed general features (e.g., *has two eyes*, *has white teeth*). The poor semantics condition also included the taxonomic feature in the first sentence, but all other sentences only conveyed general features of the concepts, so only information about the category could be extracted, but not about the specific concept. See examples in Table 4.1 and Appendix 4.7 for more sample sentences. In the no semantics condition, sentences were created using an artificial corpus made up of English-like nonwords. See sample sentences in Table 4.1 and Appendix 4.3 for full corpus and the equivalent English translations.

Sentences in all conditions were between 5 and 10 words long and sets were matched on average sentence length, $F_1(1, 19) = 1.2$, MSE = 2.52, $p = .31$, η_p^2 .06. Sixty images of real concrete nouns were selected from the Web and later modified to complement each condition. In rich semantics, novel words were accompanied by standard-resolution images to allow access to all visual features, whereas in poor semantics participants were only exposed to pixalated versions of the images that mainly conveyed surface features such as colour or shape. In the no semantics condition, only empty rectangular shapes accompanied the words in each presentation. The size of the images was adapted so they had an area between 29cm2 and 50cm2. In total, 60 images were created for the two semantic conditions and 3 different versions of rectangular shapes were created for the condition with no semantics. See Appendix 4.4 for sample visual stimuli used in each condition. In the recognition memory task, 30 filler items were used and were equivalent to trained non-words in initial letter, and length. The full list of filler items is found in Appendix 4.5. The words trained with meaning corresponded to 10 different categories with two words per category. See Appendix 4.6 for list of categories. Twenty definitions were also used in the production task as stimuli for participants to elicit the corresponding newly learned word. See Appendix 4.8 for full list.

Table 4.1. The novel word *adertmon* presented in 4 sample sentences in the trained conditions: rich semantics, poor semantics, and no semantics.

Rich semantics	Poor semantics	No semantics
An adertmon is an animal	*An adertmon is an animal*	*Tun adertmon nel tun suzen rosnow*
An adertmon is an amphibian	*An adertmon has two round eyes*	*Tun adertmon nel tun replan*
An adertmon is found in Mexico	*An adertmon has fairly small legs*	*Tun adertmon heerds melb tun evryn*
An adertmon lives in water	*An adertmon has white teeth*	*Tun adertmon mels groud incrates*

Procedure

The experiment took place over the course of a week with 4 sessions in total. The first 3 sessions occurred each on a different consecutive day whereas the last session took place a week later (day 8). On days 1 and 2, participants completed a training session that consisted of 3 parts and lasted approximately 50 minutes. On Day 3, they were required to complete a test that included recognition memory, word naming, semantic categorization, and production. On Day 8, the same test was repeated.

Training procedure

Before the first training session, participants signed a consent form which contained a brief description of the study. Then participants were asked to sit in front of the computer and type in their participant number, group, age, and session to start. In the first part of training session 1, participants were simultaneously presented with a spoken word and a picture (RS), a spoken word and a pixelated image (PS), or a spoken word and an empty rectangular shape (NS). Each word was presented four times in random order, and participants were required to say the words aloud immediately after the presentation of the stimuli (spoken word and image). There were three different versions of each image to show a different angle or different context of the concept (object, animal, plant, etc) or a slightly different rectangular shape, in the NS condition. The visual stimuli were displayed for 4 seconds preceded by a fixation cross for 500 milliseconds.

In the second part of the training session, participants were first presented with a visual stimulus (standard resolution image, pixelated image, or rectangular shape) and both the spoken and written forms of the novel words. They were required to repeat each word after they heard it or saw it on the screen. Then the target word was presented in 4 different sentences that required participants to read them carefully. After the presentation of the four sentences, the target word appeared again in isolation, as in the previous presentation, and participants were asked to pronounce it aloud. Finally, four more short sentences followed. The procedure was the same for every word. Conditions were blocked and the sets of words were rotated around conditions and across participants. The words in written form, displayed either in isolation or in sentences, were presented in lower-case, 18-point black Courier New font on white background. For the oral modality, participants were required to put on headphones to hear the novel words.

The third part of the training session required participants to type in the words after the presentation of their phonological and written forms, and the corresponding visual stimuli. The aim of this task was to reinforce the association between the novel words' components: phonology – orthography – semantics in the conditions with meaning and phonology – orthography in the condition without meaning.

On day 2, a slightly different training session was provided. Part 1 was exactly the same as its day 1 equivalent except that participants also saw the written forms of the words. The second part was also very similar, but it included a new version of the visual stimuli, so participants would see one old stimulus and a new one corresponding to the same concept (in RS and PS) or rather similar shape (in NS). Part 3 on day 2 was exactly the same as part 3 on day 1. See Figure 4.1 for structure of training session.

Overall, participants were exposed to each word four times in spoken modality accompanied by a visual stimulus, 10 times in spoken and written modality in association with a visual stimulus, and 16 times in linguist contexts. They also typed in each word two times. Frequency was kept constant in all three trained conditions.

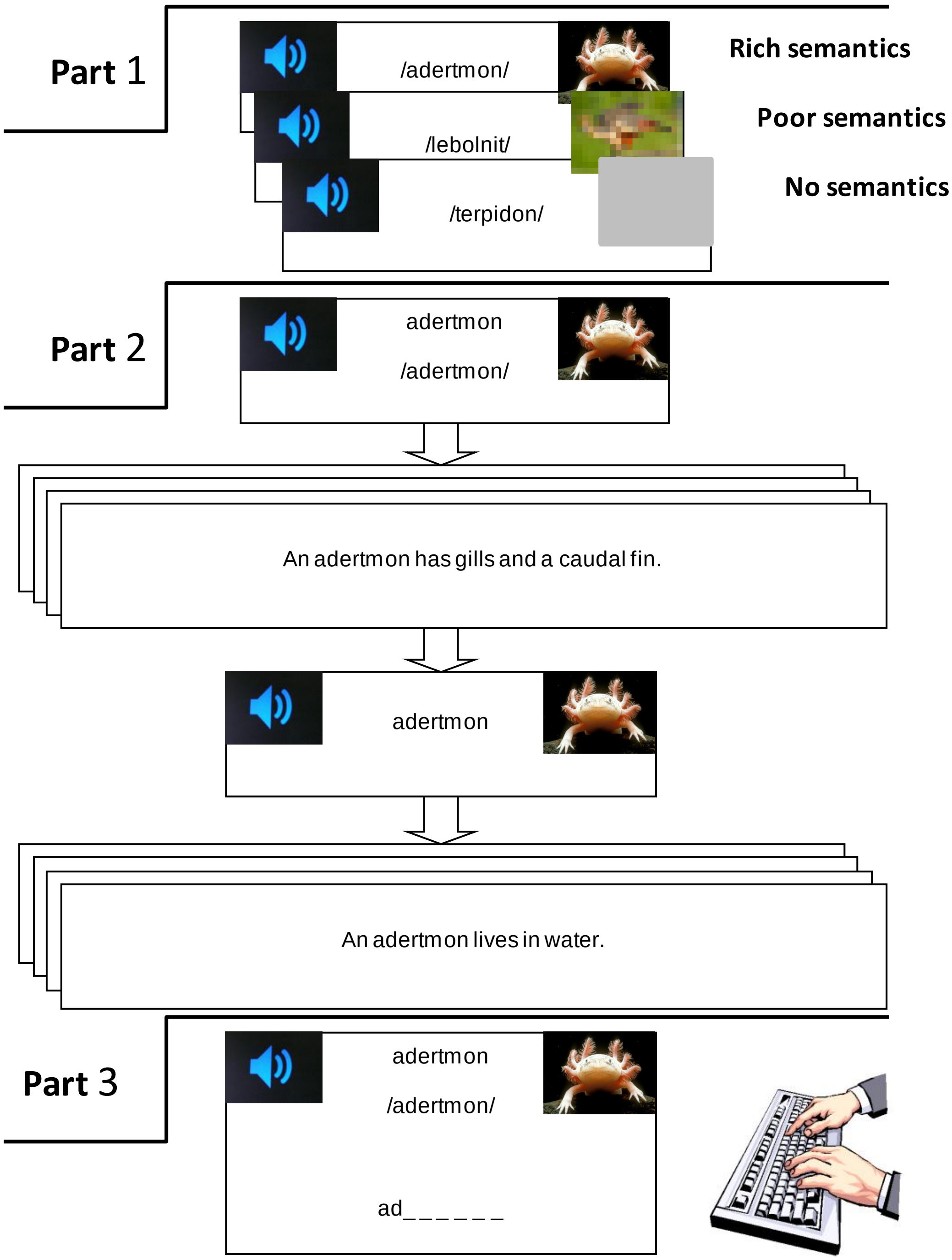


Figure 4.1. Structure of training procedure in Experiment 4 (day 1). Part 1 shows 3 sample words being presented, each representing a different condition. Parts 2 and 3 only show the rich semantics condition, but the training for the other two conditions was exactly the same.

Testing procedure

The experiment was run on a PC computer using E-Prime software (Schneider et al., 2002). All verbal stimuli were presented visually in 18-point black Courier New font on white background, except for the definitions in the production task, which were displayed in 16-point. The word naming task included rich semantics, poor semantics, no semantics, and no training conditions. The inclusion of the latter condition responded to the need of having a baseline to compare all trained conditions to it. In the recognition memory task, participants were only tested in the trained conditions by deciding whether a letter string was new or had been presented during the training. The semantic categorization and production tasks were performed on both sets of words trained with meaning (RS and PS). The first task presented was recognition memory followed by naming then the two semantic tasks – semantic categorization and word production. Participants were first tested on day 3 and then again on day 8.

Recognition memory

Before the actual presentation of the experimental stimuli, participants were asked to complete 10 practice trials in order to get used to the task. The task procedure consisted of the presentation of a fixation cross for 1 second followed by the target item for 3 seconds or until participants made a response. Finally, a blank screen was displayed for 1 second to indicate the end of the trial. Items were presented in random order which was different for each participant and on each day. Responses were made on an E-Prime response box by pressing either the left button for words learned during the training session, or the right button for untrained nonwords. Responses including accuracy and latencies from the onset of the stimulus presentation to the onset of the button press were recorded for analysis.

Word naming

As in the recognition memory task, each word was preceded by a fixation cross for 1 second. The target word appeared immediately after and remained on the screen for 2 seconds or until participants produced a response, then a blank screen was presented for 1 second to mark the end of each trial. All items were presented in a different random order to each individual. Participants were instructed to read the items aloud as soon as they appeared on the screen, and as quickly and accurately as possible. Five practice trials, including real low-frequency words, were introduced before the experimental trials, so participants could familiarise themselves with the task. Naming latencies were collected by means of a microphone connected to a voice key. Software package, Audacity (version wi-1.3.0b) (Audacity Development Team, 2010) was used to record and edit audio files in order to identify pronunciation errors.

Semantic categorization

The semantic categorization task started with the presentation of 4 practice trials that required participants to categorize familiar words in order to get to know the task. The procedure was exactly the same as that of the actual experiment, which began with the presentation of a fixation cross for 1 second, followed by a blank screen for 500 milliseconds. A trained novel word with meaning was then displayed for 2 seconds (e.g., *duntrane*) followed by two category labels (e.g., *fruit – weapon*) for 3 seconds or until a response was made. Participants were required to indicate which category the word belonged to by pressing “1” (for the category appearing on the left) or “2” (for the category on the right) on the response box. Responses including accuracy and semantic categorization latencies were recorded for analysis.

Production

Participants were presented with a blank screen for 500 milliseconds. A short definition of a word was then displayed and stayed on for 8 seconds. The precise definitions employed in this task had not appeared in the training. Individuals were instructed to read the definition and think about a novel word that could correspond to it. Immediately after the presentation of the definition, participants saw a fixation cross for 500 milliseconds and then the first letter of the target word was displayed, accompanied by a dashed line indicating the number of letters missing (e.g., a_ _ _ _ _ _ _). Participants were asked to type in the complete word as soon as the initial letter was displayed. Accuracy of response was measured by counting the number of correct letters typed in the right positions.

4.2.2 Results

The data analysis was performed on all 21 participants. They were tested twice to assess performance over the course of a week. Error and latency data were analysed with analyses of variance (ANOVAs) using both subject ($F1$) and item ($F2$) test statistics. Pairwise comparisons were performed using Bonferroni correction for multiple comparisons. Only reaction times (RTs) for correct responses were included and RTs below 300 milliseconds were regarded as outliers and removed from the analysis. This was also the case for RTs over 2.5 SD from the mean. When Mauchly's test of sphericity was significant (sphericity not assumed), Greenhouse-Geisser degrees of freedom was used to assess the significance of F. The means for trimmed, correct RTs (recognition memory, word naming, and semantic categorization tasks) and percent accuracy of response (word production task) are shown in Table 4.2.

Table 4.2. RTs corresponding to recognition memory, word naming, and semantic categorization; percent accuracy of response for word production on day 3 and day 8.

	Day 3				Day 8			
	RS	PS	NS	NT	RS	PS	NS	NT
				Recognition memory				
Mean RT	689	699	698	---	632	651	676	---
SD	117	141	130	---	102	133	117	---
% errors	3.8	9.5	6.7	---	2.4	5.7	7.6	---
				Word naming				
Mean RT	553	556	558	662	557	560	567	657
SD	78	81	85	128	74	83	73	117
% errors	3.8	3.3	7.1	13.8	4.3	4.3	2.9	12.4
				Semantic categorization				
Mean RT	767	1087	---	---	723	950	--	---
SD	225	464	---	---	191	367	---	---
% errors	6.2	21.9	---	---	4.3	20.5	---	---
				Word production				
% accuracy	88.8	59.0	---	---	92.4	75.6	---	---
SD	17.0	27.9	---	---	12.2	28.9	---	---

Note. RS, rich semantics; PS, poor semantics; NS, no semantics; NT, no training.

Recognition memory

Participants were given exactly the same test on day 3 and day 8. A total of 630 responses were collected on each occasion. On day 3, 42 (6.7%) RTs were removed from the analysis, of which 23 (3.7%) corresponded to errors, while the remaining 19 (3.0%) corresponded to RTs situated beyond the range of 2.5 SD from the mean. On day 8, a total number of 33 (5.2%) responses were eliminated, 17 (2.7%) were errors and 16 (2.5%) corresponded to RTs over 2.5 SD from the mean.

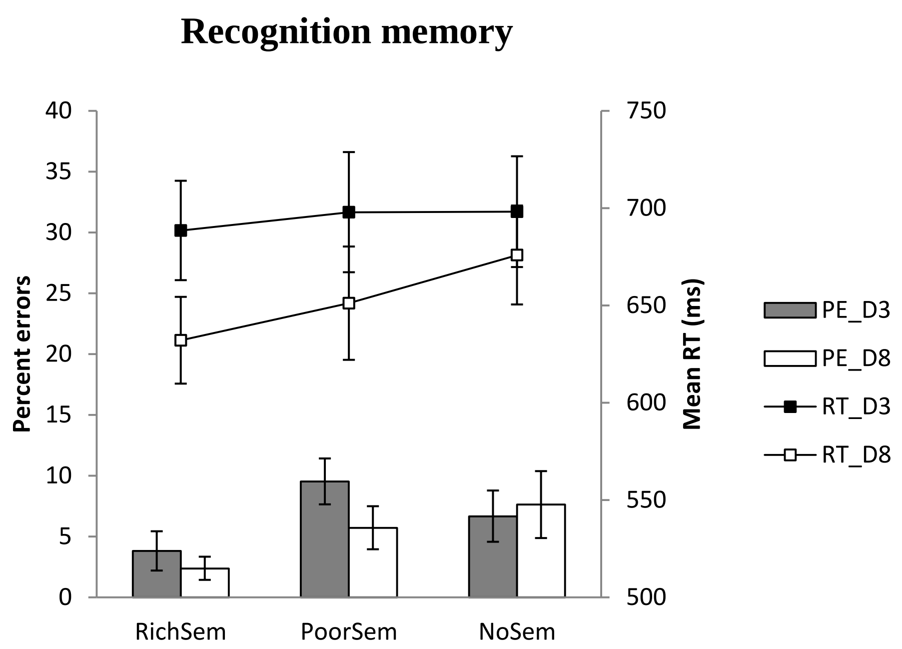


Figure 4.2. Percent errors (PE) and reaction times (RT) in rich semantics (RichSem), Poor semantics (PoorSem), and No semantics (NoSem) on day 3 (D3) and day 8 (D8). Error bars represent standard error (SE) of the mean.

RTs

A two-way repeated measures ANOVA was first conducted on RTs for correct responses with day and conditions as the main factors. Results revealed a significant main effect of day, with overall RTs being faster on day 8 than on day 3, $F_1(1, 20) = 6.21$, MSE = 8888.93, $p = .02$, $\eta_p^2 = .24$; $F_2(2, 29) = 44.98$, MSE = 1809.08, $p < .001$, $\eta_p^2 = .61$. There was also a significant effect of conditions in the by-participants analysis, $F_1(2, 40) = 3.94$, MSE = 1927.80, $p = .03$, $\eta_p^2 = .17$, but not in the by-items analysis, $F_2(2, 58) = .54$, MSE = 15137.71, $p = .59$, $\eta_p^2 = .02$. The interaction between day and conditions was also significant by-participants, $F_1(2, 40)$

= 3.59, MSE = 1288.37, p = .05, η_p^2 = .15, but did not reach significance by-items, $F_2(2, 58)$ = 2.01, MSE = 24442.21, p = .14, η_p^2 = .07.

In order to further explore the main effects and the interaction found, separate one-way repeated measures ANOVAs were also conducted on RTs for responses collected on day 3 and day 8. The ANOVA on the day 3 data showed no main effect of conditions, $F_1(1, 20)$ = .52, MSE = 1261.91, p = .60, η_p^2 = .03. However, the ANOVA on RTs collected on day 8 revealed a highly significant main effect of conditions, $F_1(1, 20)$ = 6.49, MSE = 1571.55, p = .01, η_p^2 = .25. In order to assess the results of direct comparisons between conditions within day 8, Bonferroni corrected paired-samples t-tests (α = .05) were also run on the data. Results showed no significant differences between RS and PS. However, a significant difference emerged for RS versus NS, and PS versus NS, with slower RTs in NS. Bonferroni corrected t-tests (α = .05) were also conducted on each condition across days in order to assess improvement over time. They revealed significantly faster RTs for RS on day 8 compared to RS on day3. However, no reliable difference between day 3 and day 8 was found for PS and NS.

Errors

As in the RT analysis, a factorial repeated measures ANOVA was conducted on errors with day and conditions as the main factors. Results showed no main effect of day, $F_1(1, 20)$ = 1.79, MSE = 35.952 p = .20, η_p^2 = .08; $F_2(1, 29)$ = 1.99, MSE = 57.08, p = .17, η_p^2 = .06. There was no significant main effect of conditions in the by-subjects analysis, $F_1(2, 40)$ = 2.52, MSE = 102.86 p = .09, η_p^2 = .11; but this difference was significant in the by-items analysis, $F_2(2, 58)$ = 4.11, MSE = 89.17, p = .02, η_p^2 = .12. The interaction between day and conditions was not significant, $F_1(2, 40)$ = .29, MSE = .46.20 p = .29, η_p^2 = .06; $F_2(2, 58)$ = 1.58, MSE = 54.50, p = .21, η_p^2 = .05.

Summary

The factorial ANOVA conducted on recognition latencies showed a significant main effect of day and conditions, and an interaction between the two factors in the analysis by participants. Separate ANOVAs for each day revealed no main effect of conditions on day 3, but a significant effect emerged on day 8 in the absence of major changes in error rates. Pairwise comparisons on day 8 data revealed no reliable differences between the two semantic conditions (RS and PS), but both showed faster RTs than NS. However, the comparisons across days showed reliable differences in RS (day3) compared with RS (day8), with faster RTs on day 8. The PS and NS conditions showed no reliable differences between day 3 and day 8 performances.

Word naming

Word naming was the second task of the testing session with 840 responses recorded on each day. On day 3, 59 (7.0%) RTs were eliminated from the analysis due to various reasons. These included 12 (1.5%) voice key errors, 31 (3.7%) mispronunciations, and 16 (1.9%) outliers (below 300 ms or 2.5 SD above the mean). On Day 8, 50 (6%) RTs were deleted from the analysis, of which 6 (0.7%) corresponded to voice key errors, 16 (1.9%) to mispronunciations, and 28 (3.3%) to outliers. RTs and error analyses were conducted using two-way ANOVAs with day (day 3, day 8) and conditions (RS, PS, NS, NT) as the main factors.

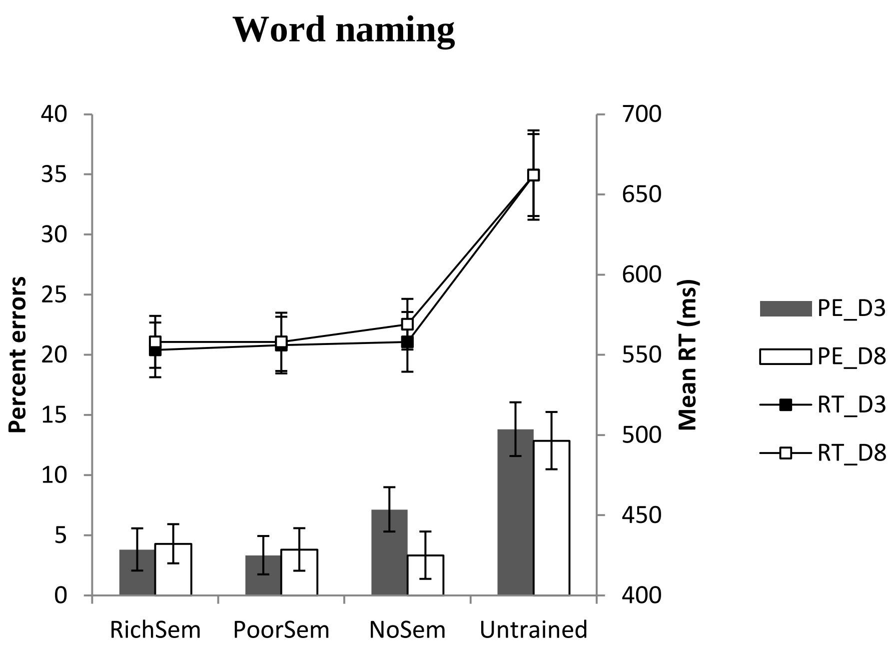


Figure 4.3. Percent errors (PE) and reaction times (RT) on day 3 (D3) and day 8 (D8) in rich semantics (RichSem), poor semantics (PoorSem), no semantics (NoSem), and untrained. Error bars represent standard error (SE) of the mean.

RTs

The factorial ANOVA conducted on correct, trimmed naming latencies revealed no main effect of day, $F_1(1, 20) = .18$, MSE = 2360.48, $p = .68$, $\eta_p^2 = .01$; $F_2(1, 29) = 1.65$, MSE = 881.37, $p = .21$, $\eta_p^2 = .05$. There was a significant main effect of conditions, $F_1(3, 60) = 54.60$, MSE = 1968.40, $p < .001$, $\eta_p^2 = .73$; $F_2(3, 87) = 76.29$, MSE = 1960.92, $p < .001$, $\eta_p^2 = .73$. The interaction between day and conditions was not significant, $F_1(3, 60) = .72$, MSE = 505.96, $p = .55$, $\eta_p^2 = .04$; $F_2(3, 87) = .34$, MSE = 976.28, $p = .80$, $\eta_p^2 = .01$. Since participants performed almost identically on day 3 and day 8, the data from both days were merged to explore comparisons regarding the four conditions. Bonferroni corrected t-tests ($\alpha = .05$) only showed a reliable difference between each of the trained conditions (RS, PS, NS) and the untrained condition (NT), with faster RTs for the trained conditions.

Errors

A two-way ANOVA was also conducted on errors. Results showed no main effect of day, $F_1(1, 20) = .48$, MSE = 100.71, $p = .49$, $\eta_p^2 = .02$; $F_2(1, 29) = 1.11$, MSE = 49.16, $p = .30$, $\eta_p^2 = .04$. A significant main effect of conditions was found, $F_1(3, 60) = 13.54$, MSE = 60.99, $p < .001$, $\eta_p^2 = .02$; $F_2(3, 87) = 13.15$, MSE = 95.65, $p < .001$, $\eta_p^2 = .31$. The interaction between day and conditions was not significant, $F_1(3, 60) = 1.48$, MSE = 55.62, $p = .23$, $\eta_p^2 = .07$; $F_2(3, 87) = 1.20$, MSE = 51.26, $p = .32$, $\eta_p^2 = .04$. As in the RT analysis, the error data were merged across days to investigate further comparisons. Bonferroni corrected paired-samples t-tests ($\alpha = .05$) revealed the same pattern of results as in the RT analysis with higher error rates for the untrained condition (NT) compared to each of the trained conditions (RS, PS, NS). No differences were found for comparisons between trained conditions.

Summary

Overall, both latencies and errors showed the same pattern of results. There was no main effect of day, a significant main effect of conditions, and no interaction between the two conditions. Pairwise comparisons revealed that all trained conditions (RS, PS, and NS) showed faster RTs and lower error rates than the untrained condition (NT), but no differences between the trained conditions emerged.

Semantic categorization

Analyses included RTs and errors to the semantic categorization of newly learned words. As in previous tasks, participants were tested on day 3 and day 8, but only the conditions in which participants learned words with meaning were included in this task. On day 3, a total number of 420 responses were collected, but 60 (14.3%) were deleted from the analysis. These included participants' errors, which reached 50 (11.9%), and 10 (2.4%) outliers. On day 8, the same 420 responses were recorded, of which 57 (13.6%) were discarded from the analysis. They included 43 (10.2%) errors and 14 (3.3%) outliers.

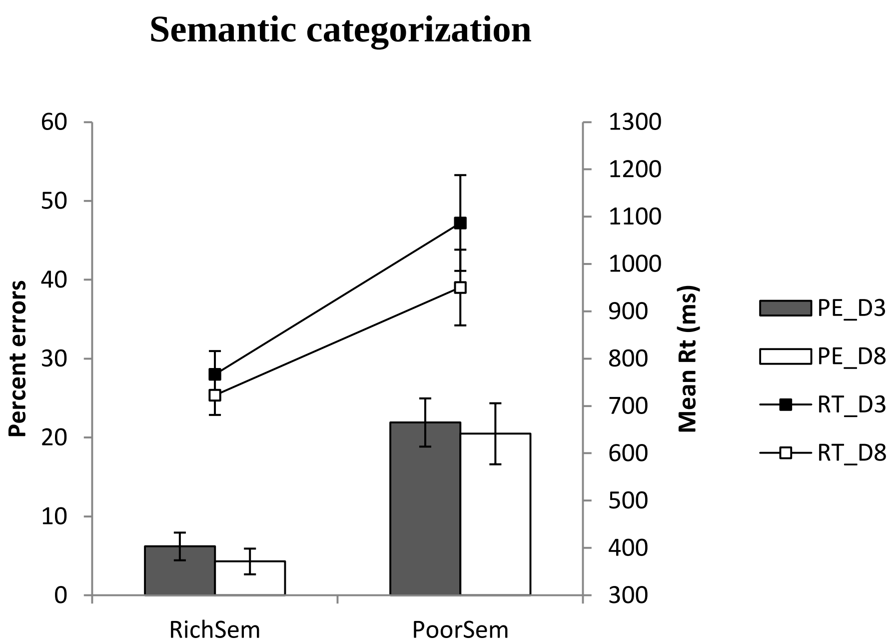


Figure 4.4. Percent errors (PE) and reaction times (RT) in the semantic categorization task performed on day 3 (D3) and day8 (D8). Error bars represent standard error (SE) of the mean.

RTs

A factorial repeated measures ANOVA was first conducted on RTs for correct responses. There was a significant effect of day, $F_1(1, 20) = 4.23$, MSE = 40498.41, $p = .05$, $\eta_p^2 = .17$; $F_2(1, 29) = 9.87$, MSE = 19493.59, $p = .01$, $\eta_p^2 = .25$. There was a highly significant effect of conditions, $F_1(1, 20) = 18.88$, MSE = 83268.10, $p < .001$, $\eta_p^2 = .49$; $F_2(1, 29) = 37.73$, MSE = 64200.73, $p < .001$, $\eta_p^2 =$

.52. No interaction between day and conditions was found, $F_1(1, 20) = 2.90$, MSE = 15426.76, $p = .10$, $\eta_p^2 = .13$; $F_2(1, 29) = 2.28$, MSE = 24169.04, $p = .14$, $\eta_p^2 = .07$.

Errors

A factorial repeated-measures ANOVA was also run on errors with day and conditions as factors. Results showed no effect of day, $F_1(1, 20) = 1.00$, MSE = 58.33, $p = .33$, $\eta_p^2 = .05$; $F_2(1, 29) = .60$, MSE = 136.07, $p = 45$, $\eta_p^2 = .02$. There was a significant main of conditions, $F_1(1, 20) = 19.86$, MSE = 269.05, $p < .001$, $\eta_p^2 = .50$; $F_2(1, 29) = 30.01$, MSE = 272.68, $p < .001$, $\eta_p^2 = .51$. There was no interaction between day and conditions, $F_1(1, 20) = .02$, MSE = 66.19, $p = .89$, $\eta_p^2 = .00$; $F_2(1, 29) = .02$, MSE = 76.08, $p = .89$, $\eta_p^2 = .00$.

Summary

The RT data showed a significant effect of day with better performance on day 8 than on day 3, and a significant effect of conditions with faster RTs for RS than PS. The error data did not show effect of day, but the effect of conditions showed the same pattern as in the RT data.

Word production

A total number of 420 responses were collected both on Day 3 and Day 8. The number of correct letters in each word was used as a measure to compare performance across days and conditions. Since the first letter was provided, the maximum number of correct letters per word was 7, so the total number of possible correct letters was 2940. On Day 3, participants produced a total number of 2173 (73.9%) correct letters. On Day 8, there was an overall improvement and production reached 2470 (84%) letters. Only accuracy of response was recorded for this task.

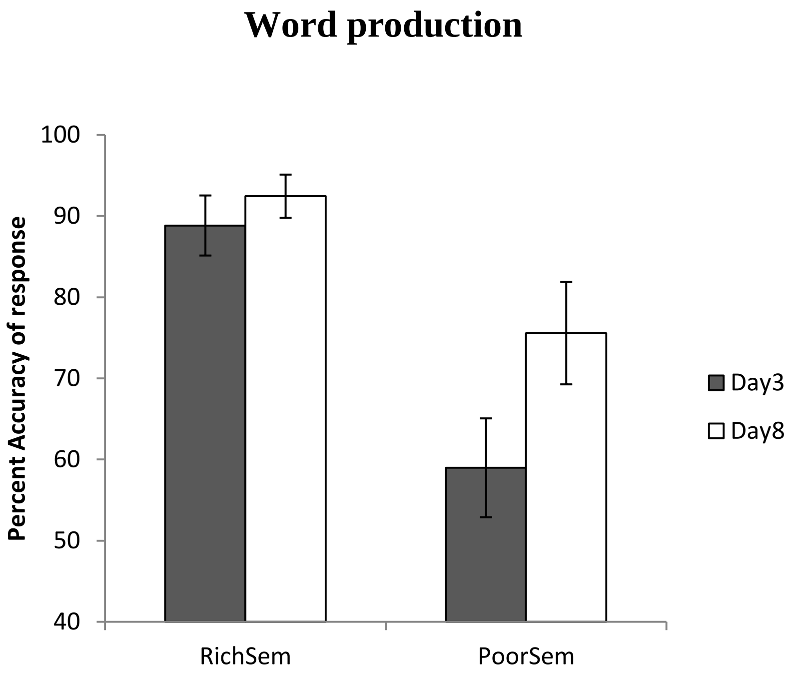


Figure 4.5. Percent accuracy of response for rich semantics (RichSem) and poor semantics (PoorSem) on day 3 and day 8 in the production task. Error bars represent standard error (SE) of the mean.

Percent accuracy

A two-way factorial ANOVA was first conducted on percent accuracy of response revealing a significant main effect of day, $F_1(1, 20) = 12.25$, MSE = 176.43, $p = .01$, $\eta_p^2 = .38$; $F_2(1, 29) = 15.21$, MSE = 202.53, $p < .001$, $\eta_p^2 = .34$. The effect of conditions was also highly significant, $F_1(1, 20) = 23.42$, MSE = 490.10, $p < .001$, $\eta_p^2 = .54$; $F_2(1, 29) = 37.69$, MSE = 435.87, $p < .001$, $\eta_p^2 = .57$. An interaction between day and conditions was also found, $F_1(1, 20) = 9.00$, MSE = 101.06, $p = .01$, $\eta_p^2 = .31$; $F_2(1, 29) = 6.17$, MSE = 209.84, $p = .02$, $\eta_p^2 = .18$. Bonferroni corrected paired-samples t-tests ($\alpha = .05$) on the data collected for each day revealed a significant difference between the conditions on both days with higher accuracy for RS than PS. The Bonferroni corrected t-tests ($\alpha = .05$) on each condition across days showed no significant difference between RS (day3) and RS (day 8), but a reliable difference between PS (day 3) and PS (day8), with higher accuracy on day 8.

Summary

The results showed that participants performed better on day 8 than on day 3 in both conditions. The RS condition showed better performance than PS on both days. The PS condition exhibited highly significant improvement over time, while RS showed a ceiling effect on day 3 and consequently no improvement over time.

4.2.3 Discussion

Predictions in this experiment were made for effects of semantic richness and consolidation/retention over time. Effects of semantic richness were expected to be replicated in recognition memory, semantic decision, and word production with better performance associated with richer meaning. No clear predictions were made regarding word naming since no effects of semantics were found in Experiment 3 and previous word learning experiments have also failed to find a semantic effect on reading aloud (e.g., McKay et al., 2008; McKague et al., 2001). Across all tasks, better consolidation over the course of a week was expected for novel words learned with rich semantics than for novel words learned with poor or no semantics. This prediction was made based on the levels-of-processing framework (e.g., Craik & Lockhart, 1972; Craik, 2002) which postulates that retention is relative to the degree of semantic processing during encoding, with deeper semantic analysis associated with better retention. The results of the current experiment are first discussed separately for each task, except for semantic categorization and word production which are presented together.

4.2.3.1 Recognition memory

In the recognition memory task, the expected overall effect of semantic richness was found in RTs and errors, but the interaction showed that this was only significant on day 8 for the RT data, with only a trend on day 3. This result is partially consistent with the results of Experiment 3 because the effect of semantics here only emerged on day 8 and not on day 3 as in Experiment 3. There were a few differences in the methodology used in this experiment with respect to previous experiments, which can explain the lack of effect on day 3. For instance, sentences were shorter so they did not allow as much contextual information as in Experiment

3, and the name of the category (taxonomic feature) was presented in both semantic conditions. This might have promoted the inference of features that were not explicitly presented in the sentences, which could have diminished the distinction between rich and poor semantics. Regarding the no semantics condition, it might be that participants spent more time reading the target words than in the semantic conditions due to the fact that the other words in the sentences were nonwords (see Appendix 4.7). The increased shift of attention towards the target words in this condition might have compensated for the lack of semantics and produced similar performance in comparison with the semantic conditions. However, over the course of a week, performance seems to have decayed or stayed the same for the NS condition, which might be due to the shallow encoding of novel words without meaning. Planned comparisons on RTs for the data collected on day 8 showed that the difference was only significant for each of the semantic conditions (RS, PS) versus the no semantics condition (NS) with faster RTs for the semantic conditions. These results suggest that the effect of semantics seems to emerge with the passing of time, but only reaches significance in comparisons between semantic and nonsemantic conditions. These findings are not consistent with Experiment 3 since they did not show a difference in performance between RS and PS. As explained earlier, there were differences in the methodology that could have produced the lack of advantage for RS, particularly the inclusion of taxonomic features in both conditions (e.g., *is an animal*) that could have caused automatic inference of extra features that were not controlled. Overall, the data are consistent with previous experiments regarding the effect of semantics on recognition memory, but failed to show a processing distinction between rich and poor semantics.

Regarding the effect of day, there was no difference in performance in error rates, but RTs were much faster on day 8 than on day 3. Planned comparisons across days for each condition revealed that only RS showed significantly faster RTs on day 8 than on day 3 whereas PS and NS did not show improvement over time. These results fit in well with the predictions for this experiment and support previous findings showing improvement in performance over time in explicit tasks (e.g, Dumay et al., 2004; Gaskell & Dumay., 2003). The fact that improvement was only significant in RS is very striking and is in line with the levels-of-processing framework because it suggests that words learned with rich semantics might consolidate better over time due to deeper semantic analysis during encoding (e.g., Craik & Lockhart, 1972; Craik, 2002; Craik, Moscovitch, & McDowd, 1994). This finding is further supported by a recent combined behavioural and fMRI study conducted by Fliessbach et al. (2010). They tested participants using 2 semantic tasks and 1 nonsemantic task. The semantic tasks included size comparison (whether the word represented a concept that was bigger or smaller than a shoebox) and animacy decision (whether the concept was a living or a nonliving thing). In the nonsemantic task, the decision included alphabetical or nonalphabetical order of the first and the last letter of each word. Behaviourally, they found better recognition in an old/new judgment task for the words previously presented in the semantic tasks in comparison with the nonsemantic task. At a neural level, results showed that words previously presented in a semantic task elicited more activation in the left anterior fusiform gyrus (an area involved in object processing) than words presented in the nonsemantic task. Fliessbach et al.'s interpretation of the findings was that the presentation of words in the semantic tasks led to the processing of objects features, which induced additional activation of object-processing brain areas contributing to enhanced memory encoding. In the current experiment, words were presented with visual and verbal information in the semantic conditions (RS, PS) and with no explicit semantic information in the nonsemantic condition (NS). The lack of improvement over time in NS might be attributed to lack of semantic processing during encoding, which is likely to produce shallow learning that is more prone to be forgotten with the passing of time (e.g., Craik & Tulving, 1975). Despite the fact that novel words learned in PS underwent semantic processing, this processing was likely to be superficial in comparison with RS due to the difference in the number of semantic features associated with novel words in each condition. In RS, words were encoded with more visual and verbal features than in PS, so the enhanced semantic elaboration in RS during learning might explain why RS showed significant improvement over time while PS only a trend.

4.2.3.2 Word naming

In the naming task, results were very consistent across days showing no improvement over time and no differences between any of the trained conditions. These results were highly consistent with Experiment 3. First, the lack of semantic involvement in reading aloud newly learned words with regular spelling is not surprising since this effect has been demonstrated earlier (e.g., McKay, 2008; McKague, 2001). Furthermore, in studies with familiar words, semantic involvement in reading aloud generally shows reliable effects for low-frequency words (e.g., Hino & Lupker, 1996) or words with irregular spelling (Rodd, 2004) but not for regular words. This has been further assessed in languages with more transparent orthography than English such as Spanish where effects of semantic variables have not showed any contribution to reading (e.g., Alija & Cuetos, 2006) unlike other variables such as frequency and age of acquisition. Overall, previous evidence seems to suggest that semantics only contributes to reading when a direct mapping between orthography and phonology cannot be achieved automatically. Thus, its role seems to contribute to disambiguation when the reader has to re-evaluate the correct pronunciation of a given word. Second, the lack of improvement over time in this task might have been due to a ceiling effect. Results showed very good performance on day 3 with overall error rates below 5% and RTs of around 550 ms in all conditions, except for the untrained condition.

4.2.3.3 Semantic categorization and word production

Results for semantic categorization and word production were very similar and largely fit the predictions for the current experiment. As in the semantic decision task and the cued recall task of Experiment 2 and Experiment 3, RS showed an overall advantage in comparison with PS. The inclusion of more semantic features through verbal (sentences) and nonverbal (pictures) information in RS might have produced a more decontextualised acquisition of meaning. Regarding semantic categorization, this is in line with the evidence presented in the discussion of Experiment 2, which suggested that an advantage for familiar words with high number of features is normally found in semantic tasks (e.g., Pexman et al., 2008; Grondin et al., 2006; Pexman et al., 2003; Pexman et al., 2002). In the word production task, the same explanation provided in Experiment 2 for the cued recall task can fit the results here, even though the tasks differed slightly. Learning a novel word with many semantic features (RS) can allow more accurate mapping between the newly learned word and its correspondent category, which could elicit more accurately responses in the word production task.

Regarding the effect of day, Improvement over time was found in both tasks with better performance on day 8. In semantic categorization both conditions improved equally over time while in the production task, there was more improvement in PS than RS. This final result is against predictions since it was expected that improvement over time would be modulated by semantic richness, with more improvement associated with more semantics. As in the naming task, this lack of significant improvement in RS was probably due to a ceiling effect since accuracy reached almost 90% on day 3 and slightly over 90% on day 8. On the contrary, in PS accuracy was only around 60% and improved over 70% on day 8. Certainly, there was more room for improvement in PS than in RS, which might explain the differential pattern of improvement for the conditions.

4.2.3.4 Summary and conclusion

The results presented above showed and overall effect of semantic richness on semantic categorization and word production. Regarding recognition memory, this effect was only found between the semantic and the nonsemantic conditions, but no differences between rich and poor semantic conditions emerged. The word naming task did not show any difference across trained conditions. These findings showed that learning conditions can have differential effects across different tasks. The amount of semantic information acquired during learning can clearly affect semantic categorization and word production, but has less of an impact on recognition memory and no effects at all on word naming. Assessment of performance over time also showed different patterns depending on the task. In semantic categorization, there was equal improvement over time in both conditions whereas in word production, improvement was only significant for the poor semantics condition, most likely due to a ceiling effect in the rich semantics condition on day 3. The naming task did not show any difference over time whereas performance for word recognition showed improvement only in the rich semantics condition.

4.3 Experiment 5

The aim of Experiment 5 was to further explore the two conditions in which words were presented with meaning in the previous experiment. In the recognition memory task, an advantage for words with semantics (either poor or rich) over words without semantics was only found on day 8. The difference between RS and PS did not reach significance, but showed a clear trend towards better performance for RS over PS, especially on day 8. Another important finding was that RS showed significant improvement in performance from day 2 to day 8 whereas PS did not. These two findings suggest that a difference between the two semantic conditions might be found and with a stronger effect on day 8 if the manipulation of semantic richness is more tightly controlled. In the discussion of Experiment 4, it was suggested that the presentation of taxonomic features or category label in both conditions could have produced activation of additional semantic features that were not controlled. For instance, if a participant in a word learning experiment reads the following sentence: *an adertmon is an animal*; semantic features that are common to many animals might be automatically activated, so more semantic information can be inferred, which can contaminate the manipulation of semantic richness. In a study conducted by McRae et al. (2005), which aimed at collecting semantic features for over 500 living (*cat*) and nonliving (*table*) concepts, they calculated the overall number of semantic features with and without taxonomic features. However, for other variables derived from *number of semantic features* such as *feature distinctiveness* and *cue validity*, they excluded taxonomic features. They argued that taxonomic features are different from other types of features such as those referring to *parts* or *function* and might target a different type of information. Given the above, in Experiment 5 taxonomic features were excluded from sentences in the poor semantics (PS) condition in order to avoid conveying extra features that could not be controlled. In the rich semantics (RS) condition, only taxonomic features that did not correspond to the actual category label were included. See Appendix 4.9 for sample sentences. Regarding the use of visual stimuli, these were the same used in Experiment 4 for the semantic conditions including standard resolution images in RS and pixalated images in PS. See Appendix 4.4 for sample images in RS and PS. The main objective of the new manipulation was to create a bigger distance in semantic richness between the conditions so that the effect of semantic richness could be observed more clearly.

Apart from the recognition memory task, the current Experiment also incorporated a semantic categorization and a production task. The semantic categorization task was exactly the same as in Experiment 4, but the production task was slightly modified. In Experiment 4, participants were provided with a definition and they were asked to elicit the newly learned word that corresponded to that definition. Since the definition of the words included all types of semantic features, this might have biased the responses towards better performance in RS than in PS. In order to avoid this confound, in the current experiment participants were presented with the category label and the first two letters of the newly learned word, instead of a definition. Since the category name was not presented in the training session, the new design does not bias responses for any of the conditions and should more accurately assess performance.

Predictions in the new study are similar to those of the previous study. First, an advantage for RS over PS is expected in the recognition memory task, particularly on day 8 since the passing of time seems to favour memories that have been stored with more elaborate semantic processing (e.g., Craik & Lockhart, 1972; Craik, 2002). Second, based on the findings of Experiment 4, an improvement over time is expected in both conditions but it might be higher for RS due to more elaborate semantic processing, which contributes to better consolidation of the new memories over time (e.g., Fliessbach et al., 2010). Regarding the semantic categorization and production tasks, a close replication of the effects found in Experiment 4 is expected, that is, better performance in RS than PS on both days and improvement over time for both conditions.

4.3.1 Method

Participants

Twenty-two native English speakers (16 female; mean age 20.5, SD 4.0) recruited from the University of York community took part in the experiment. All individuals had normal or corrected-to-normal vision and had never been diagnosed with any language problems.

Materials and design

Twenty nonwords from Experiment 4 (set A and set B) were used in the new experiment (see Appendix 4.1). Both sets of nonwords were assigned meaning, which corresponded to the same concepts used in Experiment 4 (See Appendix 4.2). As in Experiment 4, the original written forms of these nouns were replaced with the nonwords in set A and set B. Two-hundred and eighty sentences were used as linguistic contexts during the training sessions. These sentences were taken from Experiment 4, but the first sentence from each set was replaced because it contained the name of the category or superordinate feature (e.g., *animal*, *plant*, etc.). The new sentence only contained a general feature (e.g., *has legs*, *has eyes*, etc). Thus, in Experiment 5 none of the sentences included the category label, so participants were required to infer the category the novel words belonged to, which made the task harder and allowed more control of the semantic information conveyed in each condition. The rest of the sentences were the same as the ones used in Experiment 4 as well as the images. Filler items used in recognition memory, and category labels used in the semantic categorization, and production tasks were all taken from Experiment 4.

Procedure

The procedure was also similar to that of Experiment 4 with some changes regarding the time of training and the tasks used for testing. Like Experiment 4, the current experiment took place over a week, but had only one training session, instead of two. The training session took place on day 1 and consisted of two main parts, which combined lasted approximately 90 minutes. On day 2, participants were required to complete 3 tasks including recognition memory, semantic categorization, and word production. On day 8, participants were asked to return for the final testing session, which was the same as on day 2.

Training procedure

The training procedure was exactly the same as in Experiment 4, except that there were only two conditions (RS and PS) and both sessions took place on day 1 with a 10-minute break between sessions.

Testing procedure

Testing in Experiment 5 included recognition memory, semantic categorization, and word production. The recognition memory task was exactly the same as that of Experiment 5, except that participants were tested on two conditions (RS and PS), instead of 3. The semantic categorization was exactly the same as in Experiment 4. Unlike Experiment 4, the production task in the current experiment did not present participants with definitions but with the category label the word belonged to (e.g., *animal*, *plant*, *weapon*, etc.). Another difference was that the first two letters were included as orthographic cues, instead of just the first letter. The rest of the procedure was exactly the same as in Experiment 4.

4.3.2 Results

Analyses were performed on all 22 participants. The same procedure used in Experiment 4 regarding outliers was used in the current experiment.

Table 4.3. RTs for recognition memory and semantic categorization; percent accuracy of response for word production on day 2 and day 8.

	Day 2		Day 8	
	RS	PS	RS	PS
	Recognition memory			
Mean RT	759	748	691	702
SD	175	178	131	133
% errors	9.1	9.6	4.6	5.5
	Semantic categorization			
Mean RT	1066	1473	960	1264
SD	346	559	342	427
% errors	8.14	30.91	7.7	29.5
	Word production			
% accuracy	70	60	84	76
SD	21	22	15	19

Note. RS, rich semantics; PS, poor semantics.

Recognition memory

Participants were given exactly the same test on day 2 and day 8. A total of 440 responses were collected on each occasion. On day 2, 41 (9.3%) RTs were removed from the analysis, of which 27 (6.1%) corresponded to errors, while the remaining 14 (3.2%) corresponded to RTs beyond the range of 2.5 SD from the mean. On day 8, a total number of 23 (5.3%) responses were eliminated, 7 (1.6%) were errors and 16 (3.7%) corresponded to RTs over 2.5 SD from the mean.

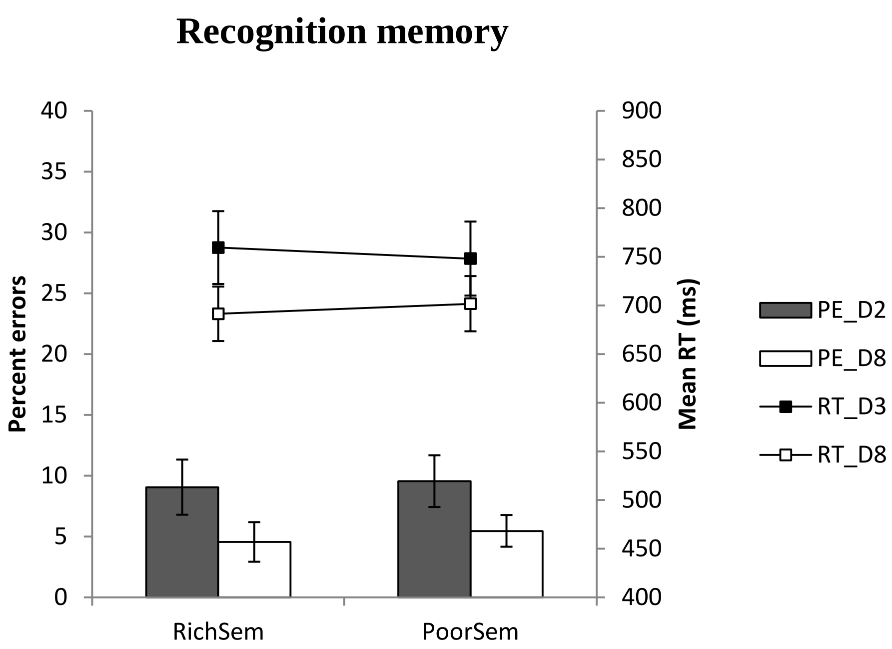


Figure 4.6. Percent errors and RTs in the recognition memory task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

RTs

A two-way repeated measures ANOVA was first conducted on RTs for correct responses revealing a significant main effect of day, $F_1(1, 21) = 9.64$, MSE = 7489.49, $p = .01$, $\eta_p^2 = .32$; $F_2(1, 19) = 61.79$, MSE = 1124.84, $p < .001$, $\eta_p^2 = .77$, with better performance on day 8 than on day 3. There was no significant effect of conditions, $F_1(1, 21) = .00$, MSE = 3211.64 $p = .97$, $\eta_p^2 = .00$; $F_2(1, 19) = .14$, MSE = 4835.25, $p = .72$, $\eta_p^2 = .01$. The interaction between day and conditions was also not significant, $F_1(1, 21) = 1.65$, MSE = 1570.36 $p = .21$, $\eta_p^2 = .07$; $F_2(1, 19) = .84$, MSE = 2247.17, $p = .37$, $\eta_p^2 = .04$.

Errors

A factorial repeated measures ANOVA was also conducted on errors with day and semantic conditions as main factors. Results showed a main effect of day, $F_1(1, 21) = 6.47$, MSE = 62.80 $p = .02$, $\eta_p^2 = .24$; $F_2(1, 19) = 7.96$, MSE = 39.43, $p = .01$, $\eta_p^2 = .30$. There was no significant main effect of conditions, $F_1(1, 21) = .13$, MSE = 81.20 $p = .72$, $\eta_p^2 = .01$; $F_2(1, 19) = .30$, MSE = 26.11, $p = .59$, $\eta_p^2 = .02$. The interaction between day and conditions was also not significant, $F_1(1, 21) = .01$, MSE = .72.58 $p = .91$, $\eta_p^2 = .00$; $F_2(1, 19) = .02$, MSE = 44.86, $p = .89$, $\eta_p^2 = .00$.

Summary

Both RTs and errors showed an effect of day with better performance on day 8 than on day 3. No interaction or effect of conditions was found.

Semantic categorization

Analyses included RTs and errors to the semantic categorization of newly learned words. On day 2, a total number of 440 responses were collected, of which 75 (17%) were deleted from the analysis. These corresponded to 70 (15.9%) errors, 5 (1.1%) outliers. On day 8, the same number of responses were recorded and 73 (16.6%) were discarded from the analysis. Most of the RTs deleted corresponded to errors, reaching 72 (16.4%) with only 1 (0.2%) corresponded to outliers.

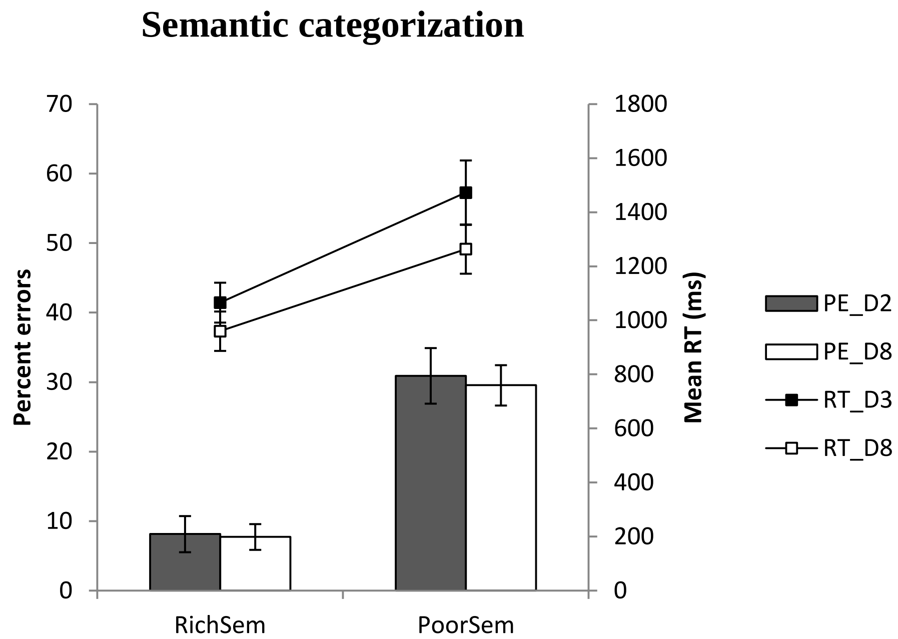


Figure 4.7. Percent errors and RTs in the semantic categorization task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

RTs

A factorial within subjects ANOVA was first conducted on RTs for correct responses. There was a significant effect of day, $F_1(1, 21) = 5.01$, MSE = 108686.16, $p = .04$, $\eta_p^2 = .19$; $F_2(1, 19) = 13.55$, MSE = 31693.60, $p = .01$, $\eta_p^2 = .42$, with faster RTs on day 8. There was a highly significant effect of conditions, $F_1(1, 21) = 36.46$, MSE = 76198.51, $p < .001$, $\eta_p^2 = .64$; $F_2(1, 19) = 49.87$, MSE = 52653.90, $p < .001$, $\eta_p^2 = .72$, with RS showing faster RTs than PS. No interaction between day and conditions was found, F_1 (1, 21) = 1.46, MSE = 39991, $p = .24$, $\eta_p^2 = .07$; $F_2(1, 19) = 1.13$, MSE = 37788.12, $p = .30$, $\eta_p^2 = .06$.

Errors

A factorial repeated measures ANOVA was also conducted on errors with day and conditions as factors. Results indicated no effect of day, F_1 (1, 21) = .12, MSE = 139.43, $p = .73$, $\eta_p^2 = .01$; F_2 (1, 19) = .03, MSE = 138.77, $p = .88$, $\eta_p^2 = .00$. There was a significant main effect of conditions with higher error rates in PS than RS, F_1 (1, 21) = 51.97, MSE = 210.44, $p < .001$, $\eta_p^2 = .71$; $F_2(1, 19) = 28.48$, MSE = 342.25, $p < .001$, $\eta_p^2 = .60$. There was no interaction between day and conditions,

$F_1(1, 21) = .10$, MSE = 49.54, $p = .75$, $\eta_p^2 = .01$; $F_2(1, 19) = .24$, MSE = 57.76, $p = .63$, $\eta_p^2 = .01$.

Summary

RTs analyses showed an overall effect of day with participants performing better on day 8 than on day 2. However, this effect was not found in the errors analyses. There was a significant effect of conditions with better performance in RS than PS in both RTs and error rates. No interaction between day and condition was found.

Word production

A total number of 440 responses were collected both on day 2 and day 8. The number of correct letters in each word was used as a measure to compare performance across days and conditions. Since the first two letters of each word were provided, the maximum number of correct letters per word was 6, so the total number of possible correct letters was 2640. On day 2, participants produced a total number of 1717 (65.0%) correct letters. On day 8, there was an overall improvement and participants' correct letters reached 2103 (79.7%).

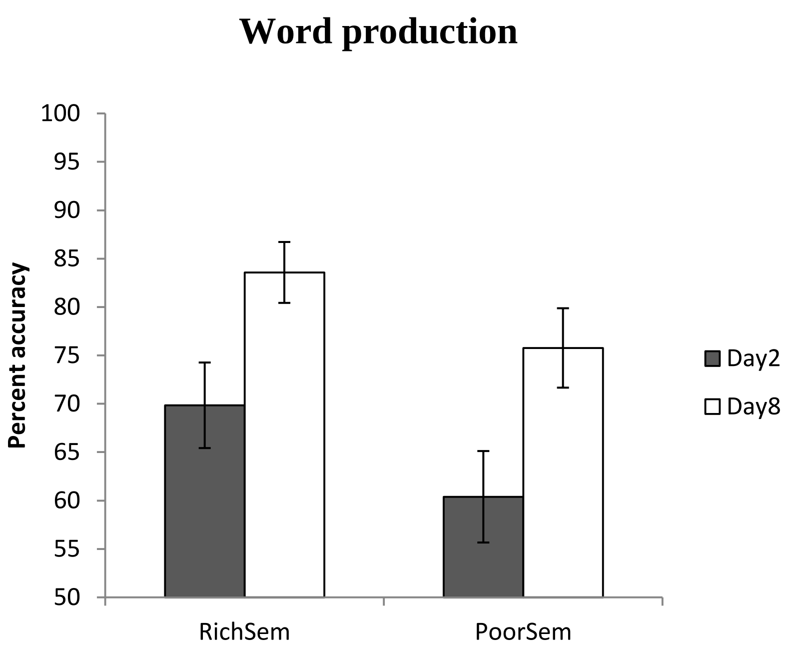


Figure 4.8. Percent accuracy of response in the production task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

Percent accuracy

A two-way factorial ANOVA was first conducted on accuracy of response revealing a significant main effect of day, $F_1(1, 21) = 56.28$, MSE = 82.19, $p < .001$, $\eta_p^2 = .73$; $F_2(1, 29) = 50.33$, MSE = .31, $p < .001$, $\eta_p^2 = .73$. The effect of conditions was also highly significant, $F_1(1, 21) = 7.20$, MSE = 227.98, $p = .01$ $\eta_p^2 = .54$; $F_2(1, 19) = 11.55$, MSE = .46, $p = .01$, $\eta_p^2 = .38$. There was no interaction between day and conditions, $F_1(1, 21) = .20$, MSE = 65.92, $p = .66$, $\eta_p^2 = .01$; $F_2(1, 19) = .08$, MSE = .51, $p = .78$, $\eta_p^2 = .00$.

Summary

The results showed that participants performed better on day 8 than on day 2 in both conditions. There was also an effect of conditions with RS showing a clear advantage over PS. The interaction between day and conditions was not significant.

4.3.3 Discussion

Experiment 5 looked at effects of semantic richness on recognition memory, semantic categorization, and word production. It also assessed performance over time to investigate the role of levels of semantic processing on the consolidation of newly learned words.

In the recognition memory task, predictions regarding an effect of semantic richness were not confirmed since no difference in performance was found between RS and PS. This was rather surprising since a clear difference was found in semantic categorization, and word production. The effect of day found in Experiment 4 was replicated but no interaction between days and conditions emerged, which suggests that improvement in performance over time in RS and PS did not differ significantly. This finding differs from that of Experiment 4 which showed a significant interaction with improvement in performance only affecting RS but not PS. In the current Experiment only a trend towards more improvement over time was found in RS. Thus, no clear evidence was collected to support the view that deeper semantic analysis during encoding produces better consolidation over time as suggested by the levels-of-processing literature (e.g., Craik, 2002; Fliessbach et al., 2010).

In the semantic categorization task, the results replicated those of Experiment 4, with better performance on day 8 than on day 2 and a clear advantage for RS over

PS on both days. Both conditions showed equal improvement over time. These results confirmed that semantic richness affects the categorization of newly learned words, which is consistent with previous evidence found in the manipulation of familiar words with high and low number of semantic features (e.g., Pexman et al., 2008; Grondin et al., 2006; Pexman et al., 2003; Pexman et al., 2002). As stated in Chapter 1, it has been suggested that words with high number of semantic features produce faster meaning activation and consequently faster semantic settling, which accelerates semantic categorization responses (e.g., Pexman et al., 2008).

In the production task, results were largely consistent with those of Experiment 4. However, overall performance was much lower due to increased difficulty of the task. Participants were required to recall a word based on the category name and orthographic cues, but there was no definition involved as in Experiment 4. Both RS and PS showed similar improvement over time and there was a significant difference between the two conditions with better accuracy for RS. These results are consistent with Experiment 3 and Experiment 4 and confirmed that participants learned better when exposed to more semantic features during training. Additionally, they showed that performance over time improved equally for RS and PS, which suggests that the advantage for RS over PS emerges early and remains stable with the passing of time.

In summary, the results of Experiment 5 showed significant improvement in performance over time affecting both conditions in a similar fashion since no interaction was found in any of the tasks. Effects of semantic richness were found in semantic categorization and production, but did not reach significance in recognition memory.

4.4 General discussion

The main purpose of Chapter 4 was to investigate the time course of word learning and examine effects of semantic richness at two different time points.

First, the analyses of the two experiments presented above have provided evidence that performance increases over time as measured by explicit tasks including recognition memory, semantic categorization, and word production. This is consistent with previous findings in word learning studies where this effect has been demonstrated using implicit and explicit tasks (e.g., Davis et al., 2008; Gaskell &

Dumay, 2003; Dumay et al., 2004; Dumay & Gaskell, 2007). In the study conducted by Davis et al., performance on explicit tasks such as repetition, recognition memory, and meaning rating was better one day after training than on the same day of training. They also showed a lexical competition effect; that is, the recognition of familiar words was affected by similar-sounding newly learned words. The fact that this effect only emerged one day after training and not on the same day was interpreted as a proof that the integration of newly learned words in the mental lexicon is mediated by sleep. In the study conducted by Gaskell and Dumay recognition of newly learned words increased significantly from day 2 to day 5 and lexical competition effects appeared only on the third day. This suggests that both explicit and implicit memories take time to consolidate as better performance is observed with the passing of time. The current experiments replicated this effect and extended the findings to semantic categorization and word production. Since participants in most previous studies and in the current experiments have been retested on the same sets of words, it is not certain whether the improvement over time is only due to the passing of time or to the fact that the same words were presented in subsequent testing sessions. This is further examined in Chapter 5.

An interesting finding concerning improvement over time was the fact that words learned with rich semantics showed significant improvement over time in the recognition memory task whereas words with poor and no semantics did not (Experiment 4). As stated earlier, this seems to suggest that novel words that are associated with rich meaning (many semantic features) developed more stable memory traces over time in comparison with words with poor or no meaning. This finding is consistent with the levels-of-processing literature (e.g., Craik & Lockhart, 1972; Brown & Craik, 2000), which suggests that retention is a function of the depth in semantic processing during encoding. In Experiment 4, participants associated novel words with many features in RS, few features in PS, and no features in the meaningless condition. This implies that more semantic information was processed in RS than PS or NS, which involved higher levels of semantic analyses during the encoding of the RS novel words. This might have produced better consolidation of RS words over time than in the other two conditions, which was reflected in improved performance in recognition memory. A main concern for this proposal is that Experiment 5 did not replicate the finding in Experiment 4. Novel words with rich and poor semantics showed similar levels of consolidation over time, with only a trend towards better consolidation in RS. It might be that the effect only holds for comparisons of semantic conditions versus nonsemantic conditions as in the study conducted by Fliessbach et al. (2010) which showed better recognition for words that underwent semantic encoding in comparison with words that were only encoded perceptually. Thus, the findings of both experiments in this chapter are not necessarily inconsistent with the levels-of-processing framework (e.g., Craik & Lockhart; Craik et al., 1994; Craik, 2002).

Regarding semantic richness, the effect was very consistent across experiments for semantic categorization and word production, which replicated the findings of Experiment 2 and Experiment 3. This suggests that exposing participants to novel words with high number of features (rich semantics) produces an advantage in performance with respect to low number of features (poor semantics) in semantic tasks. Increased semantic activation in RS allows faster and more accurate responses in semantic categorization (e.g., Pexman et al., 2003). As for word production, the acquisition of more semantic features in RS allows more accurate matching between the definition (Experiment 4) or the category (Experiment 5) and the newly learned word to be elicited. In the recognition memory task, a reliable significant effect was only found for semantic conditions versus meaningless conditions. This suggests that a substantial difference in semantic richness is needed for effects to be reflected in recognition memory.

In summary, the experiments in the current chapter suggests that newly learned words consolidate over time as measured in recognition memory, semantic categorization, and word production. Particularly in the recognition memory tasks, results showed that improvement over time might be a function of the amount of meaning acquired during training, with more meaning producing more improvement. Finally, semantic richness seems to affect all tasks with very consistent effects in semantic categorization and word production, and much weaker effects in recognition memory.

Chapter 5 – Effect of testing on the consolidation of newly learned words

5.1 Introduction

The experiments presented in Chapter 4 found that participants' performance improved over time in recognition memory, semantic categorization, and word production. A possible confound in these experiments was that the same words were tested on day 3 and day 8, so the improvement might have been due to the additional exposure provided by the first test on day 3, and not simply to the passing of time. In fact, it has been suggested that taking a memory test is not only an instance of assessing one's knowledge, but also the opportunity to enhance the retention of new material (e.g., McDaniel & Manson, 1985; Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003; Roediger & Karpicke, 2006). It is also important to mention that the effect of testing is not seen immediately after learning, but seems to emerge in the long term. In the study conducted by Roediger and Karpicke, three groups of participants were required to study 2 passages. One of the passages was studied twice while the other one was studied only once and then tested using a free recall test. One group of participants was given a free recall test 5 minutes after the learning session, another group took the test 2 days later, and the last group was examined after one week had passed. Results showed that participants tested immediately after the training session showed better recall for the passage they studied twice as compared to the passage they studied once and then tested. However, the effect was reversed for participants who were tested 2 days or a week later showing better recall for the study-plus-test passage than for the study-plus-study passage. In a follow-up experiment, Roediger and Karpicke found that performance a week after training was even higher if more free recall tests were given during the learning session. This evidence confirms that tests do not only fulfil the purpose of assessing studied material, but also help to increase long-term retention.

In order to examine the possible effect of testing in the experiments presented in Chapter 4, two more experiments are presented in the current chapter. Participants in both experiments learned novel words under the same learning conditions and were tested using the same tasks: recognition memory, semantic categorization, and word production. In Experiment 6 they were tested on all stimuli on day 2 and then retested on the same stimuli on day 8, whereas in Experiment 7 participants were tested on half of the words on day 2 and the other half on day 8.

5.2 Experiment 6

Experiment 6 mainly aims to replicate the findings of previous experiments regarding improvement over time when participants are tested on the same stimuli a week after training. It also examines differences in improvement over time depending on whether novel words are trained with semantics or without semantics. Regarding effects of semantics, it attempts to replicate the advantage for words learned with semantics over words without semantics. The tasks used in this experiment included recognition memory, semantic categorization, and word production. In order to increase the number of items per condition and to have more statistical power, conditions were reduced to only two: a condition with rich semantics (meaningful condition) and a condition with no semantics (meaningless condition). Specific predictions were made for each task. In the recognition memory task, participants were expected to show improvement over time in both conditions either because of consolidation over time or simply because of the effect of test given on day 2. However, performance should show further improvement over time in the meaningful condition consistent with the levels-of-processing literature suggesting better retention of material that undergoes deep semantic analysis (Craik & Lockhart, 1972; Brown & Craik, 2000). The effect of semantics, which only showed a trend in the first test of Experiment 4 but reached significance on the second test, should now show a significant effect on day 2 due to more statistical power - the contrast now includes the double number of words per condition. The semantic effect should be consistent across days with further improvement on day 8, as suggested earlier. The incorporation of the semantic categorization and the word production tasks has the unique purpose of assessing improvement over time since only one semantic condition was used in this experiment. Consistent with previous studies, performance for both tasks is expected to increase significantly over time. Whether or not this is simply due to the passing of time or the effect of testing will be further assessed in the following experiment.

5.2.1 Method

Participants

Twenty English native speakers (14 female; mean age 20.4, SD 2.5) enrolled at the University of York participated in the study. Individuals had normal or corrected-to-normal vision and had no history of any language disorders.

Materials and design

The stimuli used in Experiment 6 were mostly taken from previous experiments. The same 40 nonwords used in Experiment 4 were used again in Experiment 6 (see Appendix 4.1). Half of these words were learned with rich semantics (meaningful condition) and the other half with no semantics (meaningless condition). The meaning corresponded to the obscure nouns used in Experiment 4 (See Appendix 4.2). As in Experiment 4, the original name of the nouns was replaced with a nonword from Appendix 4.1. Sentences corresponding to rich semantics and no semantics conditions in Experiment 4 were used in the current experiment (see Appendix 4.6 for sample sentences). All visual stimuli used in Experiment 4 for rich semantics and no semantics were also used in the current experiment (see Appendix 4.4 for sample visual stimuli). Forty filler items (30 taken from Experiment 4) were required for the recognition memory task (see Appendix 4.8 for full list). The same categories used in Experiment 4 were used again here in the semantic categorization task and the production task (see Appendix 4.6).

Procedure

The experiment took place over a week with training on day 1 and testing on day 2 and day 8. Participants received training in all novel words and were tested on the full list of words on day 2 and then again on day 8.

Training procedure

The training procedure was similar to that of Experiment 5 including only one long training session on day 1. However, unlike Experiment 5, it included 20 words with rich semantics and 20 words with no semantics and there was no poor semantics condition.

Testing procedure

The testing session was also very similar to that of Experiment 5 and included recognition memory, semantic categorization, and production. The only difference was the conditions.

5.2.2 Results

Analyses were performed on all 20 participants. The procedure regarding data filtering was the same as in Experiment 4 and Experiment 5. Also the same criteria for statistical tests used in previous experiments were used here again.

Table 5.1. RTs corresponding to recognition memory and semantic categorization; percent accuracy of response for word production on day 2 and day 8.

	Day 2		**Day 8**	
	Meaningful	Meaningless	Meaningful	Meaningles
		Recognition memory		
Mean RT	861	894	812	912
SD	167	162	164	183
% errors	14.5	20.0	9.5	24.8
		Semantic categorization		
Mean RT	1220	---	1051	---
SD	438	---	325	---
% errors	19.3	---	17.3	---
		Word production		
% accuracy	48.8	---	61.0	---
SD	28	---	26	---

Recognition memory

Participants were tested on all words on day 2 and day 8. A total of 800 responses were collected on each day. On day 2, 138 (17.3%) of the RTs were deleted from the analyses. They included 119 (14.9%) errors and 19 (2.4%) outliers. On day 8, 139 (17.4%) RTs were removed and included 123 (15.4%) errors and 16 (2.0%) outliers.

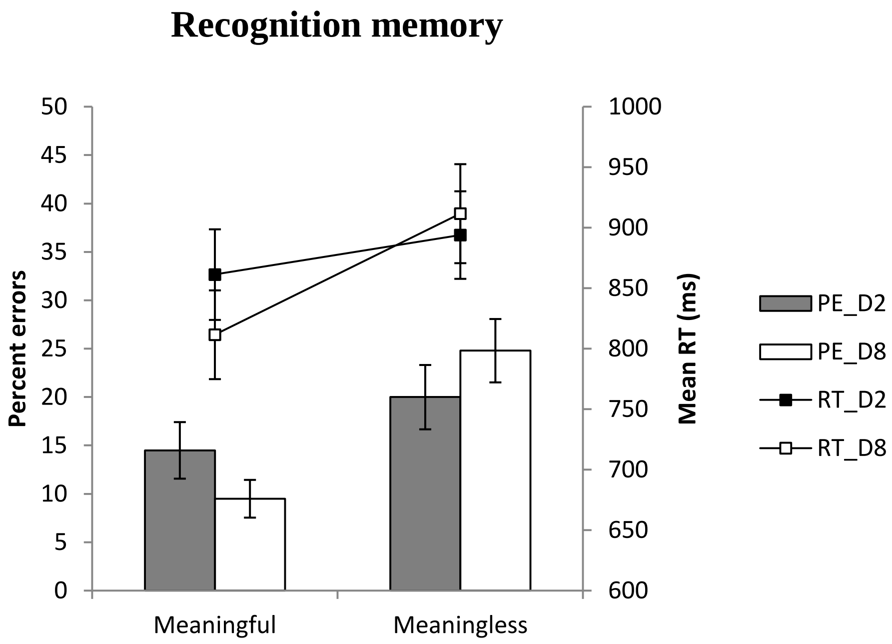


Figure 5.1. Percent errors and RTs in the recognition memory task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

RTs

A two-way repeated measures ANOVA was conducted on RTs for correct responses and revealed no effect of day, $F_1(1, 19) = .17$, MSE = 12209.50, $p = .69$, $\eta_p^2 = .01$; $F_2(1, 39) = 2.85$, MSE = 4452.43, $p = .10$, $\eta_p^2 = .07$. There was a significant effect of conditions, $F_1(1, 19) = 21.52$, MSE = 4615.22, $p < .001$, $\eta_p^2 = .53$; $F_2(1, 39) = 3.34$, MSE = 24.34, $p < .001$, $\eta_p^2 = .38$. The interaction between day and conditions was highly significant, $F_1(1, 19) = 10.00$, MSE = 2870.39 $p = .01$, $\eta_p^2 = .35$; $F_2(1, 39) = 10.59$, MSE = 6716.87, $p = .01$, $\eta_p^2 = .21$. Bonferroni corrected paired-samples t-tests ($\alpha = .05$) were also run on the data collected on day 2 and day 8 separately. The t-test on RTs collected on day 2 showed a marginal difference between the meaningful and the meaningless condition with faster RTs for the meaningful condition. The analysis on the data for day 8 showed a highly significant difference between the conditions, with faster RTs for the meaningful condition. Bonferroni corrected paired-samples t-tests ($\alpha = .05$) were also conducted per condition across days. Results in the meaningful condition did not reach significance, but showed a trend towards better performance on day 8. Results in the meaningless condition showed no significant difference between day 3 and day 8 with a trend in the opposite direction (better performance on day 3).

Errors

The two-way repeated measures ANOVA on errors showed no effect of day, $F_1(1, 19) = .01$, MSE = 1.802, $p = .93$, $\eta_p^2 = .00$; $F_2(1, 39) = .01$, MSE = .89, $p = .93$, $\eta_p^2 = .00$. There was a significant effect of conditions, $F_1(1, 19) = 22.37$, MSE = 3.85, $p < .001$, $\eta_p^2 = .26$; $F_2(1, 39) = 20.63$, MSE = 2.29, $p < .001$, $\eta_p^2 = .01$. The interaction between day and conditions was also significant, $F_1(1, 19) = 15.55$, MSE = 1.22 $p < .001$, $\eta_p^2 = .45$; $F_2(1, 39) = 20.31$, MSE = .57, $p < .001$, $\eta_p^2 = .34$. Bonferroni corrected t-tests ($\alpha = .05$) were also conducted for each separate day. The t-test on day 2 showed no significant difference between the two conditions, but day 8 analyses showed that the difference was highly significant, with lower error rates for the meaningful condition. Bonferroni corrected t-tests for each condition across days showed a significant difference between day 2 and day 8 regarding the meaningful condition with better performance on day 8. There was also a difference between days in the meaningless condition, but the effect was reverse (performance was better on day 2).

Summary

No overall effect of day was found in either RTs or errors (because the effect was different in different conditions). However, a main effect of conditions was found for both RTs and errors with better performance in the meaningful condition than in the meaningless condition. A significant interaction was also found in both RTs and errors. Pairwise comparisons conducted on RTs showed a marginal advantage for the meaningful condition on day 3. However, day 8 analyses showed significantly faster RTs for the meaningful condition. Error rates between the conditions did not differ on day 3, but were significantly lower for the meaningful condition on day 8. Pairwise comparisons on RTs for each condition across days showed that none of the conditions differed from day 2 to day 8, but they showed trends on opposite directions. The error data showed significantly lower error rates on day 8 for the meaningful condition, but significantly higher for the meaningless condition (reverse effect).

Semantic categorization

For obvious reasons, only the meaningful words were tested in this task. A total of 400 responses were recorded on each day. On day 2, 77 (19.3%) responses were removed from the analysis. These corresponded to 73 (18.3%) errors and 4 (1.0%) outliers. On day 8, 69 (17.3%) responses were eliminated, which included 63 (15.8%) errors and 6 (1.5%) outliers.

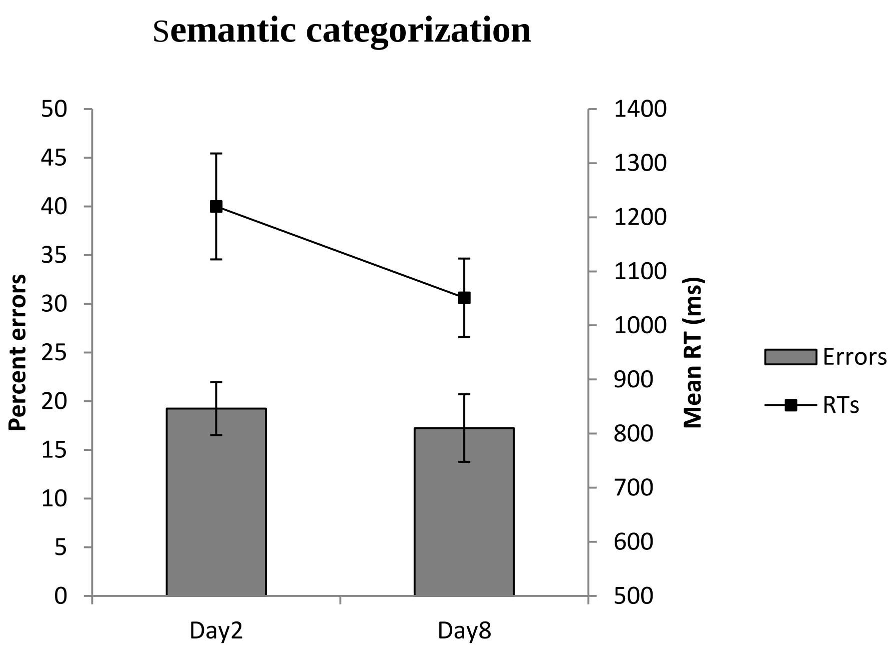


Figure 5.2. Percent errors and RTs in the semantic categorization task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

RTs

There was a significant difference between day 2 and day 8, with faster RTs on day 8, t_1 (19) = 2 .64, p = .02; t_2 (39) = 5.03, p < .001.

Errors

There was no difference between day 2 and day 8, t_1 (19) = .687, p = .50; t_2 (39) = .81, p = .42.

Summary

The data showed faster RTs for words learned with meaning in comparison with words learned without meaning. However, the error data did not show any difference between the conditions.

Word production

A total of 400 responses were recorded in the production task. Performance was measured in number of correct letters and participants could type a maximum of 6 letters per word. Thus, the overall maximum number of correct letters was 2400. On day 2, participants produced 1172 (48.8%) correct letters and on day 8, this number increased to 1465 (61.0%).

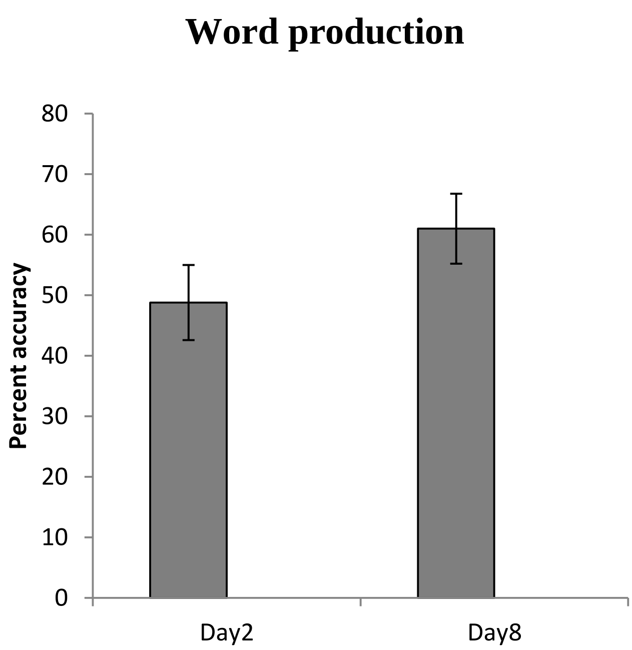


Figure 5.3. Percent accuracy of response in the production task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

Percent accuracy

There was a significant difference between day 2 and day 8, t_1 (19) = 4 .53, $p < .001$; t_2 (39) = 5.43, $p < .001$, with better accuracy on day 8 than on day 2.

5.2.3 Discussion

The results of the current experiment were mostly in line with the predictions based on the findings of Experiment 4 and Experiment 5. In the recognition memory task, there was no overall effect of day regarding both RTs and errors, but this was due to the fact that the meaningful condition showed better performance over time (with a trend in RTs but reaching significance for errors), whereas the meaningless condition showed a decline in performance (also showing a trend in RTs but significance in errors). This was confirmed in the significant (cross-over) interaction found for RTs and errors. An overall effect of semantics was also found for RTs and errors, with pairwise comparisons revealing a marginal effect on day 3 (RTs) and a highly significant effect on day 8, with faster RTs for the meaningful condition. The error data showed no difference between the conditions on day 3, but significantly lower error rates on day 8 for the meaningful condition. In the semantic categorization and word production tasks, results were very consistent with previous findings showing improved performance in the second test, which took place a week later.

First, the significant improvement over time found in the recognition memory task only for the meaningful condition (in accuracy) is consistent with the results in Experiment 4, which only showed significantly better performance on the second test for words learned with rich semantics. However, it is important to note that words in the meaningful condition here were tested three times on day 2 (because they were presented in all 3 tasks), while words in the meaningless condition were only tested once since they only appeared in the recognition memory task. Thus, unlike Experiment 4 and Experiment 5 where there was also a poor semantics condition, here it is hard compare improvement over time across conditions. What is clear from the current findings is that words learned with meaning show significant improvement over time (in accuracy) when they are tested across 3 different tasks on day 2 and then tested again on day 8. On the contrary, meaningless words show a decline in accuracy and no differences in RTs over the course of a week.

Second, in the current experiment, a marginal effect of semantics was found on day 2 (after correcting for multiple comparisons) and became highly significant on day 8. This result was slightly different from that of Experiment 4 which only found an effect of semantics on day 8. This finding suggests that the effect of semantics can be found much earlier and without a mediating testing session. As mentioned in the introduction, the design of this experiment might allow more statistical power due to the number of words per condition (20 versus 10 in previous experiments). This might have been one of the reasons for the early emergence of the (marginal) semantic effect, which in previous experiments only showed a trend in the first test and reached significance in the second test (day 8).

Third, the effects of improvement in performance from day 2 to day 8 in semantic categorization and word production were largely consistent with previous experiments employing these tasks. Given that meaningful words presented in the second test had been tested three times on day 1 and at least one time on day 8 before participants performed the semantic categorization and production tasks, the increased performance seen on the second test might be due to testing and not to the fact that words simply consolidated with the passing of time. This is further addressed in the general discussion.

5.3 Experiment 7

The main aim of Experiment 7 was to assess performance over time, but instead of testing participants on the same sets of words twice as in all previous experiments, here a different set of words was used each time. The advantage of this paradigm is that it allows assessing whether improvement in performance in previous experiments was simply due to consolidation over time involving sleep (e.g., Dumay et al., 2004; Dumay & Gaskell, 2007) or was mainly due to a testing effect (e.g., Wheeler et al., 2003; Roediger & Karpicke, 2006). It is important to mention that all studies reported earlier regarding consolidation of explicit memories over several days have also used retesting over different sessions. Only studies that have looked at the effect of consolidation following only one night of sleep have shown increased performance on items tested next day compared to items tested on the same day (e.g., Dumay et al., 2009). However, there are no studies that I know of that have found consolidation effects over a period of a week without the mediation of previous testing sessions. If consolidation over a week occurs simply because explicit memories for words become stronger with the passing of time, then improved performance for words tested on day 8 in comparison for words tested on day 2 should be expected. On the contrary, if the effect of the first test is key to improve performance over time, maybe equal or a decline in performance should be expected for the set of words tested on day 8.

5.3.1 Method

Participants

As in Experiment 6, participants in Experiment 7 were twenty English native speakers (13 female; mean age 21.3, SD 2.8) from the University of York. All individuals had normal or corrected-to-normal vision and had never been diagnosed with any language problems.

Materials and design

The stimuli used in Experiment 7 were exactly the same as in Experiment 6.

Procedure

As in Experiment 6, the current experiment took place over a week and included the same conditions. The training session of both experiments was exactly the same. However, in Experiment 7, participants were tested on half of the words on day 2 and the other half on day 8.

5.3.2 Results

Analyses were performed on all 20 individuals. The procedure regarding data filtering and statistical tests followed the same criteria as in previous experiments. The results are shown in Table 5.2.

Table 5.2. RTs for recognition memory and semantic categorization; percent accuracy of response for word production on day 2 and day 8.

	Day 2		Day 8	
	Meaningful	Meaningles	Meaningful	Meaningles
		Recognition memory		
Mean RT	807	865	834	911
SD	154	190	124	202
% errors	11.5	18.0	15.5	21.0
		Semantic categorization		
Mean RT	1118	---	1195	---
SD	449	---	390	---
% errors	11.1	---	25.3	---
		Word production		
% accuracy	71.8	---	61.4	---
SD	18	---	23	---

Recognition memory

Unlike Experiment 6, participants were tested on different words in each condition on day 2 and day 8. A total of 400 responses were collected on each day. On day 2, 59 (14.8%) of the RTs were removed from the analyses. These included 49 (12.3%) errors and 10 (2.5%) outliers. On day 8, the number of removed RTs reached 73, including 67 (16.8%) errors and 6 (1.5%) outliers.

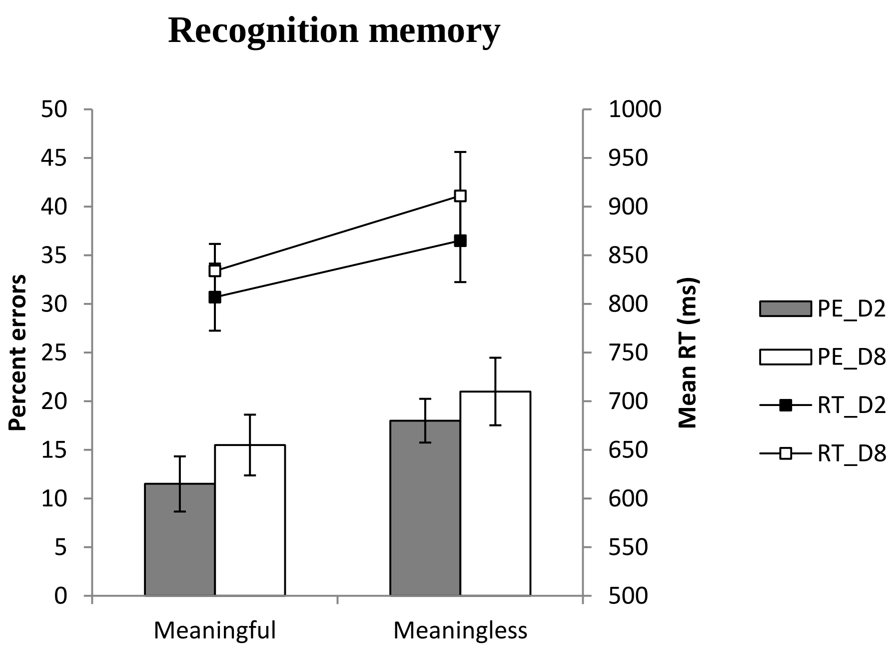


Figure 5.4. Percent errors and RTs in the recognition memory task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

RTs

A two-way repeated measures ANOVA was first conducted on RTs for correct responses with day and conditions as the main factors. A marginal effect of day was found, $F_1(1, 19) = 3.64$, MSE = 7160.84, $p = .07$, $\eta_p^2 = .16$; $F_2(1, 39) = 3.34$, MSE = 12963.70, $p = .08$, $\eta_p^2 = .08$, with better performance on day 2. There was a significant effect of conditions, $F_1(1, 19) = 10.15$, MSE = 8908.12, $p = .01$, $\eta_p^2 = .35$; $F_2(1, 39) = 9.38$, MSE = 20115.19, $p = .01$, $\eta_p^2 = .19$, with faster RTs for the meaningful condition. The interaction between day and conditions was not significant, $F_1(1, 19) = .19$, MSE = 8865.38 $p = .66$, $\eta_p^2 = .01$; $F_2(1, 19) = .58$, MSE = 16338.42, $p = .45$, $\eta_p^2 = .02$.

Errors

A two-way repeated measures ANOVA was also conducted on errors with day and conditions as the independent variables. No effect of day was found, $F_1(1, 19) = 2.11$, MSE = 116.05, $p = .16$, $\eta_p^2 = .10$; $F_2(1, 39) = 1.16$, MSE = 1.06, $p = .29$, $\eta_p^2 = .03$. There was a significant effect of conditions, $F_1(1, 19) = 6.74$, MSE = 106.84, $p = .02$, $\eta_p^2 = .26$; $F_2(1, 19) = 6.41$, MSE = .56, $p = .02$, $\eta_p^2 = .14$. The interaction between day and conditions was not significant, $F_1(1, 19) = .07$, MSE = 65.53 $p = .78$, $\eta_p^2 = .00$; $F_2(1, 39) = .02$, MSE = 1.42, $p = .90$, $\eta_p^2 = .00$.

Summary

The RT data showed a marginal effect of day with better performance on day 2. Both the RT and error data showed better performance in the meaningful condition than in the meaningless condition.

Semantic categorization

Unlike Experiment 6, half of the words were tested on day 2 and half on day 8. On day 2, a total of 200 responses were collected and 54 (27.0%) were deleted from the analysis. These included 50 (25.0%) errors and 3 (1.5%) outliers. On day 8, the same number of responses was collected but deleted responses only reached 24 (12%) and included 22 (11.0%) errors and 2 (1.0%) outliers.

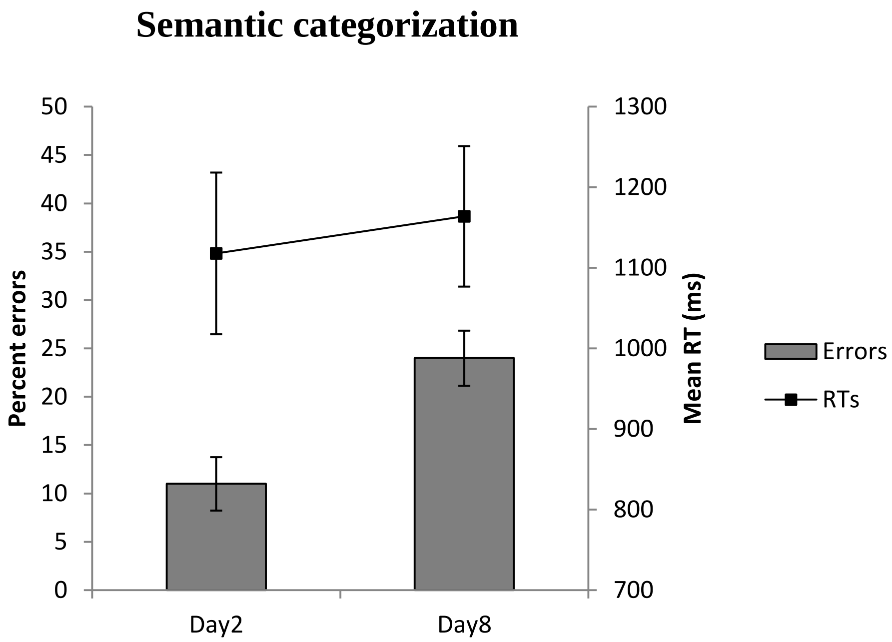


Figure 5.5. Percent errors and RTs in the semantic categorization task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

RTs

A t-test was conducted on RTs for correct responses. There was no difference between day 2 and day 8, t_1 (19) = .78, p = .44; t_2 (39) = .20, p = .84.

Errors

There was a highly significant difference between day 2 and day 8, t_1 (19) = 4 .15, p < .001; t_2 (39) = 2.61, p = .01, with higher error rates on day 8.

Summary

There was no difference between words tested on day 2 and day 8 regarding RTs. However, error rates were much higher on day 8 than on day 2.

Word production

A total number of 200 responses were collected on each day. As in Experiment 6, performance was measured in number of letters. This implies that the maximum number of correct letters in this task was 1200 on each day since participants were required to type 6 letters for each word. On day 2, the number of correct letters reached 861 (71.8%) while on day 8, this number decreased to 737 (61.4%).

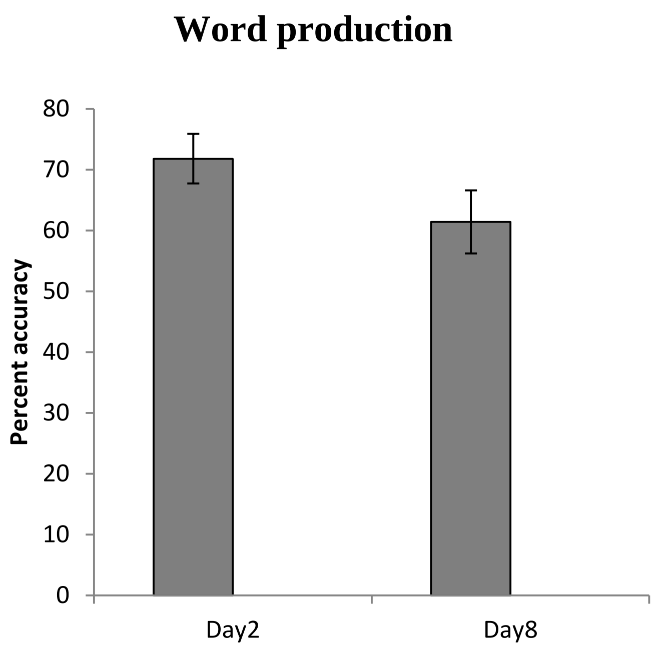


Figure 5.6. Percent accuracy of response in the production task on day 2 and day 8. Error bars represent standard error (SE) of the mean.

Percent accuracy

There was a significant difference between day 2 and day 8 with higher accuracy on day 2, t_1 (19) = 2 .25, p = .04; t_2 (39) = 3.16, p = .01.

5.3.3 Discussion

First, the results of Experiment 7 replicated the semantic effects found in the previous experiment even though only half of the items were tested on each day. This confirms once again that semantics has an effect on recognition memory. Another important factor to mention is that the effect seems very stable over time. This might suggest two things: words that participants learned associated with meaning are recognized better than words without meaning because meaning affects visual word recognition as in previous findings involving newly learned words (e.g., McKay et al., 2008) and real words (Pexman et al., 2003; Grondin et al., 2006). It might also reflect better retention of words with meaning due to the fact that they were encoded associated with semantic information (high level processing) and not only perceptual knowledge (shallow processing), which is consistent with the levels-of-processing literature (e.g., Brown & Craik, 2000). This is further supported by the effect of conditions found also in errors.

Second, an overall drop in performance from day 2 to day 8 was found in all 3 tasks. These findings suggest that memory traces for newly learned words decay over time if no instance of recall is provided after training. These findings are consistent with a study that showed an exponential decline in performance when different groups of participants were tested 5 minutes, 2 days, or a week after training (Roediger & Karpicke, 2006). In another study conducted by Wheeler et al. (2003), retention of novel words over time was better when participants were given a test immediately after training in comparison to simply restudying the material. This suggests that the improvement in performance a week after training observed in previous experiments of this thesis might have been due to the effect of the first test and not simply to the passing of time. It is important to mention that improvement over time without the mediation of a test has only been found when comparing sets of words tested immediately after training versus one day later (e.g., Davis et al., 2008). This means that other studies, in which subsequent testing of the same novel words took place over the course of a week, the improvement observed was probably due to an effect of testing and not simply to the passing of time (e.g., Dumay et al., 2004). Since the current study only used explicit tasks (recognition memory, semantic categorization, and word production), the current findings do not deny the possibility that more implicit forms of memory (e.g., perceptual learning) can consolidate over time without the need of testing or subsequent practice.

The reason why newly learned words tend to be forgotten over time rather than consolidated might be due to retroactive interference. The interference theory states that the presentation of subsequent material will disrupt the consolidation of previously presented information (Wixted, 2004). However, when memories survive initial stages of forgetting, they become more and more robust and harder to be forgotten. This is particularly useful to explain the current and previous findings. If words are trained during a certain period of time and tested a day later, the test might involve the reactivation of the words' memory traces making them stronger and less likely to be disrupted by future learning. However, if no reactivation occurs, memories become weaker over time and do not survive the disruption caused by subsequent memory traces.

In summary, regarding the consolidation of explicit memory traces for novel words, if subsequent days pass without the reactivation of the stored memories as in a test or maybe another instance of learning, the passing of time is likely to cause forgetting. However, if novel words are tested or maybe just reactivated before final assessment, performance will show less decay or improvement over time.

5.4 General discussion

The main purpose of the experiments presented in this chapter was to assess whether newly learned words consolidate over time due to the effect of a previous test or simply due to the passing of time. Highly consistent with previous experiments, Experiment 6, which tested all words on day 2 and again on day 8) found that newly learned words showed better performance over time in semantic categorization, word production, and recognition memory (meaningful condition only). On the contrary, Experiment 7, which assessed different sets of words on day 2 and day 8, showed an overall decline in performance from day 2 to day 8. Overall, these results revealed that improvement in performance over time in Experiment 6 and other previous experiments was mainly due to the effect of testing. As discussed earlier, the effect of testing is widely known and can enhance learning even beyond the restudying of material (e.g., Wheeler et al., 2003; Roediger & Karpicke, 2006). The decay in performance in Experiment 7 was explained in terms of the retrograde interference theory (e.g., Wixted, 2004), which suggests that the learning of subsequent information can affect the consolidation of previously learned knowledge. Another possible explanation might have to do with retrieval processes inhibiting the subsequent recall of information (e.g., Brown, 1968; Roediger, 1973; Anderson, Bjork & Bjork, 1994), For instance, in the study conducted by Anderson et al., which had participants practice retrieving half of the items from each of several categories, found that a subsequent recall of all items from each category showed significant inhibition of the non-practiced items relative to appropriate control. Anderson and his colleagues explained that nontarget items were inhibited or suppressed during the first retrieving practice and this inhibition persisted in the second session. Overall, previous retrieval facilitated the recall of target items, but suppressed the recall of nontarget items in following instances of retrieval. In Experiment 7 participants were trained on all words but were then tested on half of the words, which might have inhibited memory traces corresponding to the set of words that was tested later. This might have influenced the drop of performance on day 8.

The results of these two experiments have implications on previous findings. Particularly in Experiment 4 and Experiment 5, significant improvement over time was found for novel words learned with rich semantics (RS), but not for novel words with poor semantics (PS) in the recognition memory task. Since it has been concluded that the effect of testing is very powerful, it is important to add to these findings that RS might have benefited more from retrieval than PS due to the hierarchical difference in processing (deeper semantic processing for RS), which allowed better performance on the subsequent test.

Regarding effects of semantics, both experiments showed a consistent advantage for words trained with meaning in comparison with meaningless words. As suggested in previous studies, this might be simply due to the fact that words with semantics have a processing advantage over words with no semantics, which is consistent with numerous studies of familiar words in which an advantage is generally found for words with rich meaning over words with poor meaning in lexical decision tasks (e.g., Pexman et al., 2002; Borowsky & Masson, 1996; Buchanan et al., 2001). Because the current experiments involved the learning of new words, the advantage in recognition memory for meaningful over meaningless words might also be due to the quality of encoding. Brown and Craik (2000) argued that the cognitive system is hierarchical with lower level processes representing sensory aspects and higher level processing representing significance or meaning of objects or events. Shallow levels of processing can be accessed bottom-up while deeper levels of processing can be accessed both bottom-up and top-down. The fact that lower level representations are not easily accessed by top-down processes makes them less likely to be maintained and rehearsed than higher-level processing. In the current experiments, novel words with no meaning were more likely to be processed in a shallow way than novel words with rich meaning during the learning phase. This difference in the quality of encoding might have contributed to the effects found in the recognition memory task.

In summary, the experiments in this chapter showed that improvement over time in all tasks is necessarily mediated by the effect of testing. On the contrary, if no test is given to participants a day after training, performance declines dramatically after a week. Differences in performance regarding semantics versus no semantics can be explained by the classic processing advantage in favour of words with rich semantic representations over words with poor representations, and a hierarchical difference in the level of processing during encoding.

Chapter 6 - Neural correlates of semantic richness for familiar and novel words

6.1 Introduction

The five preceding chapters of this thesis have presented behavioural experiments aimed at understanding the process of word learning and the contribution of factors such as linguistic context, semantics, and consolidation over time. The main findings of these experiments have shown that participants can easily acquire information from sentence contexts and, consequently, infer the meaning of new words. Follow-up experiments have also shown that semantic richness (the amount of semantic information associated with a novel word) can influence the quality of learning and the processing of newly learned words. This has been demonstrated in a number of tasks including recognition memory, semantic decision/categorization, and word production. These experiments were the first to manipulate semantic richness as participants learned new words and the results seem to support previous findings in word learning studies that have compared semantic conditions versus nonsemantic conditions (McKay et al., 2008; McKague et al., 2001), and studies with familiar words which have manipulated semantic richness, particularly using the number of semantic features (e.g., Pexman et al., 2002; Pexman et al., 2003; Grondin et al., 2006; Pexman et al., 2008).

The manipulation of semantic richness in this thesis has undergone some changes from one study to the next in order to find the best possible way to create conditions of rich semantics, poor semantics, or no semantics. Changes have primarily involved the number and type of semantic features novel words were associated with during learning. As reviewed in Chapter 1, the idea that word meaning is represented as features has been around for at least four decades. The Feature Comparison Model of Smith et al. (1974) was probably the first to propose that word meaning was made up of different features combined together and not of single non-decompositional units as in the holistic approach (e.g., Fodor et al., 1980; Level, 1989; Roelofs, 1997). The featural view of semantics has been applied to a wide range of fields including neuropsychology (e.g., Patterson et al., 2007) computational modelling (e.g., Plaut & Shallice, 1993; Seidenberg, 2005), and behavioural psycholinguistics (e.g., Pexman et al., 2002; Grondin et al., 2006). Particularly important for the development of this thesis has been the study conducted by Pexman et al. (2002) since it was pioneering in introducing a semantic variable derived from the featural approach: *the number of semantic features*. As reviewed in earlier passages, this variable has been shown to influence naming (e.g., Pexman et al., 2002), lexical decision (e.g., Pexman et al., 2002; Grondin et al., 2006), and semantic categorization (Pexman et al., 2003). The number of semantic features has been established as a valid semantic variable and its application has gone beyond the use of concrete nouns, including also verbs (Vinson & Vigliocco, 2008).

As reviewed above, the number of semantic features has been used in a number of studies and has already been established as a valid variable in the manipulation of semantic richness. However, no previous studies have manipulated the number of semantic features during word learning. It is well-known that real words can be influenced by an infinite number of lexical and semantic variables, so the role of one single variable is particularly hard to isolate. As reviewed in Chapter 1, a word learning paradigm offers a reliable methodology regarding the manipulation of variables since effects of nontarget variables can be more tightly controlled. The behavioural experiments in the current thesis have been the first to show that a semantic richness effect measured in number of semantic features can also be demonstrated using a word learning paradigm. The evidence has been very consistent in showing that novel words acquired with many semantic features (visual and verbal) have a processing advantage in comparison with words learned with few features in semantic categorization and cued recall/word production. Results have also shown a semantic richness effect in recognition memory, but the results have been less consistent than for the other two tasks.

Since the evidence regarding both behavioural word learning experiments in this thesis and studies with familiar words is very consistent in showing the number-of-features effect, Chapter 6 aims to take behavioural experiments one step forward and examines the neural basis of semantic richness using familiar and novel words as stimuli. There are currently no previous fMRI studies of familiar or novel words that have explored the neural correlates of the number-of-semantic-features effect. Thus, it is relevant to conduct an fMRI study in order to identify brain regions within the semantic system that respond more to words with many semantic features or few semantic features.

Chapter 6 first presents a short review of relevant neuroimaging studies of word learning followed by studies of semantics. Then, a detailed theoretical motivation is outlined in order to support the current investigation. The first experiment of this chapter (Experiment 8) aims at collecting semantic features for 100 words using a similar method as in McRae et al. (2005). See Methods section of Experiment 8. As a result of Experiment 8, forty words (with high and low number of features) will be then selected to be used in Experiment 9. The latter is a combined behavioural and fMRI study looking at the neural basis of familiarity and semantic richness.

6.2 Review of relevant studies

First, a review of word learning studies that have used functional magnetic resonance imaging (fMRI) is presented followed by relevant studies of semantic memory.

6.2.1 Neuroimaging studies of word learning

The number of word learning studies that have used fMRI or other neuroimaging techniques is relatively limited. However, the research questions of the few studies available seem to be very diverse. As reviewed in Chapter 1, behavioural word learning studies have also very diverse aims and, as a consequence, the term 'word learning' is often misleading. It can referred to phonological word-form learning (no meaning involved), contextual word learning (with emphasis on incidental acquisition of meaning), paired-associative learning (with emphasis on word-concept mapping), and so on. The diversity of objectives has also been transferred to neuroimaging, primarily because researchers have explored the neural correlates of previous behavioural findings (e.g., Dumay et al., 2008; Breitenstein et al., 2005). This narrow relationship between behavioural and neuroimaging studies of word learning makes it hard to separate approaches.

In Chapter 1, a review of a wide range of word learning studies was provided including, for instance, phonological word-form learning studies. These studies have shown that new words can acquire word-like characteristics after a few exposures but only if they undergo a process of consolidation that requires sleep (Gaskell & Dumay, 2003; Dumay & Gaskell, 2007; Tamminen & Gaskell, 2008). Furthermore, these researchers argue that we temporarily store new lexical entries in the hippocampus, but during sleep new information is transferred to and integrated in neocortical areas for long-term storage. In an fMRI study that used the same paradigm as the behavioural experiments reviewed in Chapter 1, Davis et al. (2008) investigated the neural mechanisms responsible for the learning and consolidation of novel spoken words. They found that consolidated novel words (learned one day before scanning) showed reduced activation in classical language areas such as the left inferior frontal gyrus (IFG) and the left superior temporal gyrus (SFG) compared to unconsolidated novel words (learned on the same day of scanning) (see Figure 6.1). These results again demonstrated that new words need an incubation period in order to become fully integrated in the mental lexicon and behave more like real words. It is important to mention that the studies above have mainly focused on the learning and integration of new phonological forms and have not addressed the learning of semantics, which is the main component of words.

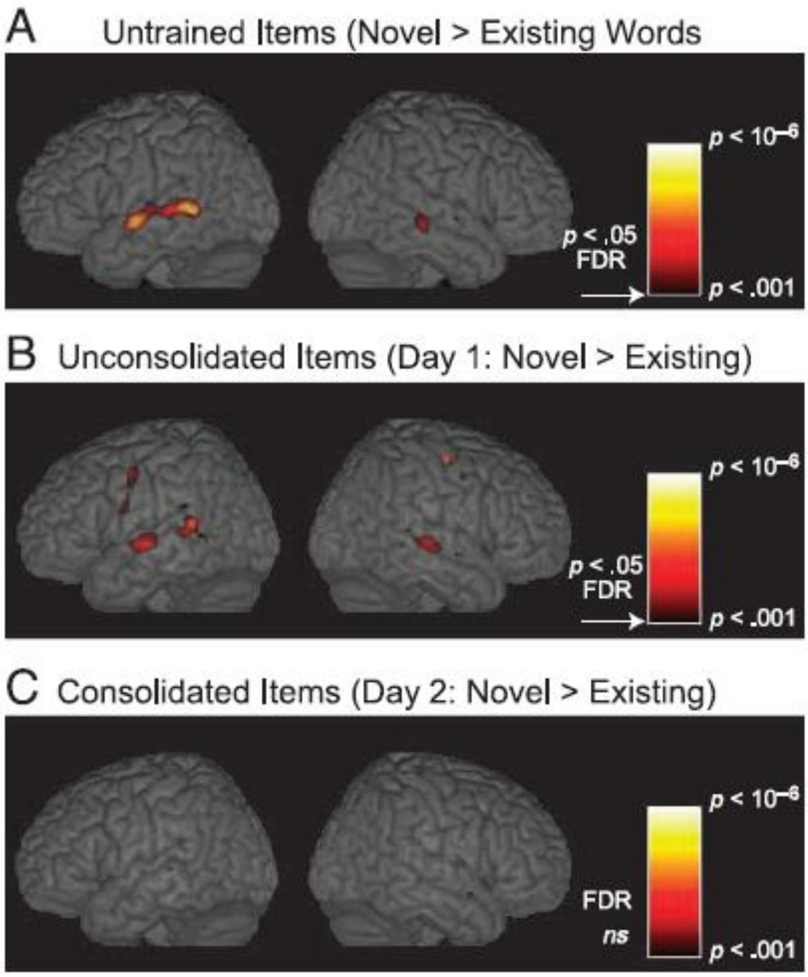


Figure 6.1. Brain regions showing response differences between novel and existing words. As novel words get consolidated over time (C), the difference in activity between novel and existing words disappears. Image adapted from Dumay et al. (2008).

One of the few studies that looked at semantic word learning was carried out by Breitenstein et al. (2005). They aimed to elucidate the brain regions that are engaged in the online process of novel word learning. A picture-word association task was used in order to teach participants the meaning of new words as they were lying in the scanner. The researchers identified a number of areas involved in the acquisition of new vocabulary, particularly its semantic component. They hypothesised that different regions of the hippocampus are involved in acquiring episodic and semantic information regarding the new lexicon. Additionally, as participants learned the new vocabulary, an increasing disengagement of the hippocampus could be observed, which is consistent with previous studies of artificial grammar learning (Opitz & Friederici, 2003; Strange et al., 1999), and face-name association (Zeineh et al., 2003). The main criticism to this study is perhaps the fact that they used familiar concepts such as *book* for which participants only had to learn new word-forms, so claims regarding semantic development cannot be made.

Another series of recent word learning fMRI studies include those of Mestres-Missé and colleagues (2007, 2008a, 2008b, 2010). These studies have used a different paradigm from that of Breitenstein and colleagues, but with a very similar focus. Both paradigms have mainly studied the neural changes associated with the online process of new vocabulary acquisition as participants learn in the scanner. However, in Mestres-Missé et al.'s studies, novel words are presented embedded in sentences and participants are asked to infer their meaning based on the contextual cues provided by each sentence. Thus, they aim to unravel the mechanisms involved in the word-to-meaning mappings generated via sentential context. They have proposed that word learning is a process that involves a large network of brain regions such as the left inferior frontal gyrus (IFG), superior temporal gyrus (STG), the fusiform gyrus (VWFA), the parahippocampal gyrus, and a number of subcortical structures (see Figure 6.2).

Even though Mestres-Missé and colleagues have provided clear evidence on the process of mapping new words to meaning, their studies do not shed much light on the acquisition of new concepts since participants are only required to learn new labels for highly familiar concepts (e.g., *lankey* = car). Their approach is more applicable to second language learning and can only partially explain the acquisition of new lexical entries.

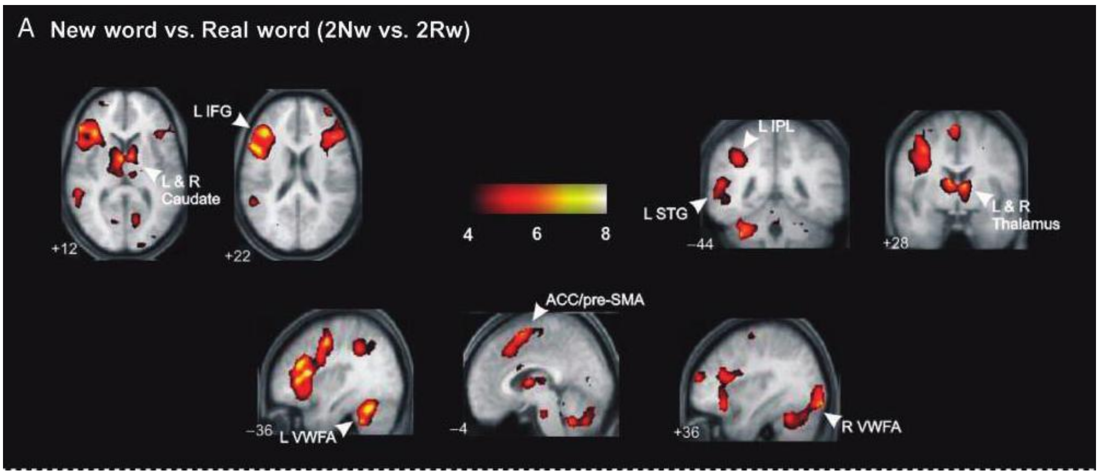


Figure 6.2. Brain regions showing increased activation for new words compared to real words as participants learn meaning from context. Images adapted from Mestres-Missé et al. (2008).

Given the above, there is certainly a lack of fMRI studies investigating neural correlates of trained novel words associated with previously unknown concepts. Furthermore, no word learning studies have manipulated the amount of semantic information associated with a novel word, and most studies have focused on the process of learning rather than the consequences of learning.

6.2.2 Studies of semantic memory

One of the main lines of research into semantic memory comes from the field of neuropsychology. Studies of brain-injured patients with semantic disorders have contributed to the understanding of how semantic knowledge is stored and organized in the brain (Martin & Chao, 2001, Damasio et al., 2004; Patterson et al., 2007). These studies have also provided insights into the manipulation and retrieval of semantic knowledge, which involves different processes that are context and task dependent (Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011). The reason why retrieval processes are important is because we have large quantities of knowledge stored, but not all this knowledge is relevant during a specific task. Hence, a system that allows task-relevant manipulation of knowledge can ensure access to the right information at the moment required, without the need to access all possible semantic knowledge.

Researchers have drawn a line between two semantic systems: a semantic representation system, which stores semantic information, and a semantic control system, which involves executive mechanisms that direct semantic activation according to the task being performed. The distinction between these two systems was proposed based on studies of patients with semantic dementia, who have atrophy of the anterior temporal lobes (e.g, Mummery et al., 2000; Nestor, Fryer, & Hodges, 2006; Jefferies & Lambon Ralph, 2006), and stroke aphasia patients, who have suffered infarcts to the prefrontal or temporal-parietal region (e.g., Chertkow, Bub, Deaudon, & Whitehead, 1997; Berthier, 2001).

A study conducted by Jefferies and Lambon Ralph (2006), which directly compared semantic dementia and stroke aphasia patients, found qualitative differences between the two groups regarding performance in semantic tasks. Both groups showed poor performance in tasks such as picture naming or word-picture matching. However, semantic dementia patients showed a high correlation between scores in the different semantic tasks they performed, while stroke aphasia patients showed poorer performance in tasks that involved higher demands of executive control. These findings suggest that damage to the anterior temporal lobes in semantic dementia patients result in increasing loss of semantic knowledge which is not task dependent, whereas in stroke aphasia patients damage to the prefrontal and temporal-parietal regions results in impairment of executive functions during access to conceptual knowledge and not to conceptual knowledge itself. The findings by Jefferies and Lambon Ralph had very important theoretical implications since they proposed that semantic cognition, defined as our ability to use semantic knowledge efficiently in different situations, requires two interacting systems. One system comprises amodal semantic representations built up with input from various association areas whereas the other one corresponds to a semantic control system in charge of regulating the activation of semantic information.

The fact that in the neuropsychology literature the anterior temporal lobe is considered the main and almost exclusive store of semantic knowledge does not mean that it is the only area of the brain devoted to semantic representation. Other brain regions such as the angular gyrus and ventrolateral middle temporal gyrus are also candidates for storage of semantic knowledge. Thanks to functional neuroimaging, great advances in the understanding of brain regions and their role in semantics have emerged in the last decade. Functional neuroimaging studies have helped to complement neuropsychological findings, particularly regarding the distinction between the conceptual representation and the semantic control systems. There is currently agreement that semantics is widespread in the brain with different brain regions processing different types of information and with perhaps one or several regions acting as amodal or integration areas. While some have proposed the anterior temporal lobe as the main ‘hub’ for semantic integration (e.g., Patterson et al., 2007), others seem to suggest that there are different ‘convergent zones’, where meaning is abstracted as a result of input from different areas. This view arises from recent models of featural representation suggesting that concepts are distributed across different brain regions specialised for processing information from different sensorimotor modalities (e.g., McNorgan, Reid, & McRae, 2011). Multimodal semantic models of this kind assume a hierarchy of ‘convergence zones’ over which information is integrated. According to this view, there is no one single ‘semantic hub’ as proposed by Patterson et al. (2007), but rather several areas organized in a cascading fashion. This view gives room to an extension of conceptual/semantic representation areas beyond the anterior temporal lobes.

Another region that has been referred to as 'amodal', receiving input from different other regions, is the angular gyrus (AG). Despite being rarely mentioned in the neuropsychological literature, neuroimaging studies have shown that the AG is consistently activated by a wide range of semantic tasks including auditory (e.g., Démonet et al., 1992) and visual (e.g., Vandenberghe et al., 1996) stimuli. A number of reviews have reported the angular gyrus as one of the main semantic areas (e.g., Cabeza & Neiberg, 2000; Vigneau et al., 2006; Binder et al., 2009). For instance, Vigneau et al. (2006) reported activation peaks in the AG for semantic contrasts of reading, categorization, and semantic association tasks. The fact that the AG shows increased activation for words in comparison to nonwords (e.g., Binder et al., 2003), suggests that it might have a role in semantic representation rather than in semantic control. Fluent access to the conceptual representation of a word might be reflected in increased activation in this region. Since nonwords do not have a lexical representation, automatic access to a lexical representation is virtually impossible, so less activation is found in the AG. Due to the fact that the AG is a very large area, it is very likely that different parts of the area are involved in different types of processing. This was explored in a recent study conducted by Seghier, Fagan, and Price (2010) which found that only the ventrolateral portion of the AG responded uniquely to semantic processing in matching and production tasks. Unlike the dorsal angular gyrus (dAG), the vAG was not sensitive to strings of Greek letters and non-objects (see Figure 6.3). Seguier et al. (2010) interpreted these findings as the vAG being involved in later stages of conceptual identification while dAG was associated with the search for a semantic representation. These findings are very relevant for the distinction between semantic representation and semantic control areas. As follows from Seguier et al. (2010) study, only vAG would have a role in semantic representation while dAG is probably involved in some sort of semantic control processing.

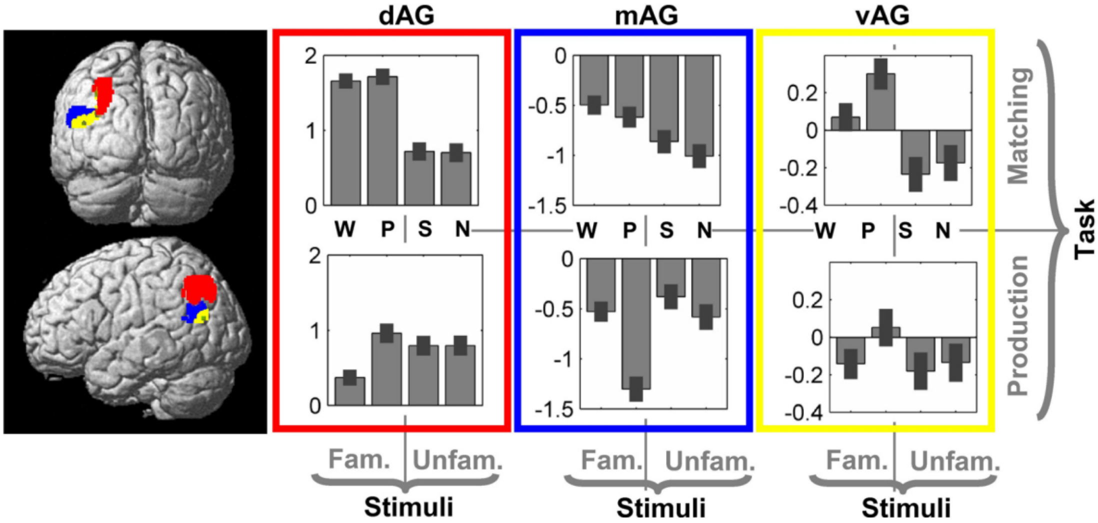


Figure 6.3. Signal level in the three subdivisions vAG (in yellow), mAG (in blue), and dAG (in red) of the AG for all tasks and stimuli: W, words; P, pictures; S, string of Greek letters, N, non-objects; Fam, familiar; Unfam, unfamiliar. Image taken from Seguier et al. (2010).

While anterior temporal pole (ATP) and left temporal regions (e.g., Vandenberghe et al., 1996; Rogers et al., 2004; Vigneau et al., 2006; Patterson et al., 2007; Binder et al., 2010), and the angular gyrus (e.g., Binder et al., 2009; Vigneau et al., 2006) can be regarded as integration areas of conceptual knowledge, the left inferior frontal gyrus (IFG) has been attributed a regulatory role in the recovery of semantic information. A number of fMRI studies have reported increased activation in this area during the performance of tasks that require high semantic control demand (e.g., Wagner et al., 2001; Badre et al., 2005; Badre & Wagner, 2007). For instance, activation increases during the retrieval of less dominant semantic features (e.g., *a banana makes you slip* versus *a banana has a peel* or *a banana is yellow*), or the subordinate meanings of ambiguous words (e.g., *a river bank* versus *investment bank)* (Bedny et al., 2008; Whitney et al., 2011; see also Gennari et al., 2007; Whitney et al., 2009). Wagner et al. (2001) suggest that the left IFG or ventrolateral prefrontal cortex (VLPFC), as it is also referred to in some studies, supports top-down (controlled) retrieval of knowledge when bottom-up (automatic) processes are not sufficient to retrieve task-relevant knowledge. According to this view, the left IFG is not a store of semantic knowledge as the ATP, but rather a semantic control area which directs semantic activation. This reinforces the idea that the semantic system is divided into semantic representation areas and semantic control areas (e.g., Patterson et al., 2007; Jefferies et al., 2007; Whitney et al., 2011).

In a recent review of over a hundred functional neuroimaging studies of general semantic contrasts (words > pseudowords, semantic tasks > phonological task, and high > low meaningfulness), Binder et al. (2009) identified a large left-lateralised network of semantic areas that were consistent across studies (see Figure 6.4.). These corresponded to the angular gyrus (AG) and adjacent supramarginal gyrus (SMG); the middle temporal gyrus (MTG) and posterior inferior temporal gyrus (ITG); the mid-fusiform gyrus and adjacent parahippocampus; the superior frontal gyrus (SFG) and adjacent middle frontal gyrus (MFG); the inferior frontal gyrus (pars orbitalis); the ventromedial and orbital prefrontal cortex; and the posterior cingulate gyrus, and adjacent ventral precuneus. According to Binder and colleagues, these regions play a role in high-level integrative processes since they receive multimodal and supramodal input from other regions. Due to their role in integrative processes, they have been referred to as heteromodal association areas.

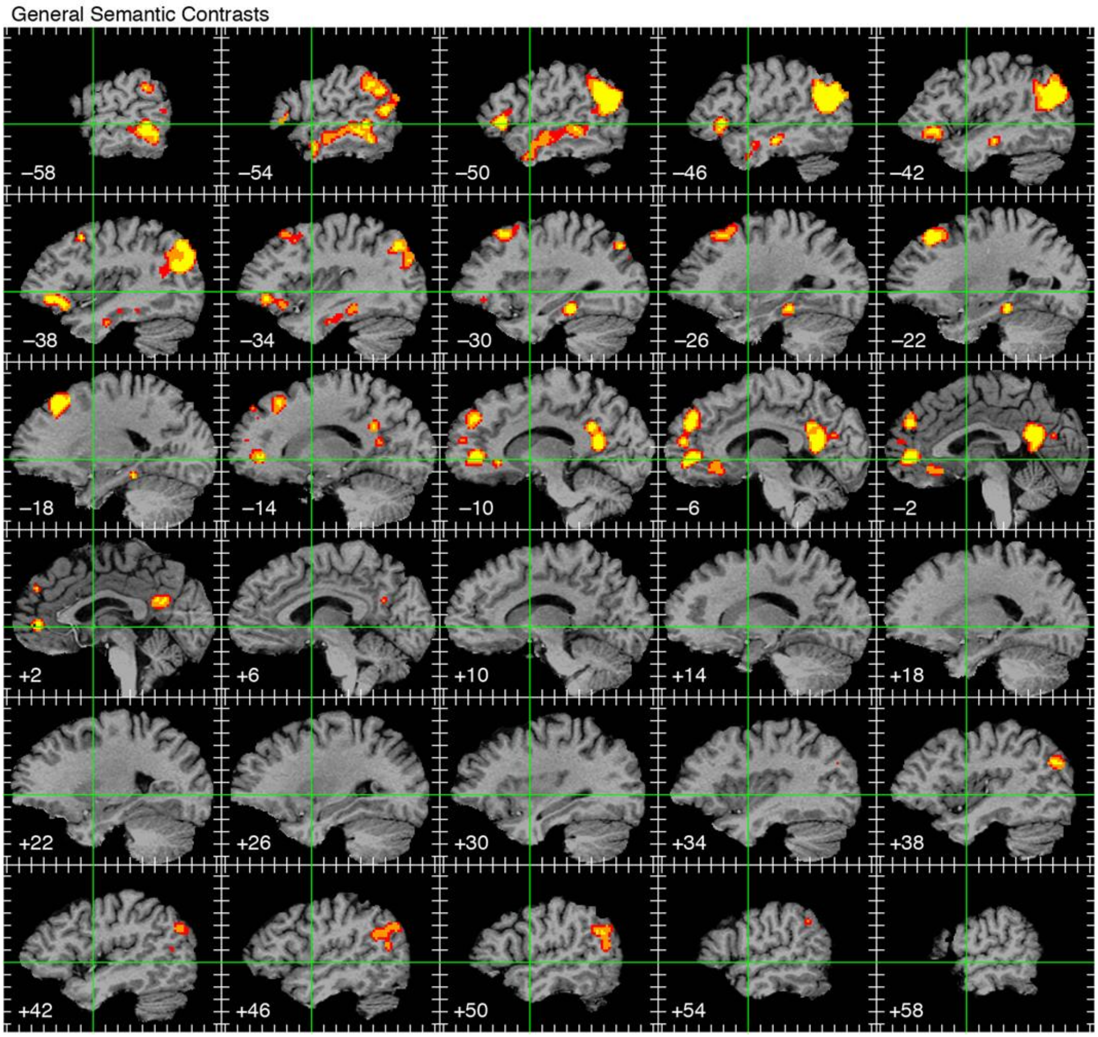


Figure 6.4. Semantic foci in Binder et al. (2009) for general semantic contrasts (words > pseudowords, semantic tasks > phonological task, high > low meaningfulness). Activations are displayed on serial sagittal sections through the stereotaxic space of Talairach and Tournoux (1988) at 4-mm intervals, with slice locations given at the lower left of each image. Green lines indicate the stereotaxic y and z axes. Tick marks indicate 10-mm intervals.

Due to the nature of the semantic contrasts in Binder et al.'s review, it can be inferred that most of the areas reported correspond to semantic representation areas and not semantic control. For instance, inferior frontal cortex activation is limited to the anterior portion, which corresponds to the pars orbitalis and not to the mid and posterior portions (pars triangularis and opercularis, respectively) most commonly linked to semantic control (e.g., Badre et al., 2005; Badre & Wagner, 2007; Whitney et al., 2011). Another reason for lack of activation in posterior inferior frontal cortex might be due to its role in phonology and it is generally more active for nonwords than words (e.g., Binder et al., 2003). This is due to the processing effort associated with the translation of unknown word forms into phonological forms. The same is true for motor regions such as the precentral gyrus, which has also been reported more active for nonwords than familiar words (e.g., Binder et al., 2003).

It is also worth noting that the anterior temporal lobe (reviewed earlier), which has a major role as a store of semantic knowledge (e.g., Patterson et al., 2007; Rogers et al., 2004), is not explicitly mentioned in Binder et al.'s review. This might be due to the fact that standard fMRI analyses generally fail to pick up activation from this region since it is close to air-filled cavities that can distort the fMRI signal, decreasing the signal-to-noise ratio (e.g., Visser et al., 2010). In a more recent study, Binder et al. (2011) also pointed out that fMRI protocols in current widespread use produce very little activation in the anterior temporal pole.

Another aspect worth clarifying is that not all areas listed in Binder et al.'s review are necessarily semantic areas. This is the case of the cluster including the posterior cingulate gyrus and the precuneus, which has been associated with episodic memory functions (e.g., Epstein et al., 2007; Vincent et al., 2006; also see Cavanna & Trimble, 2006, for review), and visual imagery (e.g., Hassabis et al., 2007) rather than semantic memory. The reason why these areas are consistently activated in contrasts that emphasize semantic processing might be due to a presumed evolutionary purpose (Binder et al., 2009). Binder and colleagues further explained that the brain has evolved as to remember better those experiences which are highly significant to us such as events that evoke associations and concepts. This is in line with the levels-of-processing literature reviewed in Chapter 4 and Chapter 5 (e.g., Craik & lockhart, 1972), which predicts better encoding for semantic versus perceptual information, or deep semantic analysis versus low semantic analysis. The fact that semantic processing is better encoded than nonsemantic processing might explain why episodic areas light up more during the presentation of semantic stimuli than nonsemantic stimuli.

6.3 Theoretical motivation and chapter outline

The sizeable number of neuropsychological and functional neuroimaging studies of semantics has provided a relatively good map of the brain regions devoted to the retrieval and activation of semantic representations. As reviewed in the previous section of this chapter, most studies of verbal semantics have primarily focused on the comparisons of semantic conditions versus nonsemantic conditions (e.g., words > nonwords or semantic tasks > phonological tasks) (see Vigneau et al., 2006; Binder et al., 2009, for review), or tasks of high versus low semantic control demand (e.g., Badre & Wagner, 2007; Badre et al., 2005). However, there are almost no studies that have investigated finer distinctions between semantic stimuli such as the processing advantage of words with rich semantics reported in previous behavioural experiments (e.g., Pexman et al., 2002; Grondin et al., 2006), and word learning experiments of this thesis.

Examining the processing dissociation between words with rich and poor semantic representations can shed light on which areas of the semantic system are devoted to conceptual representation and which can be associated with semantic control functions, without the need of using comparisons across different tasks. As discussed in previous chapters of this thesis, novel words with high number of semantic features (high-NSF) are consistently processed faster than words with low number of semantic features (low-NSF) in semantic categorization tasks. This processing advantage was interpreted as high-NSF novel words acquiring a more complete conceptual representation than low-NSF novel words. If this is the case, high-NSF novel words should activate semantic representation areas more strongly than low-NSF novel words. The reverse effect should be expected regarding semantic control areas since words learned with few features acquire a much poorer semantic representation, which would increase the processing demand during the search for a conceptual representation.

One of the few studies that has investigated semantic richness using fMRI corresponds to the work of Pexman et al. (2007). However, Pexman and colleagues did not use the number of semantic features to manipulate semantic richness, as in previous studies of this thesis. Instead, they used the number of semantic associates (NSA). Their event-related fMRI experiment directly compared high-NSA words with low-NSA words in a semantic categorization task. Results did not show any significant activation when subtracting high-NSA from low-NSA familiar words. However, increased activation was found in left precentral, left inferior frontal, and left inferior temporal gyri for the reverse comparison (low-NSA versus high-NSA). Pexman et al. (2007) argued that high-NSA words produce less activation than low-NSA words due to faster semantic settling in semantic space, which is consistent with connectionist views of word processing (Plaut & Shallice, 1993). However, they did not fully explain how the areas that were significantly more activated for low-NSA words are engaged in semantic processing. Furthermore, they did not discuss why high-NSA words did not produce any significant activation in comparison with low-NSA words.

Chapter 6 aims to explore the neural correlates of semantic richness in familiar and novel words using the number of semantic features as the manipulative variable. The chapter proposes that familiar and novel words associated with high and low number of features will show differential patterns of activation in semantic representation and semantic control areas. While high-NSF words are more likely to activate conceptual representation areas, low-NSF words are expected to show increased activation in semantic control areas. This proposal arises from the widespread evidence reviewed earlier suggesting a dissociation between two systems of semantic cognition (e.g., Jefferies & Lambon Ralph, 2006; Whitney et al., 2011), and is further supported by the behavioural findings of a number of studies of familiar words (e.g., Grondin et al., 2006; Pexman et al., 2003; Pexman et al., 2008), and the word learning studies presented in previous chapters of this thesis which have shown the number-of-features effect on word processing, particularly in semantic categorization.

Chapter 6 presents two experiments. The main purpose of Experiment 8 was to collect semantic features from British speakers for 100 familiar words. As discussed in preceding chapters, feature production norms are currently available for use, but they were collected in North America (McRae et al., 2005; Vinson & Vigliocco, 2008) and might not be valid for British speakers. This is further discussed in Experiment 8. After semantic features were collected and recorded, words with high and low number of features were selected so that they could be used in Experiment 9 to manipulate semantic richness. Experiment 9 is a combined behavioural and fMRI experiment, which first aims to replicate previous behavioural findings regarding the processing advantage of words with high number of features. And second, it explores the neural correlates of words with high and low number of features aiming to provide evidence on two different semantic networks supporting the processing of words with rich and poor semantics.

6.4 Experiment 8

The purpose of Experiment 8 was to collect semantic features from British speakers for 100 English words, which corresponded to living and nonliving things. Then, based on the results of the feature collection task, 20 words with high number of features and 20 words with low number of features were selected. These words were then used as stimuli in Experiment 9 (see Methods of Experiment 9 for details).

As reviewed in Chapter 1, semantic features are collected from participants who are asked to list features or attributes that can describe the meaning of a given word. For instance, features produced by participants for the word *duck* can include *eats, flies, swims, quacks, has a bill, has wings, is edible*, etc. There are currently two main studies that have published semantic feature production norms for large sets of words. One of these studies was conducted by McRae et al. (2005) and collected semantic features for over 500 living (dog) and nonliving (chair) concepts. The study was conducted over a long period of time and included three main phases with collection taking place at McGill University, University of Southern California, and University of Western Ontario. A more recent study conducted by Vinson and Vigliocco (2008) recruited students from the Department of Psychology at the University of Wisconsin, Madison. Vinson and Vigliocco collected semantic features for 456 words including nouns referring to objects and events, and verbs referring to events.

As presented above, the studies conducted by McRae et al. (2005), and Vinson and Vigliocco (2008) collected features for large sets of English words. However, both studies were conducted at universities in Canada or the United States, so these feature norms might not be valid for application in the United Kingdom. For instance, by having a quick look at feature lists in McRae et al.'s study, there are a number of words that have low number of features (e.g., *cod* with 8 features) and others that have high number of features (e.g., *cougar* with 18 features), which might not correspond to what British speakers would produce. The first example (*cod*) is a very common type of fish in Britain used in the preparation of *fish & chips* (a traditional dish), so British participants are likely to produce many features regarding the word *cod*. The second example (*cougar*) might not be well-known in Britain since *cougar* is an animal that lives in America, so British participants might have an impoverished representation of this word and consequently list few features and not many as American or Canadian participants do.

In order to collect semantic features for English words that could more accurately represent the conceptual knowledge of British speakers, Experiment 8 was conducted at the University of York and participants were native speakers of British English. Feature analyses include total number of features with and without taxonomic features. A correlation analysis is also performed between number of features collected from North American participants and in the current experiment in order to compare to what degree dialects differ.

6.4.1 Method

Participants

Participants were 100 native speakers of British English recruited at the University of York, UK (67 females, 33 males; mean age 22.2 years, SD 3.3). All individuals had normal or corrected-to-normal vision and had never been diagnosed with any language or mental disorders.

Feature norms

Word selection

One hundred words corresponding to living (e.g., duck) and non-living (e.g., coat) were selected for the experiment. See Appendix 6.2. These words were taken from the study conducted by McRae et al. (2005) which collected semantic feature production norms for over 500 living and nonliving concepts. The words selected for the study corresponded to high or relatively high frequency words with a mean number of features (excluding taxonomic features) of 12.3 (range 5 - 20) in McRae et al.’s study. Each word corresponded to a single English noun avoiding (as much as possible) ambiguous concept names. Living concepts included animals such as mammals, insects, birds, and fish whereas nonliving concepts included objects such as tools, weapons, clothing, and musical instruments.

Collection of semantic features

Semantic features were collected in a lab with 4 computers (2 desktop and 2 laptop computers), so a maximum of 4 participants at a time were able to take part. Participants were told that the experiment was part of an investigation into how people process familiar words, so they were expected to list attributes or features for some common or relatively common English nouns. Features should contain as few words as possible, and all of them combined should define and describe the target word as completely as possible (see Appendix 6.1 for instructions).

Participants were randomly assigned to five groups of 20. Each group was given a different set of 20 words for which they were required to list semantic features by typing them on the computer. Words were pseudorandomly assigned to each list avoiding words from the same category to occur next to each other. Words were presented on a word processor document, each on a different page appearing centred on top and followed by an empty table with 30 rows (see Appendix 6.1, second page). The order of presentation of the words was counterbalanced across participants, so each word occurred at a different position with each participant. Participants were told that on average they should spend around 2½ minutes per word and around 50 minutes in total. However, if they needed extra time, they could stay for as long as they wanted. The times were given as a reference in order to keep participants focused throughout the experiment. Participants were also instructed to type one semantic feature in each row or to separate features with a comma if more than one feature was listed in the same row. If more space was needed, they could simply expand the table and continue writing. Participants were encouraged to write as many features as possible for each word and could not move to next word if they had not finished listing features for the preceding word.

Recording and selection of semantic features

The editing and recording of semantic features followed a similar procedure to that used in previous studies (McRae et al., 2005; Vinson & Vigliocco, 2008). The raw data obtained consisted of lists of features for each word and from each participant. The cut-off criterion adopted for inclusion of features as part of a word's meaning was 4. This means that 4 out of 20 participants had to list the same feature in order for that feature to be counted and validated. If less than 4 participants produced a given feature, this was not counted. The cut-off was chosen based on McRae et al.'s study which adopted a cut-off of 5, but had 30 participants listing features for each word.

The first stage of processing included obtaining the frequency for each semantic feature in each word. The frequency corresponded to the number of participants that listed a given feature (ranging from 1 to 20). Since participants do not write exactly the same words when they list a semantic feature, a number of rules were established in order to be consistent in the process of selection. (i) Synonym features were recorded identically. For instance, *is large* and *is big* were recorded as *is big*. (ii) Quantifiers such as *generally* or *usually* were eliminated from responses since the information they conveyed is carried by the production frequency (McRae et al., 2005). (iii) Features that were made up of an adjective plus a noun (*has four legs*) were divided into two (*has legs* - *has four legs*). (iv) Disjunctive features were also divided into two, so *is black or white* became *is black* and *is white*.

The second stage involved obtaining the frequency of the features listed for each concept by counting the number of participants who produced a given feature. As stated earlier, if the number of participants was 4 or more, then the feature was selected as part of a word's feature list. Once the frequency of each feature was obtained, the number of features that exceeded the threshold (4) was also counted separately for each word. Finally, the selected features in each word were classified into taxonomic[4] or nontaxonomic features, so a measure of the number of semantic features was obtained with and without taxonomic features. Unlike previous studies, features were not further classified into smaller categories because the purpose of the experiment was only to obtain measures of the number of features of each word. However, the features collected here can be further processed to obtain several other measures regarding subdivisions of nontaxonomic features (e.g., function, surface property, material, etc). Features can also be classified into distinctive (common to few concepts) or shared (common to many concepts). This classification is also beyond the scope of this study.

[4] Taxonomic features correspond to the category name of a concept. For instance, *animal* or *reptile* for *lizard*.

6.4.2 Analysis and selection of words

After the number of features for each word was computed, the 20 words with the highest number of features and the 20 words with the lowest number of features were selected. See Table 6.1, and Appendix 6.3 for full list of words. The number of features was exclusive of taxonomic features.

Table 6.1. Mean number of semantic features (NSF) for the words selected in Experimen 8.

	Words	Mean NSF	Range	SD
High NSF	20	18	14 – 24	2.5
Low NSF	20	9.6	5 – 13	2.4

Note: NSF, number of semantic features.

Correlation across datasets

As mentioned in preceding paragraphs, one of the reasons for conducting the current experiment was that the number of features might vary for some words across the North American and the British variety of English. Thus, a correlation was run between the number of features in McRae et al. (2005) and the current experiment for the 40 words selected. The analysis showed a highly significant correlation between the two sets of words, r = .48, p = 002 (see Figure 6.5). However, 10 words showed a difference of at least 1.0 SD across dialects, with 4 of these words showing extreme differences of 1.8 SD or higher (*cod*, 1.8 SD; *spade*, 2.5 SD; *coat*, 2.0 SD; *cougar*, 3.3 SD) (see Figure 6.5). These differences in the number of semantic features across dialects had a significant impact on the classification of words into high-NSF and low-NSF words. Five words (*cheetah, cod, hyena, piano,* and *spade*), classified as high-NSF words according to the results of Experiment 8, would have been classified as low-NSF words according to McRae et al.'s study. Likewise, five low-NSF words (*beaver, coat, cougar, pelican,* and *skirt*) in the current experiment would have been classified as high-NSF words according to the feature production norms by McRae and colleagues.

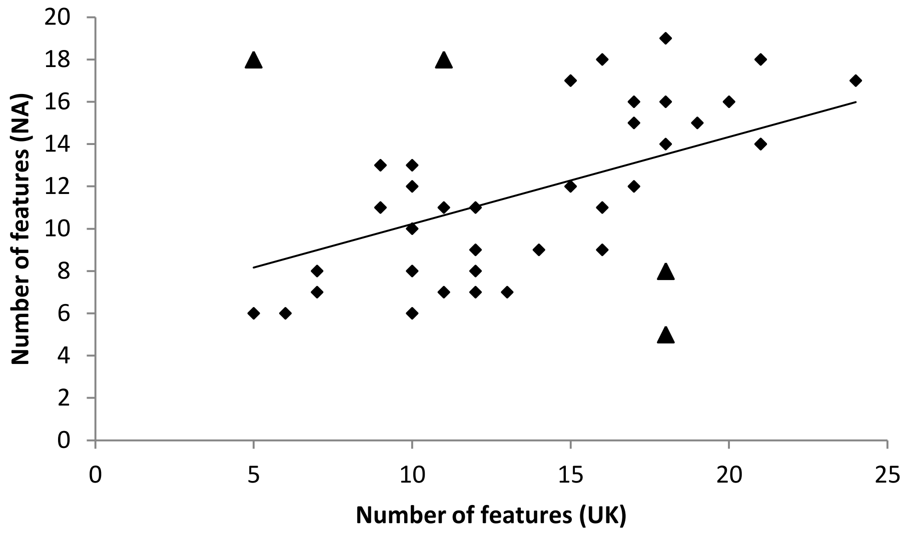


Figure 6.5. Correlation between number of features (North America) and number of features (United Kingdom). Triangles represent words with differences over 1.8 SD across dialects.

6.4.3 Conclusion

The aim of the current study was to select 20 familiar words with high number of features and 20 with low number of features to be used as stimuli for Experiment 9. The number of features for the 40 words selected correlated highly with the study of McRae et al. (2005), which collected features from North American participants. However, a number of words showed great differences across dialects which had a significant impact on the classification of words into high-NSF of low-NSF. This finding suggests that feature production norms collected in North America do not necessarily represent the conceptual knowledge of British speakers regarding certain words. The current experiment is probably the first attempt to collect semantic features from native speakers of British English and provides a valid reference for future studies.

6.5 Experiment 9

Experiment 9 is a combined behavioural and fMRI experiment. First, the study aims to replicate previous behavioural findings regarding the processing advantage of familiar (e.g., Pexman et al., 2003; Grondin et al., 2006) and novel words (experiments in this thesis) with high number of semantic features in a semantic categorization task. Second, the neuroimaging protocol of the current investigation is an event-related fMRI experiment conducted as participants performed the semantic categorization task in the scanner. The event-related fMRI experiment explores the neural correlates of familiar and novel words with high and low number of semantic features.

The main objective of the neuroimaging study is to identify which brain areas respond more strongly to words with high and low number of features. The well-established behavioural semantic richness effect is expected to be reflected in two separate components of semantic cognition: the semantic store or conceptual representation system, and the semantic control system. As reviewed earlier, this distinction has its basis in studies with semantic dementia patients, who have suffered atrophy of the temporal lobes, and consequently experience loss of conceptual knowledge (e.g., Mummery et al., 2000; Nestor et al., 2006; Jefferies et al., 2006), and stroke aphasia patients, who have damage to prefrontal and temporal-parietal areas, which leads to impairment of executive functions during semantic retrieval (e.g., Chertkow et al., 1997; Berthier, 2001; Jefferies et al., 2006). Jefferies et al. (2006) observed that stroke aphasia patients showed increasingly poorer performance as the control demands of semantic tasks increased, which suggested that impairment affected the access to the conceptual system. On the contrary, semantic dementia patients were not affected by the systematic variation in semantic control demands and did not show facilitation when cues were provided. This suggested that impairment affected the representation of concepts in semantic memory and not retrieval, as in stroke aphasia patients.

In the current investigation, words with high number of features are assumed to have rich semantic representations, so higher activation for these words should be found in brain regions associated with semantic representation. The neuropsychological literature has proposed the anterior temporal lobes as the 'semantic hub' due to its fundamental role as a store of conceptual knowledge (e.g., Patterson et al., 2007). This view has been known as the distributed-plus-hub model and suggests that information about different input modalities (e.g., actions, colours, etc.) from several brain regions converges in the anterior temporal lobes. This model falls into the category of hierarchically shallow models because it assumes a single location for semantic integration (McNorgan et al., 2011). On the contrary, hierarchically deep models support the view that there are several 'convergence zones' where features from different representational units and modalities are bind together. Early convergence zones integrate features from a single modality while later convergence zones can integrate multimodal information from distant brain regions (e.g., Simmons & Barsalou, 2003; McNorgan et al., 2011). The latter view of semantic representation opens the door to more than one semantic representation or convergence zone in the brain. As discussed in preceding sections of this chapter, along with the anterior temporal poles, the angular gyrus and ventrolateral middle temporal gyrus might also have a role in the storage of conceptual knowledge since they have repeatedly been found more active for contrasts of semantic versus nonsemantic processing (e.g., Binder et al., 2009). Thus, these two regions might also show increased activation for words with high number of semantic features. It can be hypothesized that the convergence of many features from diverse modalities is likely to produce a well-established semantic representation, which can be reflected in stronger activation in semantic representation areas such as those mentioned earlier. In contrast, words with low number of features should exhibit decreased activation in these regions due to a less elaborated lexical representation. The expected increased activation in conceptual representation areas for high-NSF words might also be accompanied with the activation of another network of areas that have been associated with episodic memory. For instance, it has been found that the posterior cingulate gyrus and the precuneus are among the areas that also show hightened activation during semantic tasks (e.g., Binder et al., 2009). As reviewed in earlier sections of this chapter, these brain regions tend to couple with semantic areas when the stimuli being encoded is semantic rather than perceptual, or requires deep semantic processing in comparison with more shallow processing. Thus, for the current experiment, more neural activity is expected in the precuneus with activation spreading towards the posterior cingulate gyrus during the presentation of high-NSF words.

While words with high number of semantic features are likely to show more activation in conceptual representation and episodic areas, words with low number of features, which are assumed to have poor semantic representations, are expected to show increased activity in semantic control areas, due to greater effort involved in the retrieval of knowledge. A well-established area involved in semantic control is the left inferior prefrontal cortex. As reviewed earlier, activation in this region increases during the performance of tasks that require high semantic control demand (e.g., Wagner et al., 2001; Badre et al., 2005; Badre & Wagner, 2007; Whitney et al., 2011). Thus, it can be hypothesised that low-NSF words, due to lack of a well-elaborated semantic representation, might put stronger demands on the semantic control system as access becomes increasingly more difficult. This should then be reflected in stronger activation in left inferior prefrontal cortex.

It is worth remembering that in the current experiment, familiar words and novel words have been defined as having rich or poor semantic representations based on the number of semantic features associated with them. For familiar words, this corresponds to the average number of different semantic features listed by 20 participants in Experiment 8 (see Methods section of Experiment 8, for details). However, novel words were defined as having many or few features based on the number of semantic features they were exposed to over 2 days of training (see Procedure section for details). The paradigm regarding novel words resembles the one used in Mestres-Missé and colleagues (2007, 2008a, 2008b, 2010) in the sense that participants learn novel vocabulary imbedded in sentences. However, in the present investigation participants are required to learn unknown concepts and words, and are not scanned during learning but only 3 days after the initial session, in order to allow enough time for consolidation (e.g., Davis et al., 2008). Therefore, the main focus here is not the online learning mechanisms of word learning, but rather the brain areas involved in long-term storage and the mechanisms involve in retrieving that knowledge.

In summary, the current investigation aims to replicate behavioural findings regarding the processing advantage of familiar and novel words with high number of semantic features. In a second step, it explores the neural correlates of the behavioural findings and predicts that words with high and low number of semantic features will be represented in two different networks of brain regions. On the one hand, high-NSF words are expected to show increased activation in conceptual representation areas (anterior temporal pole, ventrolateral middle temporal gyrus, and ventrolateral angular gyrus), and episodic memory areas (precuneus and posterior cingulated gyrus). On the other hand, low-NSF words are expected to show heightened brain response in left inferior prefrontal cortex, which is a well-established semantic control area.

6.5.1 Method

Participants

Twenty-two native speakers of British English from the University of York (seven males, fifteen females; mean age = 20.82; SD = 2.28) participated in the experiment. They received either payment or course credits for their participation. All participants were right-handed (laterality index > 80; Oldfield, 1971), had normal or corrected-to-normal vision, and had no diagnosed language problems.

Materials and design

Twenty familiar words with high number of semantic features (18 features on average; SD = 2.5) and 20 familiar words with low number of semantic features (9.6; SD = 2.4) were selected (see Appendix 6.3). The number of features was counted, exclusive of taxonomic features (e.g., an animal, a weapon, etc). Half of the words in each set corresponded to living things (birds, fish, insects, and mammals) and half to non-living things (house objects, weapons, tools/devices, clothing, containers, and musical instruments). Familiar words with high and low number of features were matched on number of letters, number of phonemes, number of syllables, bigram frequency, verbal frequency (BCN corpus), subtitles frequency, HAL frequency, orthographic neighbourhood, and phonological neighbourhood. Audio files recorded by a female native speaker of British English were created for each novel word. Forty nonwords that matched familiar words on initial letter, length (number of letters), bigram frequency, and orthographic neighbourhood were also included. These nonwords were assigned the meaning of an existing obscure concept (e.g., [5]*hoatzin*, [6]*cestus*) during the learning phase. Appendix 6.4 shows a list of concepts' names and nonwords that would become newly learned words after undergoing extensive training. Two sets of pictures were also used to show participants an illustration of each new concept. One set was modified using the Gaussian Blur filter in Adobe Photoshop CS3 Extended (Adobe Systems Incorporated, 2007) with a radius ranging from 7 to 10 pixels. Blurred images retained overall shape but lacked surface detail. See Figure 6.6 below and Appendix 6.5 for more sample images.

[5] 'Hoatzin' is the name of a bird that lives in the Amazon region.
[6] 'Cestus' was the name of a special combat glove used by gladiators.

Rich semantic | Poor semantics

Figure 6.6. Sample images used in rich semantics and poor semantics.

Nonwords were also presented in sentences that conveyed many features (rich semantics) or few features (poor semantics). See Table 6.2 below. The number of features per word was decided based on the number of features familiar words contained. See Appendix 6.6 for more sample sentences in each condition.

Table 6.2. Sample sentences in the rich semantics and the poor semantics conditions.

Rich semantics	Poor semantics
An etar lives in the Amazon, builds nests in swamps, and lays eggs. It eats all sorts of vegetables.	*An ornel is small and has brownish-red fur.*

Procedure

The experiment took place over three days with a one-hour session on each day. On day 1 and day 2 participants completed the training procedure, which consisted of different tasks aimed at teaching the phonological, orthographic, and semantic components of the novel words. On day 3, they were scanned while performing a semantic categorization task. The experiment received ethical approval from the Research Governance Committee of the York Neuroimaging Centre (YNiC) at the University of York, UK.

Training

Before starting the training session, participants were given an overview of the experiment and were asked a series of questions to check their eligibility for scanning. The training consisted of two one-hour sessions over two consecutive days. Sessions did not differ substantially from each other and were divided into two main parts, so participants could take a short break between parts. The order of presentation of the conditions was blocked and counterbalanced across individuals.

Participants were divided into two groups. Group 1 were presented first with 20 novel words with rich semantics (many features) followed by 20 novel words with poor semantics (few features). For group 2, the order of presentation of the conditions was inversed.

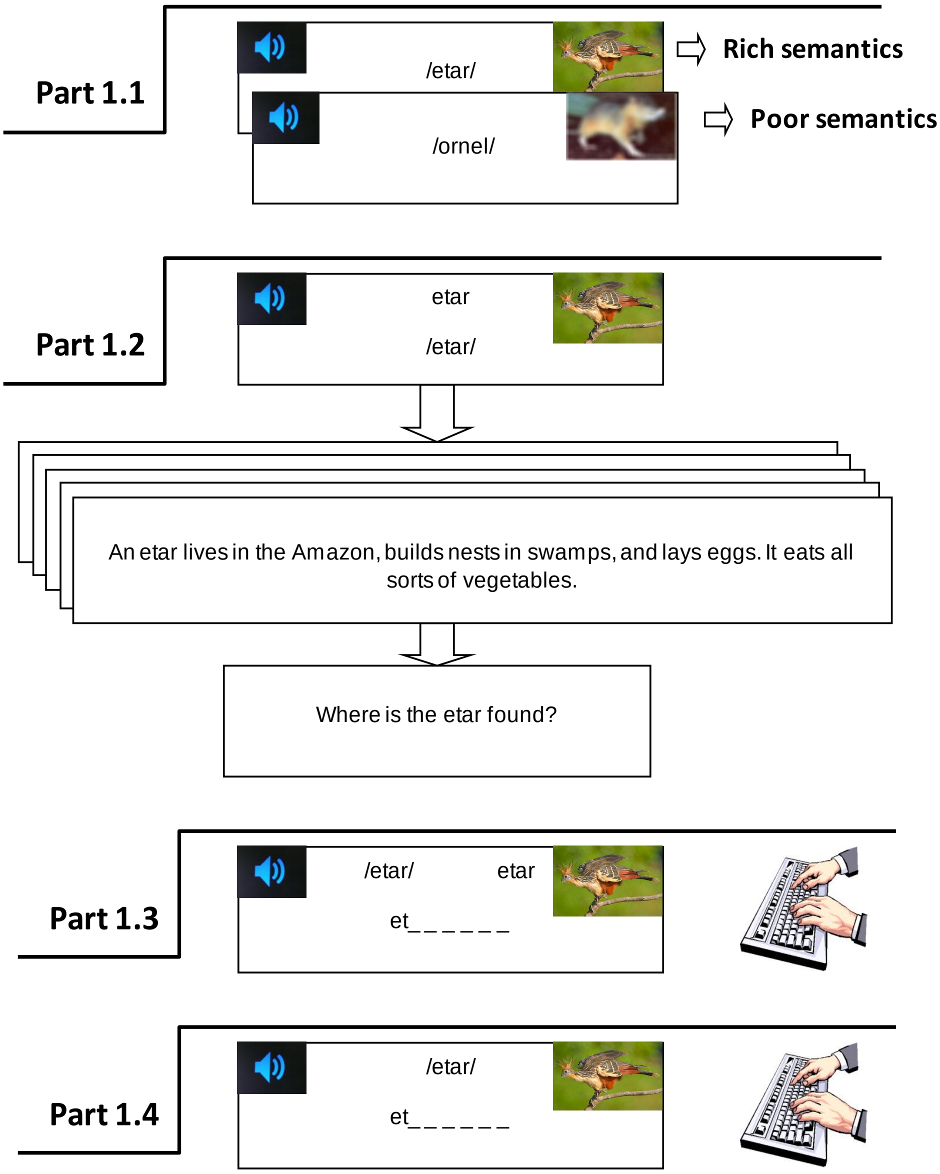


Figure 6.7. Structure of training procedure in Experiment 9 (day 1). Part 1.1 shows 2 sample words from each condition (rich semantics and poor semantics). Parts 1.2, 1.3, and 1.4 only show a sample word in the rich semantics condition.

As shown in Figure 6.7, part 1.1 on day 1 began with the oral and visual presentation of the first 20 novel words (either with rich or poor meaning, depending on the group). Each trial started with the presentation of a blank screen for 1 second followed by a picture for 4 seconds. Along with the picture, participants heard the phonological form of the novel word but were not exposed to its orthographic form. Stimuli were presented in two randomised blocks. Participants were instructed to look at the pictures that appeared on the screen and repeat the novel words aloud as they heard them. In the rich semantics condition, spoken novel words were presented accompanied with a standard-resolution picture while in the poor semantics condition, a degraded image was shown (see Figure 6.6). This was done in order to have a tighter control over the number of features participants would learn in each condition.

After reaching the end of part 1.1, participants received new instructions in order to complete part 1.2. In part 1.2 they were presented with all 20 words again but also adding the written modality. Learning trials began with the presentation of a blank screen for 1 second followed by a 4-second slide with a picture illustrating the novel concept and the written form of the new word appearing on top. As in part 1.1, participants heard the spoken form of the word and they were asked to repeat the word aloud. Then a 500-millisecond blank screen was displayed followed by 4 slides that stayed on for 12 seconds each or until the participant pressed *Enter*. Each slide presented the novel words in written modality and embedded in 5 sentences or short paragraphs (see Appendix 6.6). Each sentence conveyed a certain number of semantic features – roughly 8 or 16, depending on whether the condition was rich or poor semantics. Participants were instructed to read the sentences carefully and learn the information associated with each novel word. At the end of each set of sentences, a short question was displayed (e.g., W*here is the etar found?*) and required the participant to recall one of the semantic features presented earlier in the sentences. They had to type the correct answer and pressed *Enter* to check their responses (e.g., *In the rainforests of Malaysia and Indonesia).* The correct answer was then displayed for 5 seconds or until the participant decided to move to the next trial.

In part 1.3, participants were once again presented with a picture and the written and phonological forms of the novel words. They were asked to pay attention to each stimulus very carefully and then re-type each novel word in a space provided below the picture. The stimuli were presented only once and for a maximum of 15 seconds or until participants finished typing each word. Part 1.4 was very similar to part 1.3 except that participants were required to recall the written form and type it based on only the picture and the spoken word. Trials would start with the presentation of a blank screen for 500 ms followed by a slide containing the spoken and visual stimulus and a space for participants to type the orthographic form. At the end of the trial the correct written form was displayed for 1 second so that participants could check their responses. At the end of this part participants could take a short break before they started part 2.

Part 2 on day 1 was exactly the same as part 1, except that participants saw 20 more words in the other condition (rich or poor semantics depending on the group).

On day 2, the training session was divided into 3 parts. The first two parts were almost identical to the first two parts on day 1. However, the order of conditions was reversed, so if the novel words with rich semantics occurred first on day 1, they were presented after the set of novel words with poor semantics on day 2. Unlike the day 1 training, on day 2 participants were exposed to the written forms of the novel words from the beginning. Furthermore, in parts 1.2 and 2.2, which required participants to read sentences and then answer a question, the questions used on day 1 were replaced by new questions in order to target different features. For instance, if the question on day 1 was: W*here is the etar found?* On day 2, this question was replaced by: *What does an etar eat?* In parts 1.3 and 2.3, which required participants to type the novel words based on a picture and the spoken word, on day 2 novel words were presented twice instead of once (day 1). Finally, in part 3, participants were required to complete one more typing task (as in part 1.2 and 2.2), but including all the novel words they had learned over the two days.

Across the two training sessions, participants were exposed to each novel word 6 times in spoken modality, 7 times in written (alone) and visual modalities, and 10 times in written modality as part of sentences. In addition, they were required to type each word 4 times.

Testing

After completing the training session over two days, participants attended the York Neuroimaging Centre (YNiC) on day 3 for scanning and testing on the semantic categorization of familiar and novel words.

Prior to the scanning session, individuals were asked to complete 10 practice trials to familiarise themselves with the procedure in the scanner. The practice trials simulated the procedure in the scanner by presenting participants with words, which they had to categorize into living and nonliving things by pressing a button. If participants were not completely sure of what to do, they could repeat the procedure until they were absolutely sure. The radiographer then went through a routine checklist to be absolutely sure they were eligible for scanning. Then they then read and signed the standard YNiC consent forms (see Appendix 6.7) While in the scanner, participants were shown which buttons to press for each category and were asked one more time whether they had understood the task.

Semantic categorization

The experiment involved an event-related design with a semantic categorization task that required participants to classify familiar and novel words into *living* or *nonliving* things. Presentation software, version 12.0 (Neurobehavioral Systems, 2007) was used to deliver the stimuli, which were back-projected onto a screen that participants could visualise through a mirror attached to the scanner and located a few centimetres above their head. Stimuli were presented intermixed and in pseudorandom order, right at the centre of the screen on black background, and in 50-point courier new font. Participants were required to make semantic judgements to all 40 familiar words, half of them with high number of semantic features (e.g., *cheetah, mug*), the other half with low number of semantic features (e.g., *salmon, cork*); and 40 newly learned words learned with many semantic features (e.g., *darp, stire*) or few semantic features (e.g., *centeg, pon*). See Table 6.3 for characteristics of the stimuli. Each stimulus was presented for 2.5 seconds and participants were asked to indicate their responses by means of a button press using the index and middle fingers of their right hand. Half of the participants were asked to press “1” for *living* and “2” for *nonliving* things whereas the other half used the reversed order of buttons for each category. All stimuli were presented twice in order to increase statistical power, but the order of presentation was different each time. The mean intertrial interval from the offset of one stimulus to the onset of the next was 2.5 seconds (range 1.1–8.0 seconds) including a fixation cross for 500 ms appearing right before each stimulus.

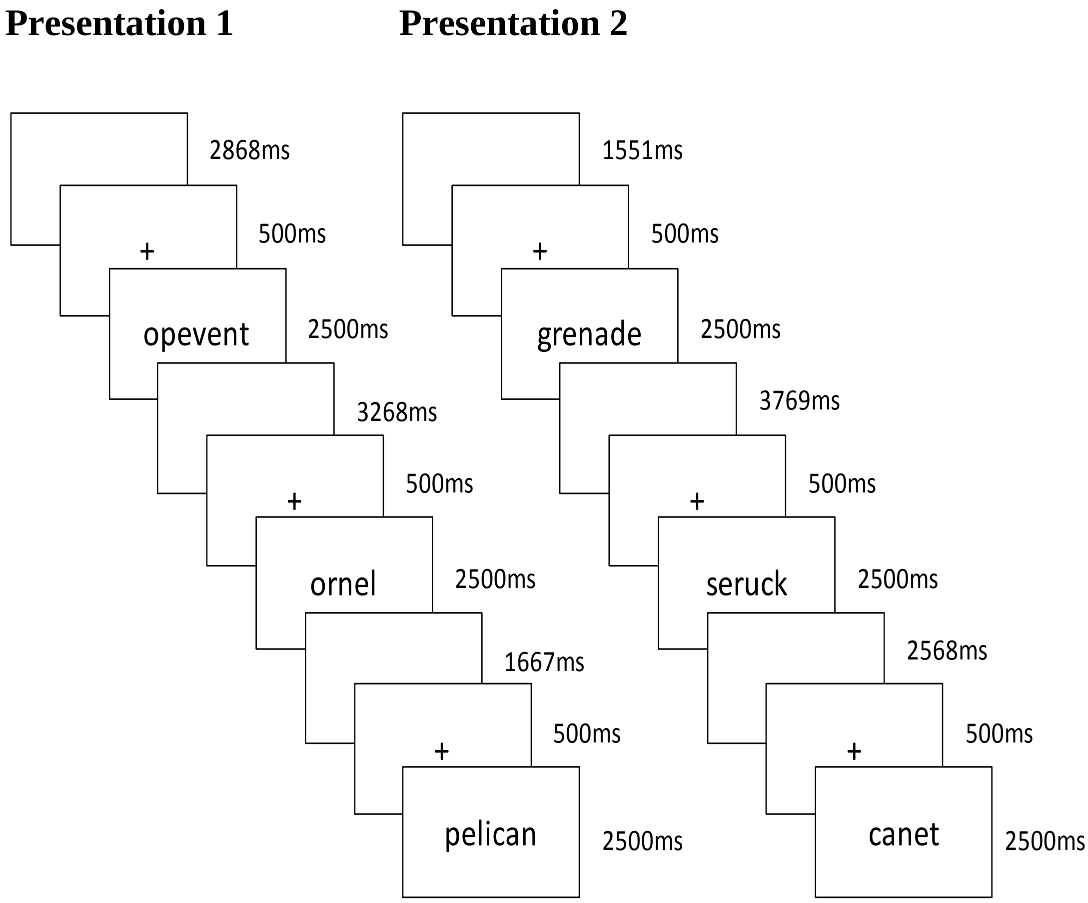


Figure 6.8. Sample presentation of the stimuli during the semantic categorization task in the scanner. Blocks correspond to the first and second presentation of the stimuli. Responses were made as soon as each stimulus was presented.

Table 6.3. Characteristics of familiar and novel words used in Experiment 9.

	familiar rich		familiar poor		novel rich		novel poor	
	Mean	*Range*	*Mean*	*Range*	*Mean*	*Range*	*Mean*	*Range*
NSF	18.0	14-24	9.6	5-13	18.0	14-24	9.6	5-13
NL	5.3	3-8	5.6	3-9	5.2	3-8	5.6	3-9
NPh	4.3	3-7	4.6	3-9	---		---	
NS	1.8	1-3	1.8	1-3	---		---	
Log F (BNC)	6.3	4-8	5.9	3-9	---		---	
HAL F	7.9	5-9	7.4	6-10	---		---	
Log SWF	2.6	2-3	2.3	2-3	---		---	
ON	3.9	0-16	3.0	0-16	3.4	0-16	2.9	0-16
PhN	7.9	0-25	6.8	0-28	---		---	
BG	1552	337-2364	1768	616-3034	1785	423-2567	1732	523-2645

Note: NSF stands for number of semantic features. For familiar words, this number corresponds to the number of features obtained as a result of Experiment 8. For novel words, it is the number of features participants were exposed to during training. (See *Materials* section for details). NL, Length measured in number of letters; NPh, Length measured in number of phonemes; NS, Length measured in number of syllables; LogF (BNC), British National Corpus log frequency; HAL F, frequency as reported by the HAL study (Lund & Burgess, 1996); Log SWF, frequency based on television and film subtitles (Brysbaert & New, 2009); ON, number of orthographic neighbours; PhN, number of phonological neighbours; BG, mean bigram frequency. The last three measures were extracted from the online version of the English Lexicon Project (Balota et al., 2007).

Feature recall task

After scanning, participants were required to complete a feature recall task. The task was run on a laptop computer using E-Prime software (Schneider et al., 2002). The full list of 40 novel words was presented in a different random order to each participant. Each word appeared on the screen for up to 1 minute or until participants press *Shift*. They were instructed to read each word and write (by typing) as many attributes as they could remember in order to describe the meaning of the word. As an example of what they were expected to write, they were shown all features listed for the word *cheese* from McRae et al. (2005)'s study (see Table 6.4). Participants took around 30 minutes on average to complete the task. This task was introduced in order to obtain an accurate measure of the number of features participants would remember for each novel word. See Appendix 6.8 for sample features recalled by participant 1.

Table 6.4. Semantic features listed for the word *cheese* in McRae et al. (2005).

Word *'cheese'*
...a dairy product, a food, eaten by mice, e.g. cheddar, is edible, is hard, is melted, is orange, is soft, is white, is yellow, made from milk, smells distinct, tastes good.

MRI data acquisition

Whole-brain structural and functional images were acquired on a 3.0 Tesla MRI scanner (General Electric HDx Excite) using an eight-channel eight-element phased-array head coil. Foam padding was used to keep participant's head stable in order to minimise movement. Participants were also required to wear earplugs to protect their ears. fMRI data were acquired using a gradient single-shot echo planar imaging (EPI) sequence (TR = 3 sec, TE = 33.7 ms, flip angle = 90º, FOV = 26 x 26, matrix = 128 x 128, continuous slice thickness = 3.5 mm). In order to facilitate localisation and co-registration of functional data to the structural image, a T1-weighted in-plane anatomical image was also acquired using a fluid attenuated inversion recovery (FLAIR) sequence [TR = 2.5 sec, TE = 9.94 sec, inversion time (TI) 1050ms, acquisition matrix = 256 x 224, FOV = 288 mm, slice thickness = 3 mm]. High-resolution T1-weighted structural images were also acquired using an inversion recovery-prepared 3-D FSPGR (Fast Spoiled Gradient Echo) pulse sequence (TR = 8.03 sec, TE = 3.07 sec, TI = 450ms, acquisition matrix = 256 x 256, Flip angle 20º, FOV = 290 mm, slice thickness = 1 mm).

fMRI data analysis

Functional imaging data were pre-processed and analysed using FEAT (FMRI Expert Analysis Tool, FMRIB, Oxford, UK; http://www.fmrib.ox.ac.uk/fsl). The data were first preprocessed using MCFLIRT motion correction, slice-timing correction with Fourier-space time series phase-shifting, spatial smoothing (Gaussian, FWHM 8 mm) and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting with sigma = 50 sec). Registration to high-resolution and standard space was carried out using FMRIB's linear registration tool (Jenkinson et al., 2002; Jenkinson & Smith, 2001).

First-level general linear model (FILM) time-series analysis was carried out using local autocorrelation correction (Woolrich et al., 2001) for each individual EPI sequence in order to produce contrasts for the group-level analysis. The data from each participant was entered into a general linear model for event-related analysis with 6 event types derived from the factorial crossing of familiarity (familiar, novel) and semantic richness (rich, poor).

Higher-level analyses across 21 participants (1 participants was deleted; see behavioural results section) were conducted using FLAME Bayesian mixed-effects analysis (Beckmann, Jenkinson, & Smith, 2003; Woolrich et al., 2004) in order to generate z-statistics based on the contrasts between the conditions presented above. Images were cluster thresholded at $Z > 2.3$ and a cluster significance threshold of $p < .05$ (Forman et al., 1995).

Regions of interest

ROI analyses were conducted to complement the whole-brain analyses and aimed to examine activation in ventrolateral angular gyrus (vAG), the precuneus, and pars opercularis. ROIs were selected based on coordinates from previous studies located in the left hemisphere. This method of selection was preferred over anatomical selection of brain areas because the latter tends to cover extremely large regions which might be involved in different types of processing (Poldrack, 2007). ROI analyses were conducted using the FEATQuery tool, which is part of FEAT-FMRI Expert Analysis Tool. A three-way factorial ANOVA with hemisphere (left, right), familiarity (familiar, novel), and semantic richness (rich, poor) was conducted on parameter estimates for percent signal change of the voxels defined by each mask described below.

(i) Ventrolateral angular gyrus (vAG)

Left hemisphere Right hemisphere

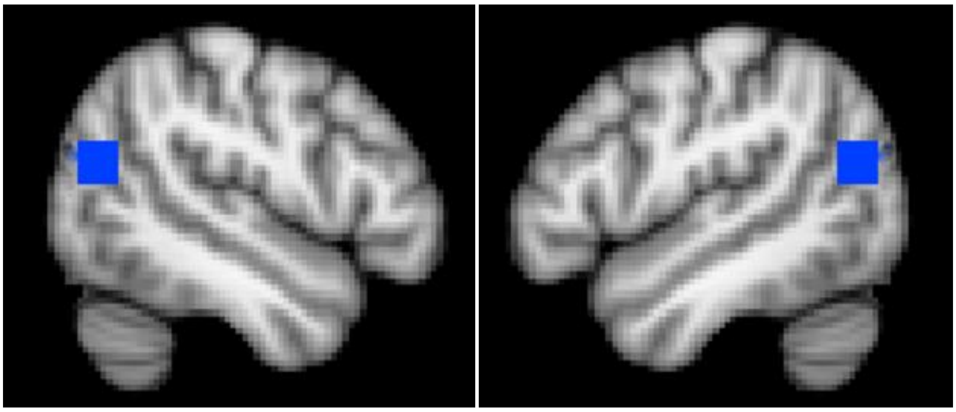


Figure 6.9. Mask showing location of ventrolateral angular gyrus (vAG) centred at MNI -48 -68 20 (left) and 48 -68 20 (right).

As reviewed earlier, activation in the AG was early reported in functional neuroimaging studies of semantic processing of auditory (Démonet et al., 1992) and visual (Vandenberghe et al., 1996) stimuli. In Binder and colleagues' (2009) review, the angular gyrus (AG) concentrated the largest proportion of activation foci for general semantic contrasts (e.g., words > nonwords). The evidence for involvement of the AG in semantic processing is very consistent and the early fMRI findings have been successfully replicated a number of times (see reviews by Vigneau et al., 2006; Binder et al., 2009). Seguier et al. (2010) proposed a subdivision of the AG in which they identified three main areas: a midregion (mAG), a dorsolateral region (dAG), and a ventrolateral region (vAG). Seguier and colleagues reported the peak for the vAG at MNI -48 -68 20. Furthermore, they found that this area responded above fixation during semantic decision to familiar stimuli, but below fixation during perceptual tasks. They concluded that vAG differed from the other two subdivisions of the AG because it responded uniquely to semantics and was involved in later stages of conceptual representation. For the current analysis, a mask of 7x7x7 mm was drawn centred at the coordinates for vAG provided by Seguier et al. Additionally, a mirror mask was created for the right vAG. Consistent with the evidence above, activation in the vAG is expected to be higher for familiar words than novel words and for high-NSF than low-NSF words.

(ii) The precuneus

Left hemisphere Right hemisphere

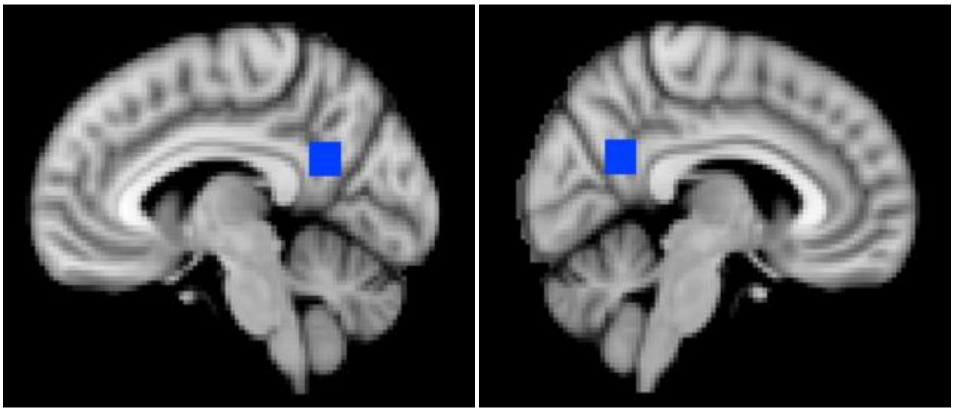


Figure 6.10. Mask showing location of left and right precuneus centred at -6 -56 24 (left) and 6 -56 24 (right).

The precuneus has been implicated in episodic memory (Vincent et al., 2006; Cavanna & Trimble, 2006), and visual imagery (Burgess, 2008; Hassabis et al., 2007; Johnson et al., 2007). Despite not being considered a semantic area, it is often activated in contrasts that emphasize semantic processing. The reason might be that it works in conjunction with semantic areas and its role is to keep a record of meaningful experiences (Binder et al., 2009). The left precuneus showed increased activation across several studies in Binder et al.'s review of semantic contrasts (e.g., semantic tasks versus phonological tasks). The approximate MNI coordinates in Binder et al.'s study were centred around -6 -56 24. In order to conduct the ROI analysis, a 7x7x7 mask was drawn around these coordinates with a mirror mask in the right hemisphere. Since the precuneus is generally activated in concert with semantic representation areas, it is expected to show increase activation for words relative to novel words and for high-NSF in comparison with low-NSF words.

(iii) Pars opercularis

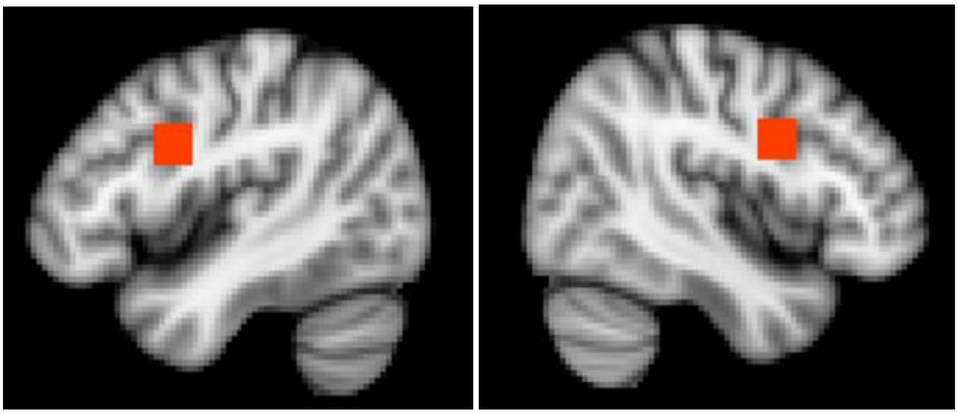


Figure 6.11. Mask showing location of left and right pars opercularis centred at MNI -44 4 26 (left) and 44 4 26 (right).

The third ROI corresponds to an area in the inferior prefrontal cortex. The name for this area has not always been consistent. In some studies, it has been referred to as posterior ventrolateral prefrontal cortex (e.g., Badre et al., 2005), here it will be referred to as pars opercularis. Pexman et al. (2007) reported the peak of activation for this area at -42 0 26 in Talairach space (roughly -44 4 26 in MNI space) in a contrast between words with low number of semantic associates (NSA) and words with high NSA. They found increased activation for low-NSA words and they attributed this effect to difficulty in semantic access. This type of semantic processing, which occurs at an early stage of semantic access, involves control and selection of semantic features rather than the activation of a conceptual representation, which occurs at a later stage (Thompson-Schill et al., 1997; Wagner et al., 2001). In order to look at activation in the pars opercularis, a 7x7x7 mask was drawn around the peak location reported in Pexman et al. (2007). A second mask was created for the same coordinates in the right hemisphere. Activation is pars opercularis is expected to be higher for novel than familiar words and for low-NSF relative to high-NSF words.

6.5.2 Results

Behavioural results

The results for the semantic categorization and feature recall tasks are presented here. One participant was removed from the analyses due to low performance in the semantic categorization task for novel words (71%, 3.2 SD below the group mean). The 21 remaining participants were entered in the analyses. The data from both tasks were analyzed using analyses of variance (ANOVAs) and t-tests using subject and items test statistics. In the semantic categorization task, only reaction times (RTs) for correct responses were included, and RTs below 300 milliseconds or over 2.5 SD from the mean (per participant) were regarded as outliers and removed from the analysis of RTs. The means for trimmed RTs are presented in Table 6.5. In the feature recall task, the number and proportion of features were analyzed. See Table 6.5.

Table 6.5. Behavioural results in the semantic categorization task.

		Presentation 1		**Presentation 2**	
		Familiar	**Novel**	**Familiar**	**Novel**
Mean RT	**Rich**	852	1073	781	888
SD		132	161	123	128
% errors		1.4	5.2	3.3	4.3
Mean RT	**Poor**	895	1155	820	930
SD		130	159	126	158
% errors		3.1	9.0	4.0	9.8

Semantic categorization

Behavioural responses for the semantic categorization task were collected in the scanner. A total number of 3360 responses were collected of which 174 (5.2%) were removed from the analysis. A hundred and thirteen (3.4%) of the removed responses corresponded to errors, 8 (0.2%) to no button press, and 53 (1.6%) to RTs above 2.5 SD from the mean.

Semantic categorization

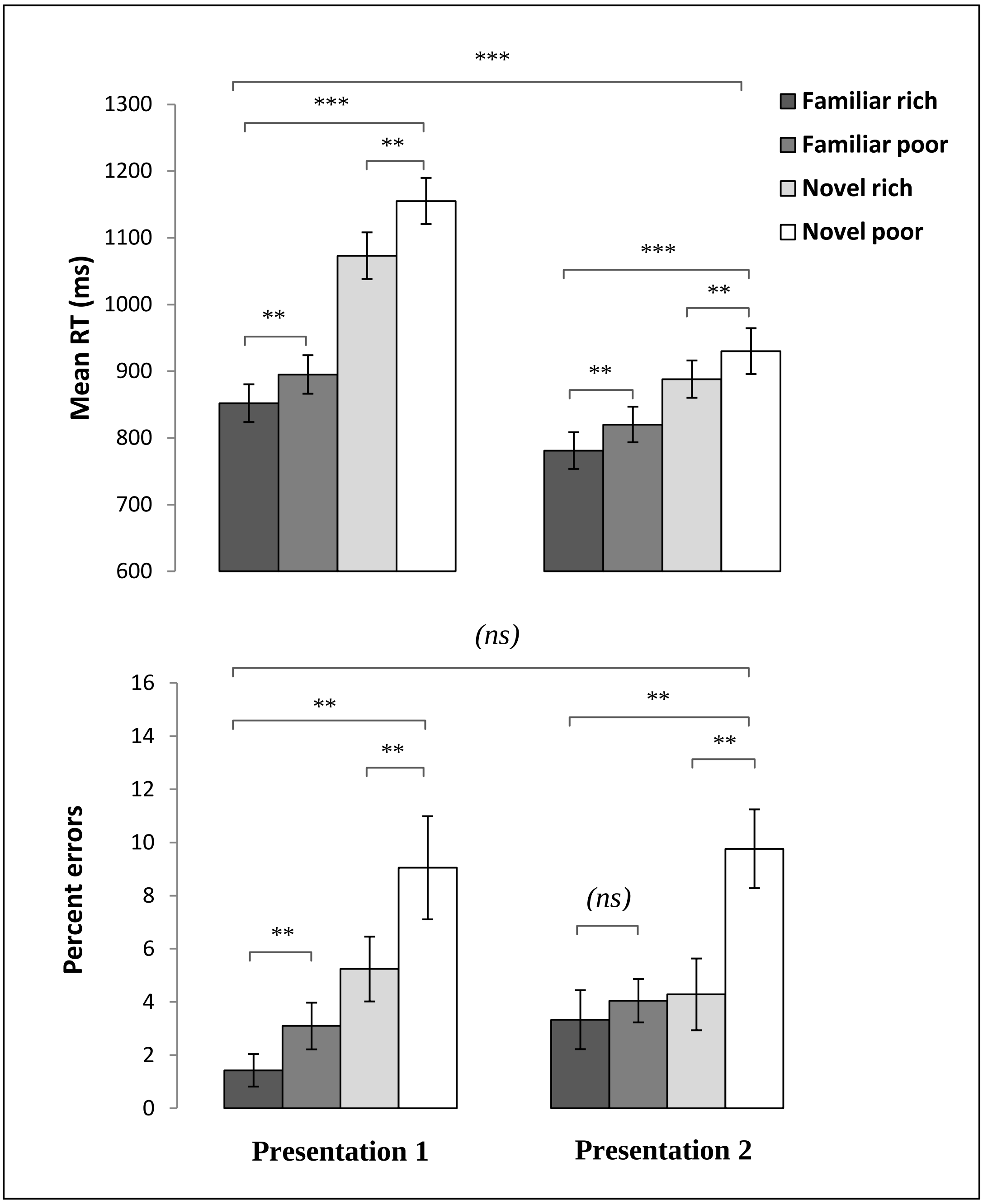


Figure 6.12. Semantic categorization RTs and errors for familiar and novel words with rich and poor semantics across two presentations. Error bars represent standard error (SE) of the mean. Statistical significance of ANOVAs (* = p < .001, ** = p < .01, *(ns)* = nonsignificant.**

RTs

A three-way repeated measures ANOVA with presentation (first and second), familiarity (familiar, novel), and semantic richness (rich, poor) as independent variables was conducted.

A significant main effect of presentation was found with faster RTs in the second presentation of the stimuli, $F_1(1, 20) = 104.63$, MSE = 780730.43, $p < .001$, $\eta_p^2 = .84$; $F_2(1, 19) = 71.49$, MSE = 1069723.23, $p < .001$, $\eta_p^2 = .79$. A reliable main effect of familiarity was also found, with familiar words showing faster RTs than novel words, $F_1(1, 20) = 63.09$, MSE = 2027096.34, $p < .001$, $\eta_p^2 = .76$; $F_2(1, 19) = 106.69$, MSE = 1202650.13, $p < .001$, $\eta_p^2 = .85$. The effect of semantic richness was also significant showing faster RTs for words with rich semantics, $F_1(1, 20) = 29.80$, MSE = 379247.73, $p = .01$, $\eta_p^2 = .60$; $F_2(1, 19) = 8.66$, MSE = 1382826.65, $p = .01$, $\eta_p^2 = .31$. There was a significant presentation x familiarity interaction, $F_1(1, 20) = 45.98$, MSE = 397511.85, $p < .001$, $\eta_p^2 = .70$; $F_2(1, 19) = 23.52$, MSE = 682256.53, $p < .001$, $\eta_p^2 = .55$. However, none of the other two-way interactions reached significance: Presentation x semantic richness [$F_1(1, 20) = 1.66$, MSE = 300650.30, $p = .21$, $\eta_p^2 = .08$; $F_2(1, 19) = .61$, MSE = 947982.63, $p = .44$, $\eta_p^2 = .03$] and familiarity x semantic richness [$F_1(1, 20) = 1.09$, MSE = 422880.49, $p = .31$, $\eta_p^2 = .05$; $F_2(1, 19) = .63$, MSE = 812305.87, $p = .44$, $\eta_p^2 = .03$]. The three-way presentation x familiarity x semantic richness interaction was also not significant, $F_1(1, 20) = 1.57$, MSE = 221211.88, $p = .23$, $\eta_p^2 = .07$; $F_2(1, 19) = .35$, MSE = 936201.11, $p = .56$, $\eta_p^2 = .02$.

In order to break down the presentation x familiarity interaction, separate one-way ANOVAs were conducted for RTs collected in the first and the second presentation of the stimuli. Results showed a significant effect of familiarity with faster RTs for familiar than novel words in both presentations, but with a bigger effect (as shown by η_p^2) in the first presentation, $F_1(1, 20) = 136.56$, MSE = 889154.30, $p < .001$, $\eta_p^2 = .87$; $F_2(1, 19) = 87.41$, MSE = 1344938.00, $p < .001$, $\eta_p^2 = .82$, than in the second presentation, $F_1(1, 20) = 16.11$, MSE = 1535453.89, $p = .001$, $\eta_p^2 = .45$; $F_2(1, 19) = 49.64$, MSE = 539971.36, $p < .001$, $\eta_p^2 = .72$. One-way ANOVAs were also conducted on familiar and novel words separately. Results showed a significant effect of presentation in both sets of words, with faster RTs in the second presentation and with a bigger effect for novel words [familiar, $F_1(1, 20) = 32.35$, MSE = 350589.34, $p < .001$, $\eta_p^2 = .62$; $F_2(1, 19) = 13.97$, MSE = 803486.78, $p < .001$, $\eta_p^2 = .42$; novel words, $F_1(1, 20) = 107.08$, MSE = 827652.94, $p < .001$, $\eta_p^2 = .84$; $F_2(1, 19) = 85.71$, MSE = 948492.98, $p < .001$, $\eta_p^2 = .82$].

Errors

A three-way repeated measures ANOVA was also conducted on errors including the same factors (presentation, familiarity, and semantic richness), as in the RT analysis. Results showed no effect of presentation, $F_1(1, 20) = .84$, MSE = 21.44, $p = .37$, $\eta_p^2 = .04$; $F_2(1, 19) = .18$, MSE = 52.24, $p = .68$, $\eta_p^2 = .01$. However, there was a significant effect of familiarity, with lower error rates for familiar than novel words, $F_1(1, 20) = 13.17$, MSE = 53.80, $p = .01$, $\eta_p^2 = .40$; $F_2(1, 19) = 10.16$, MSE = 64.07, $p = .01$, $\eta_p^2 = .35$. There was also a significant effect of semantic richness with higher error rates for words with poor semantics, $F_1(1, 20) = 7.92$, MSE = 45.10, $p = .01$, $\eta_p^2 = .28$; $F_2(1, 19) = 6.04$, MSE = 58.51, $p = .02$, $\eta_p^2 = .24$. Unlike the RT data, the presentation x familiarity interaction was not significant, $F_1(1, 20) = 1.31$, MSE = 19.21, $p = .27$, $\eta_p^2 = .06$; $F_2(1, 19) = .55$, MSE = .37, $p = .55$, $\eta_p^2 = .02$. There was also no interaction between presentation and semantic richness, $F_1(1, 20) = .07$, MSE = 20.40, $p = .80$, $\eta_p^2 = .00$; $F_2(1, 19) = 1.83$, MSE = 61.02, $p = .19$, $\eta_p^2 = .09$. The familiarity x richness interaction reached significance by-participants, $F_1(1, 20) = 4.58$, MSE = 27.34, $p = .05$, $\eta_p^2 = .19$, but not by-items, $F_2(1, 19) = 1.83$, MSE = 61.02, $p = .19$, $\eta_p^2 = .09$. Finally, the three-way presentation x familiarity x semantic richness interaction did not show a reliable difference, $F_1(1, 20) = .84$, MSE = 21.44, $p = .37$, $\eta_p^2 = .04$; $F_2(1, 19) = .72$, MSE = 38.66, $p = .41$, $\eta_p^2 = .04$.

The familiarity x semantic richness interaction (by-participants) was explored by conducting two separate ANOVAs for familiar and novel words. The ANOVA conducted on errors to familiar words showed no effect of semantic richness, $F_1(1, 20) = 1.07$, MSE = 27.89, $p = .31$, $\eta_p^2 = .05$. However, the ANOVA conducted for novel words showed a reliable effect of semantic richness, $F_1(1, 20) = 10.16$, MSE = 44.55, $p = .01$, $\eta_p^2 = .34$.

Summary

The factorial ANOVA conducted on the data showed a reliable effect of familiarity and semantic richness in both RTs and errors. The effect of presentation was only found in the RT data. The presentation x familiarity interaction was only significant in the RT analysis, and showed that familiar words produced faster RTs than novel words in both presentations, but the effect was much bigger in the first presentation. Likewise, the effect of presentation (faster RTs in the second presentation) was bigger for novel than familiar words. In the error data, a significant interaction between familiarity and richness was found (only by-participants). Further analyses revealed a semantic richness effect only for novel words. Familiar words showed ceiling effect in both conditions.

Feature recall task

The number of different features produced by each participant for each novel word was first computed. All types of different features were assigned the same weight (1) and no further distinction between the features was analyzed. As in McRae et al. (2005), features produced by participants could be classified into different categories such as taxonomic (e.g., animal, tool), functional (e.g., used for cutting), colour, size, behaviour, etc (see Appendix 6.8 for sample features produced by participant 1). However, no specific characteristics of the features were analyzed in this study. Similar criteria to that used in Experiment 8 was adopted here in order to count the number of different features per participant. For instance, synonym features were recorded identically (*is large, is big* = *is big*), quantifiers such as *generally* and *usually* were removed from participants' responses. Phrases such as *has four legs* counted as two features: *four* and *legs*. Similarly, if an object was described as *green or red*, two features were recorded (*green* and *red*).

In total, 840 correct features were collected from all 21 participants. Two different analyses were conducted: The first analysis was a direct comparison of the total number of features recalled in each condition; the second analysis included the proportion of features recalled in each condition. It is worth noting that on average participants learned more features for the novel words in rich semantics (18.0 features) than in poor semantics (9.6 features).

Table 6.6. Mean number and proportion (%) of features recalled for novel words.

	Novel rich	Novel poor
Number	6.0	3.5
SD	1.4	1.3
Proportion (%)	34	39
SD	1.7	3.0

Number of features recalled

A paired-samples t-test was conducted on the data for novel words learned with rich semantics and poor semantics. Results showed a highly significant difference between the conditions, with higher number of features recalled for rich semantics than poor semantics, $t_1(23) = 12.10$, $p < .001$; $t_2(39) = 13.49$, $p < .001$.

Proportion of features recalled

A paired-samples t-test was also conducted on the proportion of features recalled. Results showed that the proportion of features recalled was higher for poor semantics than rich semantics, $t_1(23) = 2.63$, $p = .02$; $t_2(39) = 1.52$, $p = .14$.

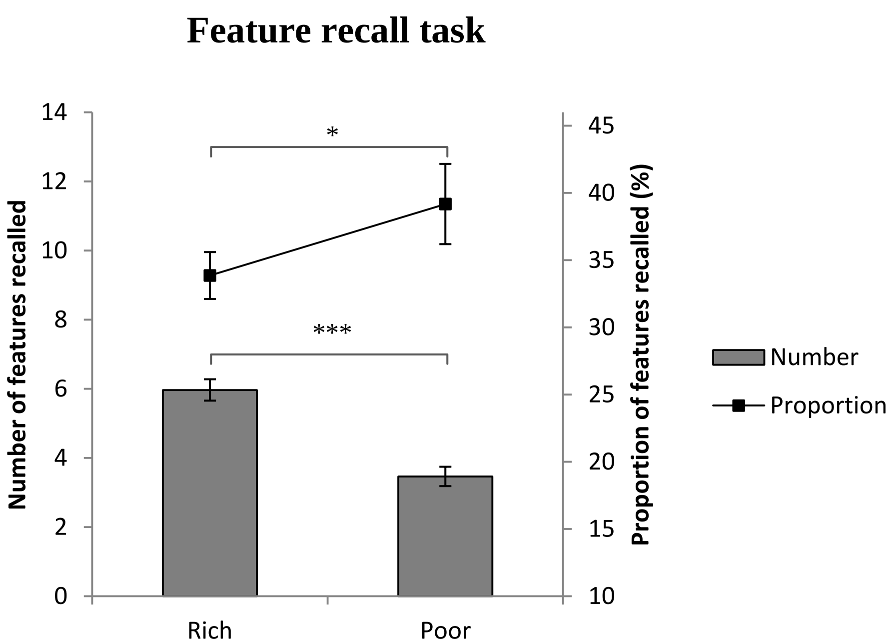


Figure 6.13. Mean number and proportion of features (%) recalled in the feature recall task for novel words. Error bars represent standard error (SE) of the mean. Statistical significance of t-tests (* = p < .001, * = p < .05).**

fMRI results

Whole-brain analysis

The whole-brain analysis included the data collected while 21 participants were performing the semantic categorization to familiar and novel words. Tables are used to display clusters of activation in descending order of size from the biggest to the smallest. Only those showing significant activation after cluster correction (Z = 2.3; *p* = .05) are included. When the distance between two peaks of activation for the same area within a cluster was more than 8mm, both peaks are displayed. When the distance was less than 8mm, only the highest peak for the given area is reported. Names of areas reported are labelled according to the Harvard-Oxford Cortical Structure Atlas and the Harvard-Oxford Subcortical Structure Atlas built into FSL view [Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library; www.fmrib.ox.ac.uk/fsl], and all coordinates are in MNI space (Evans et al., 1992).

The results included only the fMRI data for correct responses. Trials deleted in the behavioural analysis were removed for the fMRI data analysis. The maximum number of correct responses per participants was 160 across the two presentations of the stimuli. The mean number of correct trials per participants was 152 (94.8%), ranging from 143 (89.4%) to 158 (98.8%).

Contrasts of interest included familiar words versus novel words, familiar words with rich semantics (familiar rich) versus familiar words with poor semantics (familiar poor), novel words with rich semantics (novel rich) versus novel words with poor semantics (novel poor).

Familiar words > novel words

The contrast familiar words versus novel words yielded widespread bilateral activation clustered around nine different brain regions (see Table 6.7 and Figure 6.14). Clusters within frontal and prefrontal regions included (i) left frontal pole with activation extending to the adjacent superior frontal gyrus (SFG); (ii) left frontal orbital cortex, frontal pole, and temporal pole; (iii) right inferior frontal gyrus (IFG) in the pars triangularis with activation extending into the adjacent frontal orbital cortex. Clusters in the temporal lobe covered (iv) the posterior and anterior divisions of the left middle temporal gyrus (MTG), the insular cortex and the temporal pole; (v) the right posterior region of the MTG and adjacent planum polare, and the temporal pole. Clusters located in parietal and occipital regions comprised (vi) left angular gyrus (AG) with activation extending further into MTG, and lateral occipital cortex (LOC); (vii) right LOC and neighbouring portions of the MTG, AG, and anterior supramarginal gyrus (SMG); (viii) bilateral precuneus overlapping with the posterior cingulate gyrus and extending to the superior portion of the right LOC; and (ix) bilateral posterior cingulate gyrus covering portions of the right precuneus.

The areas reported above correspond roughly to the same areas in Binder et al. (2009)'s review of semantic contrasts (e.g., words versus nonwords). However, activation foci in the current study are more bilateral.

Table 6.7. MNI coordinates for peak voxels showing increased activation for familiar words versus novel words.

Brain region	Cluster size (voxels)	x	y	z	Z
L frontal pole (superior division)	10525	-4	60	18	5.37
R superior frontal gyrus		4	52	32	5.11
L frontal pole (superior division)		-12	52	28	5.05
L superior frontal gyrus		-8	52	38	4.93
R inferior lateral occipital cortex	4964	60	-62	8	4.98
R temporooccipital MTG		60	-54	10	4.92
R angular gyrus		54	-58	18	4.54
R anterior supramarginal gyrus		64	-32	34	4.45
L angular gyrus	4226	-60	-60	20	5.4
L angular gyrus		-44	-60	16	5.04
L temporooccipital MTG		-58	-58	10	4.87
L superior lateral occipital cortex		-52	-64	18	4.86
L precuneus	2248	-4	-58	32	5.02
R precuneus		8	-58	44	3.98
L precuneus		-12	-54	52	3.72
R superior lateral occipital cortex		12	-60	62	3.5
L anterior MTG	2066	-52	-10	-26	4.7
L posterior MTG		-62	-18	-18	4.61
L temporal pole		-36	6	-42	4.23
L insular cortex (*Heschl's Gyrus*)		-40	-16	0	4.03
L temporal pole		-50	4	-34	3.93
R posterior MTG	1446	46	-20	-2	3.75
R posterior MTG		58	-16	-14	3.59
R planum polare		54	-6	-6	3.52
R temporal pole		36	6	-42	3.48
R planum polare		40	-14	-14	3.45
R posterior cingulate gyrus	800	8	-28	42	3.88
L posterior cingulate gyrus		-10	-28	38	3.5
R anterior cingulate gyrus		2	-12	34	3.37
R precuneus		4	-42	52	3.16
R IFG (pars triangularis)	700	58	30	12	4.28
R frontal orbital cortex		42	26	-18	3.61
R IFG (pars triangularis)		48	28	0	3.35
R IFG (pars triangularis)		56	26	-8	3.27
L frontal orbital cortex	608	-44	22	-16	4.46
L frontal pole		-40	38	-18	3.98
L temporal pole		-32	12	-24	3.38

Note: L, left; R, right; MTG, middle temporal gyrus; IFG, inferior frontal gyrus

Novel words > familiar words

This comparison elicited five significant clusters of activation located on the left and right hemispheres, with the two largest clusters situated on the left (see Table 6.8 and Figure 6.14). Clusters situated in prefrontal and frontal cortex included (i) left frontal pole, left insular cortex, and left precentral gyrus; (ii) right insular cortex, frontal orbital cortex, frontal operculum, middle frontal gyrus (MFG), and right frontal pole. Other clusters included (iii) the left superior lateral occipital cortex (LOC) with activation extending to the left superior parietal lobule (SPL), the right cerebellum, and the left temporal occipital fusiform cortex (TOFC) in a region previously identified as the posterior visual word form area (VWFA); (iiv) right superior LOC and the right cuneal cortex, and (v) bilateral paracingulate gyrus and right anterior cingulate gyrus.

Table 6.8. MNI coordinates for peak voxels showing higher activation for novel words versus familiar words.

Brain region	Cluster size (voxels)	x	y	z	Z
L superior lateral occipital cortex	14783	-26	-70	44	5.87
L superior parietal lobule		-32	-56	38	5.4
R cerebellum		8	-78	-30	5.01
L posterior TOFC (posterior VWFA)		-42	-62	-16	4.91
L frontal pole	2657	-20	60	0	4.18
L frontal pole		-34	48	4	4.14
L insular cortex (*near FOC*)		-30	22	-6	4.05
L precentral gyus		-50	4	32	3.99
R insular cortex (*near FOC*)	2433	32	24	-2	4.93
R frontal orbital cortex		26	28	-6	4.35
R frontal operculum		36	24	4	4.23
R middle frontal gyrus		48	28	24	4.06
R frontal pole		38	42	2	3.74
R superior lateral occipital cortex	983	30	-58	40	5.36
R cuneal cortex		16	-66	38	3.19
L paracingulate gyrus	700	-2	14	42	3.97
R paracingulate gyrus		2	14	44	3.86
R anterior cingulate gyrus		12	30	24	3.43
R paracingulate gyrus		12	22	36	3.25

Note: L, left; R, right; TOFC, temporal occipital fusiform cortex; VWFA, visual word form area; FOC, frontal orbital cortex.

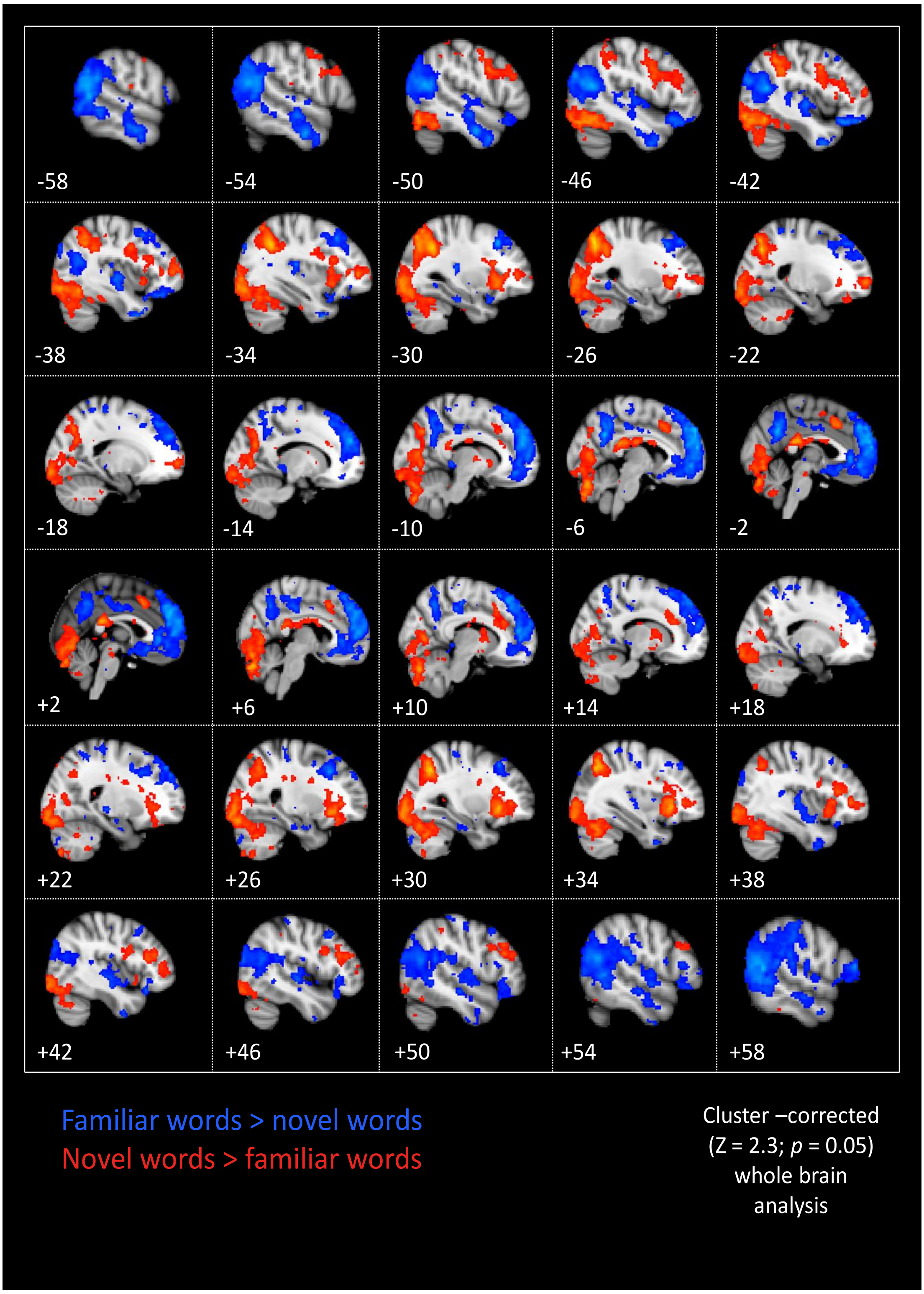


Figure 6.14. Thresholded (Z=2.3) brain images of the contrasts familiar words versus novel words (in blue) and novel words versus familiar words (in red). Areas in blue represent the nine clusters listed on Table 6.7, which included all the brain regions significantly more active for familiar words than novel words. Areas in red show the clusters on Table 6.8 and correspond to the areas that responded more strongly to novel words than familiar words. From left top corner, brain slices were taken every 4 mm starting at y = -58 and ending at y = 58.

Familiar rich > familiar poor

This contrast did not show any significant clusters of activation after correcting for multiple comparisons. However, uncorrected peaks of activation were located in bilateral frontal pole, left precuneus and posterior cingulate gyrus, left angular gyrus, right paracingulate gyrus, and right supracalcarine cortex.

Familiar poor > familiar rich

The contrast familiar poor versus familiar rich showed only one significant cluster of activation in bilateral occipital pole (see Table 6.9).

Table 6.9. MNI coordinates for corrected peak voxels showing increased brain activity for familiar poor versus familiar rich.

Brain region	Cluster size (voxels)	x	y	z	Z
L occipital pole	1485	-12	-92	-8	4.2
L occipital pole		-22	-98	-6	3.83
R occipital pole		18	-98	-12	3.32
R occipital pole		32	-92	-22	3.25

Note: L, left; R, right.

Novel rich > novel poor

The contrast novel rich versus novel poor showed increased activation in nine clusters distributed on the left and right hemispheres (see Table 6.10 and Figure 6.15). Clusters located in frontal regions included (i) right frontal medial cortex (FMC), right paracingulate gyrus, left subcallosal cortex and the left frontal pole; and (ii) left superior frontal gyrus (SFG) and left middle frontal gyri (MFG). Clusters primarily located in temporal and occipital regions included (iii) left temporooccipital fusiform gyrus, left anterior parahippocampal gyrus, and left posterior temporal fusiform cortex (TFC); and (iv) left posterior middle temporal gyrus (MTG), and superior temporal gyrus (STG). Cluster in parietal and occipital areas were found in (v) bilateral precuneus extending to the left posterior cingulate and lingual gyri; (vi) left angular gyrus (AG), bilateral superior lateral occipital cortex (LOC), and left inferior LOC; and (vii) superior LOC and right AG. Finally, two other clusters were found in (viii) bilateral cerebellum, and (ix) right cerebellum.

Table 6.10. MNI coordinates for corrected peak voxels showing increased activity for novel rich versus novel poor.

Brain region	Cluster size (voxels)	x	y	z	Z
R frontal medial cortex	3673	4	50	-8	3.98
R paracingulate gyrus		10	48	-8	3.81
L subcallosal cortex		-2	12	-6	3.77
L frontal pole		-8	56	-8	3.51
R precuneus	2767	2	-54	26	4.06
L precuneus		-8	-54	22	4.05
L posterior cingulate gyrus		-8	-48	28	4.04
L lingual gyrus		-8	-58	2	3.59
L angular gyrus	1469	-44	-62	16	3.98
L superior lateral occipital cortex		-42	-76	28	3.95
L inferior lateral occipital cortex		-52	-74	8	3.23
R superior lateral occipital cortex	1153	56	-64	32	3.81
R superior lateral occipital cortex		44	-62	26	3.55
R angular gyrus		60	-60	16	3.30
L temporal occipital fusiform cortex	697	-32	-46	-10	3.33
L hippocampus		-18	-18	-18	3.23
L parahippocampal gyrus		-20	-24	-20	3.21
L temporal occipital fusiform cortex		-30	-52	-6	3.15
L posterior temporal fusiform cortex		-28	-36	-24	3.07
R cerebellum	560	8	-58	-50	4.38
L cerebellum		-6	-60	-48	3.44
L superior frontal gyrus	518	-24	18	42	3.38
L middle frontal gyrus		-32	20	44	3.14
L posterior middle temporal gyrus	430	-60	-18	-16	3.11
L posterior superior temporal gyrus		-58	-18	-6	2.96
R cerebellum	381	34	-84	-40	3.67
R cerebellum		24	-84	-30	3.37
R cerebellum		48	-70	-40	3.02

Note: L, left; R, right.

Novel poor > novel rich

The comparison novel-poor versus novel-rich activated a large cluster that covered the left precentral gyrus and the left inferior frontal gyrus (pars opercularis). See Table 6.11.

Table 6.11. MNI coordinates for corrected peak voxels showing increased brain response for novel poor versus novel rich.

Brain region	Cluster size (voxels)	Z	x	y	z
L precentral gyrus	467	3.19	-42	6	32
L precentral gyrus		3.02	-54	4	42
L IFG (pars opercularis)		2.96	-50	16	22

Note: L, left; IFG, inferior frontal gyrus.

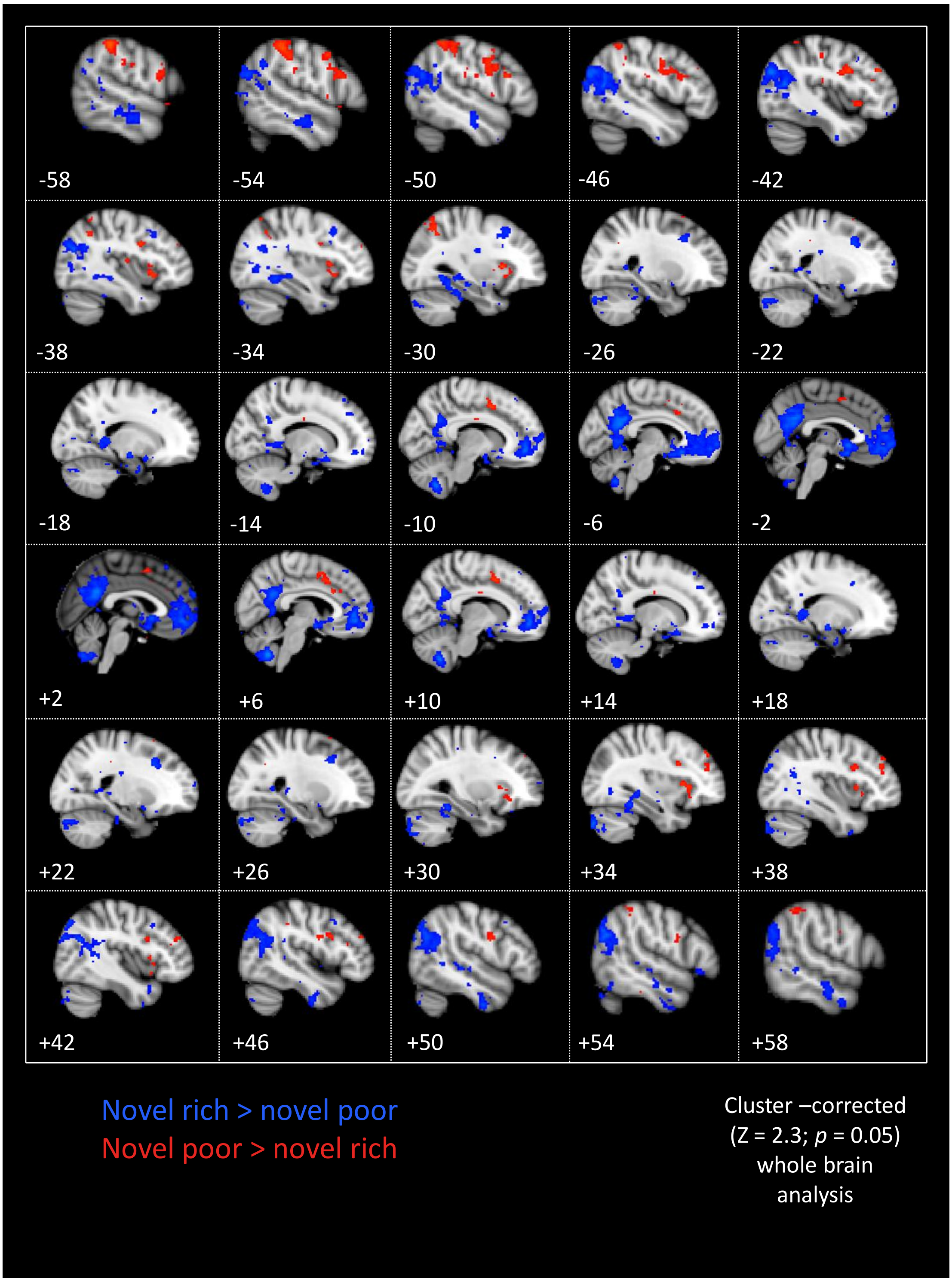


Figure 6.15. Thresholded (Z=2.3) brain images of the contrasts novel rich versus novel poor (in blue) and novel poor versus novel rich (in red). From left top corner, brain slices were taken every 4 mm starting at y = -58 and ending at y = 58.

ROI results

Percent signal change in each ROI was assessed using three-way repeated measures ANOVAs on parameter estimates (PE) obtained for each individual in each condition. The three-way ANOVA had hemisphere (left, right), familiarity (familiar, novel) and semantic richness (rich, poor) as independent factors. Two-way ANOVAs were used to break down significant or marginally significant interactions.

As mentioned in the Methods section, ROIs were selected based on coordinates from previous studies, which were localised in the left hemisphere. However, since activation in the current experiment was mostly bilateral, a mirror mask for each ROI in the right hemisphere was also included. As shown in the whole-brain analysis, a familiarity effect was found in bilateral angular gyrus and precuneus with higher activation for familiar words. The opposite contrast (novel > familiar) showed increased activation in the left precentral gyrus extending further into middle temporal gyrus and pars opercularis. The semantic richness effect was also found in bilateral angular gyrus and precuneus with higher activation for novel rich than novel poor. The reverse comparison (novel poor > novel rich) showed increased activation in the left precentral gyrus and pars opercularis. The semantic richness effect did not reach significance for familiar words after correcting for multiple comparisons, but showed the same trend as in novel words.

The following analysis includes the three ROIs selected: ventrolateral angular gyrus (vAG), the precuneus, and the pars opercularis. See Figure 6.16.

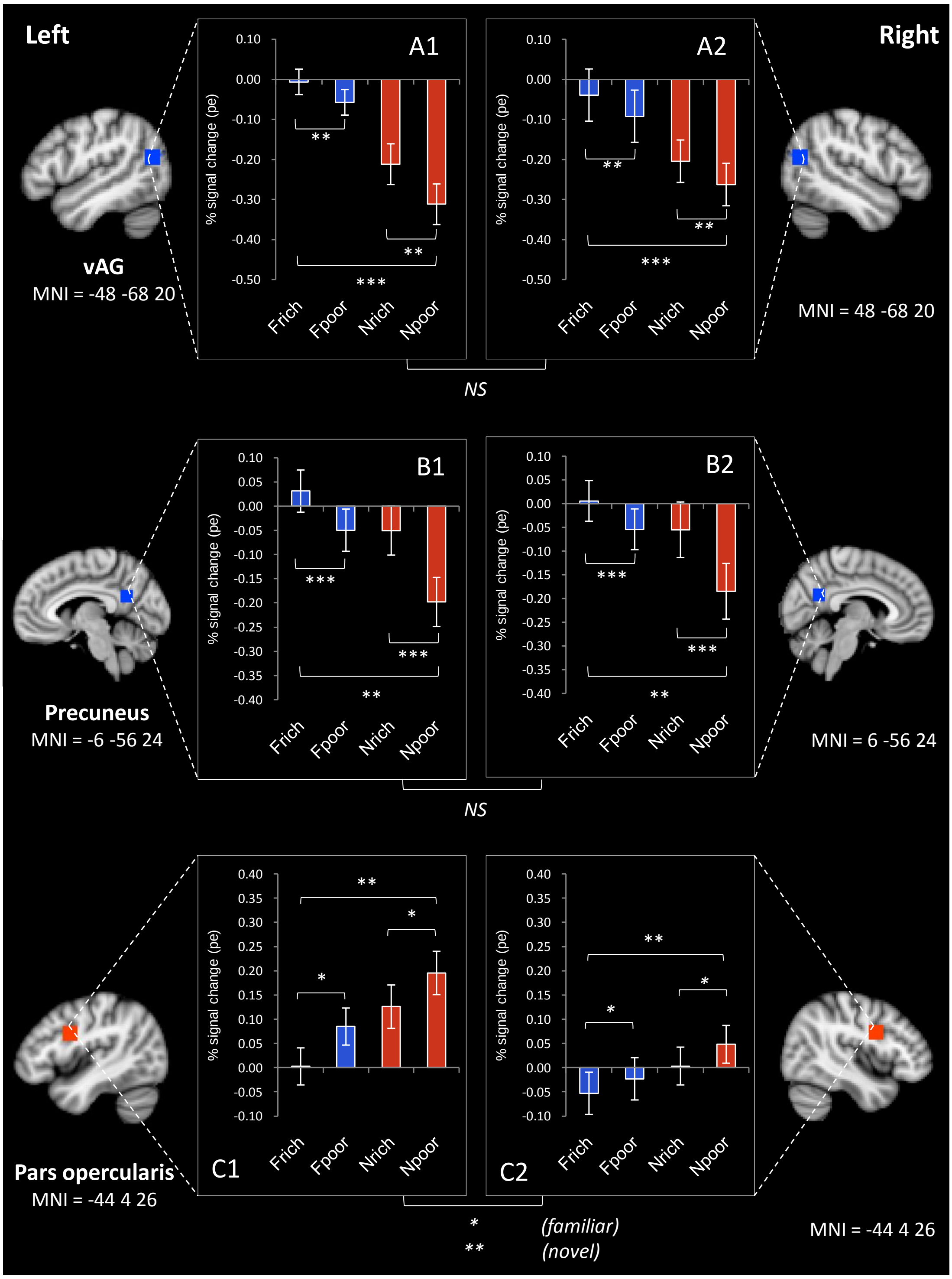


Figure 6.16. Results of ANOVAs conducted on percent signal change in three regions of interest: ventrolateral angular gyrus (vAG), precuneus, and pars opercularis. Bars in blue represent familiar words and bars in red novel words. Statistical significance of ANOVAs (**p* < .001, ***p* < .01, **p* < .05, *ns* = nonsignificant). Error bars represent 95% confidence intervals (CI).**

Ventrolateral angular gyrus (vAG)

The factorial three-way ANOVA conducted on percent signal change in the vAG region showed no main effect of hemisphere, $F_1(2, 20) = .00$, MSE = .04, $p = .95$, $\eta_p^2 = .00$. However, there was a significant effect of familiarity with higher activation for familiar words than novel words, $F_1(2, 20) = 35.88$, MSE = .05, $p < .001$, $\eta_p^2 = .64$. There was also a significant effect of semantic richness, with words with rich semantics showing increased activation in comparison with words with poor semantics, $F_1(2, 20) = 9.40$, MSE = .02, $p = .01$, $\eta_p^2 = .32$. A marginal hemisphere x familiarity interaction was found, $F_1(2, 20) = 4.13$, MSE = .01, $p = .06$, $\eta_p^2 = .17$. However, none of the other interactions showed a reliable effect: Hemisphere x semantic richness [$F_1(2, 20) = .69$, MSE = .01, $p = .42$, $\eta_p^2 = .03$], familiarity x semantic richness [$F_1(2, 20) = .38$, MSE = .02, $p = .54$, $\eta_p^2 = .02$]. The three-way hemisphere x familiarity x semantic richness interaction was not significant either, $F_1(2, 20) = 1.80$, MSE = .00, $p = .19$, $\eta_p^2 = .08$.

The marginal hemisphere x familiarity interaction was further explored by conducting two separate ANOVAs on percent signal change for familiar words and novel words with hemisphere as a factor. The ANOVA conducted on percent signal change for familiar words showed a significant effect of hemisphere, with higher activation on the left than on the right hemisphere, $F_1(2, 20) = 20.74$, MSE = .04, $p < .001$, $\eta_p^2 = .51$. However, the ANOVA conducted on novel words did not show the hemisphere effect found for familiar words, $F_1(2, 20) = .64$, MSE = .03, $p = .43$, $\eta_p^2 = .03$.

Summary

Results in the vAG region showed no overall effect of hemisphere. However, the effects of familiarity and semantic richness were significant. Increased brain activity was found for familiar words in comparison with novel words, and for word with rich semantics in comparison with words with poor semantics. There was only one marginal interaction (hemisphere x familiarity). Further analyses showed more bilateral activation for novel than familiar words. None of the other interactions showed a reliable effect.

Precuneus

The factorial three-way ANOVA conducted on percent signal change in the precuneus region showed no overall effect of hemisphere, $F_1(2, 20) = .10$, MSE = .01, $p = .78$, $\eta_p^2 = .01$. As in the vAG region, there was a significant effect of familiarity, with more activation for familiar words than novel words, $F_1(2, 20) = 9.77$, MSE = .05, $p = .01$, $\eta_p^2 = .33$; and a significant effect of semantic richness, with higher brain activity for words with rich semantics, $F_1(2, 20) = 20.59$, MSE = .02, $p < .001$, $\eta_p^2 = .51$. There was no hemisphere x familiarity interaction, $F_1(2, 20) = 1.71$, MSE = .00, $p = .21$, $\eta_p^2 = .08$. However, there was a marginal hemisphere x semantic richness interaction, $F_1(2, 20) = 3.61$, MSE = .00, $p = .07$, $\eta_p^2 = .15$. The familiarity x semantic richness interaction did not show any reliable effect, $F_1(2, 20) = 2.69$, MSE = .02, $p = .12$, $\eta_p^2 = .12$. The three-way hemisphere x familiarity x semantic richness interaction did not show significance either, $F_1(2, 20) = .02$, MSE = .00, $p = .91$, $\eta_p^2 = .00$.

In order to break down the marginal hemisphere x semantic richness interaction, two separate ANOVAs were conducted on percent signal change for words with rich semantics and poor semantics with hemisphere as the independent variable. Results for words with rich semantics showed no effect of hemisphere, $F_1(2, 20) = .80$, MSE = .01, $p = .38$, $\eta_p^2 = .04$. The same pattern of results was found for words with poor semantics, $F_1(2, 20) = .04$, MSE = .01, $p = .84$, $\eta_p^2 = .00$.

Summary

As in the vAG region, no effect of hemisphere was found, but significant effects of familiarity and semantic richness emerged. Only a marginal hemisphere x semantic richness interaction was found. Further analysis showed no hemisphere effect for words with either rich or poor semantics. No other interactions showed a reliable effect.

Pars opercularis

The factorial three-way ANOVA conducted on percent signal change in the pars opercularis region showed a significant effect of hemisphere with higher activation in the left than in the right hemisphere, $F_1(2, 20) = 7.77$, MSE = .06, $p = .01$, $\eta_p^2 = .28$. There was also a significant effect of familiarity, with increased activation for novel words in comparison with familiar words, $F_1(2, 20) = 16.23$, MSE = .02, $p < .001$, $\eta_p^2 = .45$; and a significant semantic richness effect with higher activation for words with poor semantics, $F_1(2, 20) = 7.95$, MSE = .02, $p = .01$, $\eta_p^2 = .28$. The hemisphere x familiarity interaction showed a reliable effect, $F_1(2, 20) = 4.63$, MSE = .01, $p = .04$, $\eta_p^2 = .19$. A marginal hemisphere x semantic richness interaction was also found, $F_1(2, 20) = 3.98$, MSE = .00, $p = .06$, $\eta_p^2 = .17$. However, the familiarity x semantic richness interaction was not significant, $F_1(2, 20) = .00$, MSE = .01, $p = .95$, $\eta_p^2 = .00$. The three-way hemisphere x familiarity x semantic richness interaction was not significant either, $F_1(2, 20) = .81$, MSE = .00, $p = .38$, $\eta_p^2 = .04$.

In order to breakdown the hemisphere x familiarity and the semantic richness x familiarity interactions, separate ANOVAs were conducted on percent signal change. The first interaction was explored by means of an ANOVA conducted on the left and right pars opercularis regions separately. Both ANOVAs showed a significant familiarity effect, with increased activation for novel words [left, $F_1(2, 20) = 19.42$, MSE = .02, $p < .001$, $\eta_p^2 = .49$; right pars opercularis, $F_1(2, 20) = 6.61$, MSE = .01, $p = .02$, $\eta_p^2 = .25$]. The size of the effect was bigger in the left hemisphere. Separate ANOVAs were also used to assess the effect of hemisphere on familiarity, so one analysis included familiar words and the other novel words. Both ANOVAs showed a hemisphere effect, with higher activation on the left than on the right pars opercularis [familiar words, $F_1(2, 20) = 4.71$, MSE = .03, $p = .04$, $\eta_p^2 = .19$; and novel words $F_1(2, 20) = 9.50$, MSE = .04, $p = .01$, $\eta_p^2 = .32$]. The size of the hemisphere effect was slightly bigger for novel words.

The second (marginal) interaction was further explored using ANOVAs on familiar and novel words separately with semantic richness as a factor. Results showed that the effect of semantic richness was significant for both familiar and novel words, with increased brain activity for words with poor semantics [familiar words, $F_1(2, 20) = 5.61$, MSE = .01, $p = .03$, $\eta_p^2 = .22$, and novel words, $F_1(2, 20) = 5.32$, MSE = .01, $p = .03$, $\eta_p^2 = .21$]. Separate ANOVAs were also conducted on words with rich and poor semantics with familiarity as a factor. The analyses showed a similar pattern of results in both sets, with increased activation for novel words compared to familiar words [words with rich semantics, $F_1(2, 20) = 8.44$, MSE = .02, $p = .01$, $\eta_p^2 = .30$; and poor semantics, $F_1(2, 20) = 19.24$, MSE = .01, $p < .001$, $\eta_p^2 = .49$]. The effect was bigger in the set of words with poor semantics.

Summary

Higher activation was found in the left than in the right hemisphere, for novel than for familiar words, and for poor words than for rich words. There was a hemisphere x familiarity interaction with further analyses showing a familiarity effect in both hemispheres, but with a bigger effect on the left. The hemisphere effect was found for familiar and novel words, with a slightly bigger effect for novel words. A marginal semantic richness x familiarity interaction was also found, with further analyses only showing a bigger familiarity effect for words with poor semantics. None of the remaining interactions showed a reliable effect.

6.5.3 Discussion

The current study used behavioural and fMRI protocols to investigate semantic richness effects during the processing of familiar and novel words. Existing behavioural evidence has consistently found an effect of semantic richness (number of semantic features) in semantic categorization tasks (e.g., Pexman et al., 2002; Pexman et al., 2003; Grondin et al., 2009). Experiments in the current thesis have also shown this effect for words that participants learned in a laboratory setting associated with high (rich) and low (poor) number of semantic features (NSF). However, previous neuroimaging studies of familiar and novel words have not explored this effect. This investigation proposed that two distinctive networks of areas would emerge for words with high and low number of features. Words with high number of features (high-NSF) would active semantic representation and episodic areas, whereas words with low number of features (low-NSF) would show increased activation in brain regions associated with semantic control.

6.5.3.1 Behavioural findings

Two tasks were used to assess behavioural data. The semantic categorization task, which was performed in the scanner, was investigated using a factorial ANOVA with presentation (first, second), familiarity (familiar, novel), and semantic richness (rich, poor) as the independent factors.

First, participants showed faster RTs in the second presentation of the stimuli, but error rates did not show significant variation across both presentations. The expected effect of familiarity was found across both presentations, with faster RTs and lower error rates for familiar than novel words. The significant presentation x familiarity interaction found only in the RT data, revealed that the effect of familiarity was stronger in the first than in the second presentation of the stimuli. This suggests that even though both sets of words showed an increase in performance in the second presentation, the benefit for novel words was bigger. The effect of presentation is consistent with the test-enhanced learning literature (e.g., McDaniel et al., 2007; Roediger et al., 2006) reviewed earlier, and with previous studies in this thesis which showed that participants' performance improves over time when mediated by a test. The fact that this effect was much bigger for novel than familiar words might reflect that less stable lexical representations are much harder to process without explicit memory cues. However, they can benefit substantially from a test instance since this allows the reactivation of memory traces, boosting performance in subsequent retrievals (Wixted, 2004). On the contrary, familiar words, which are well-established lexical representations, might benefit very little from retrieval or a test instance since their processing is highly fluent, even without the presentation of any contextual information.

Second, the predicted effect of semantic richness was found across RTs and error rates, with faster RTs and lower error rates for words with high number of features (rich semantics) than for words with low number of features (poor semantics). These results replicated the semantic richness effect previously found in studies of familiar words (e.g., Pexman et al., 2002; Pexman et al., 2003) and extended the findings to novel words. The latter had never been demonstrated in any previous publications and it has only been shown in previous experiments of this thesis. An interpretation for the semantic richness effect was first discussed in Chapter 2 where speakers of English as a second language showed a processing advantage for words learned with rich semantics over words learned with poor semantics in a semantic decision task (see Methods section of Chapter 2). A classic explanation for the semantic richness effect is that high-NSF words have rich semantic representations as opposed to low-NSF words, which have poor semantic representations. Hence, when participants perform a task that involves semantic processing, high-NSF words produce more semantic activation than low-NSF words (because of the extra features). This additional semantic activation allows faster mapping between semantics, orthography and phonology, which is translated into faster responses (e.g., Hino & Lupker, 1996; Pexman & Lupker, 1999; Pecher, 2001; Pexman et al., 2003). A similar explanation has been offered in the computational literature. Plaut and Shallice (1993), based on simulations with models of semantic memory, demonstrated that the system can settle faster into a more stable pattern of activation when concepts have richer representations. They argued that when semantic representations contain more semantic units, the model builds stronger attractors for those concepts in semantic space, which allows for more efficient semantic processing.

The second behavioural task was a feature recall task and only included novel words since it aimed at assessing the learning of semantic features during the two days of training. Results showed that participants recalled significantly more features in the rich semantics condition than in the poor semantics condition. This suggests that participants in the rich condition not only were exposed to more features during training, but they also learned more features. This result has important implications for subsequent analyses because it allows treating both familiar and novel words in the same way, regarding the number of semantic features associated with their corresponding lexical representations. Even though there was an advantage for high-NSF novel words with respect to the number of features participants recalled after scanning, the proportion of features recalled was slightly higher for low-NSF words. This might be simply because in the poor semantics condition participants had to recall fewer features, so less demand on memory resources was expected than in the rich semantics condition.

To sum up, the current behavioural results have replicated previous findings in familiar and novel words regarding the semantic richness effect. Additionally, they have provided a measure of the number and proportion of features participants can recall after learning novel words with high and low number of features. Taken together, these findings provide a good behavioural basis to examine the neural correlates of semantic richness.

6.5.3.2 fMRI findings

The current investigation used an event-related fMRI protocol to explore the neural correlates of semantic richness for familiar and novel words. Results were analyzed using whole-brain and region-of-interest (ROI) analyses.

First, it was proposed that high-NSF words would show more activation than low-NSF words in semantic representation and episodic memory areas. Hypothesised conceptual or semantic representation areas included anterior temporal poles (e.g., Nestor et al., 2006; Jefferies & Lambon Ralph, 2006; Patterson et al., 2007), ventrolateral middle temporal gyrus (e.g., Binder et al., 2003; Binder et al., 2009; Vigeneau et al., 2006), and ventrolateral angular gyrus (e.g., Cabeza & Neiberg, 2000; Seguier et al., 2010; Binder et al., 2009; Vigneau et al., 2006). Episodic memory areas comprised posterior cingulate gyrus and the precuneus (e.g., Binder et al., 2009; Epstein et al., 2007; Vincent et al., 2006). These areas have been proposed to show increased activation when meaningful stimuli are processed, so they usually couple with conceptual representation areas during semantic processing (e.g., Binder et al., 2009).

Second, it was also proposed that activation would increase in semantic control areas during the processing of low-NSF words in comparison with high-NSF words. Even though previous semantic studies have outlined more than one semantic control areas (e.g., Whitney et al., 2010), the current study only focuses on inferior prefrontal cortex, whose role in regulating semantic activation during retrieval is well-established (e.g., Wagner et al., 2001; Badre et al., 2005).

The whole-brain analysis included the following contrasts: familiar words versus novel words, high-NSF familiar words (familiar rich) versus low-NSF familiar words (familiar poor), and high-NSF novel words (novel rich) versus high-NSF novel words (novel poor). The ROI analyses matched the behavioural data on the independent factors of familiarity and semantic richness, but did not include presentation as a factor due to the fact that the data from both presentations were merged to increase statistical power. Before discussing the results regarding semantic richness, a report and discussion of brain regions involved in the familiarity effect is also presented.

Familiarity effect

The aim of this analysis was to replicate previous findings in studies of verbal semantics and/or word learning, particularly those that have compared words versus novel words/nonwords. It is worth noting that no previous studies have compared familiar and novel words that have been learned with meaning over extensive training sessions. Previous word learning studies have assessed familiarity as participants learned novel words without time for consolidation (e.g., Breitenstein et al., 2005; Mestres-Misse et al., 2008), or have compared meaningless novel words with familiar words (e.g., Davis et al., 2008).

The comparisons of familiar versus novel words revealed two very distinctive networks of brain areas. Familiar words showed increased activation in nine clusters that included bilateral areas in frontal, temporal, parietal, and medial regions (see Table 6.7). These results were very consistent with previous functional neuroimaging studies that have compared words versus nonwords or semantic tasks versus nonsemantic tasks, as presented in Binder et al. (2009)'s review. Results were also consistent with a number of spoken word studies published in 2009, which were reviewed by Price (2010). For instance, a study that compared spoken sentences relative to spectrally rotated words found increased activation in the angular gyrus (AG) when participants heard the intelligible speech in comparison with the rotated words (e.g., Obleser & Kotz, 2009). In another study, Davis and Gaskell (2009) contrasted spoken words with pseudowords and showed that familiar words activated bilateral anterior middle temporal cortices, posterior temporal parietal cortices, and the precuneus, with left-lateralized activation in the temporal pole, and posterior middle temporal cortex.

It is worth noting that activation in the current study was mostly bilateral (e.g., angular gyrus, anterior temporal pole, posterior middle temporal gyrus, precuneus, posterior cingulate gyrus). However, in most previous studies activation has been largely located in left-lateralised areas (see Binder et al., 2009; Vigneau et al., 2006). As discussed earlier, some of the areas reported in the studies above correspond to regions classified here as semantic representation areas (e.g., anterior temporal pole, middle temporal gyrus, angular gyrus) or episodic memory areas (e.g., posterior cingulate gyrus and the precuneus). Thus, they are also expected to show increased activation in contrasts of high-NSF words versus low-NSF words (this will be discussed in the following paragraphs). The factorial ANOVA (factors included hemisphere, familiarity, and semantic richness) conducted on signal change in ventral angular gyrus (vAG) showed that activation was bilateral for novel words, but it was higher on left vAG than on right vAG for familiar words. However, in the precuneus region the effect of hemisphere was not significant and no hemisphere x familiarity interaction was found, which showed that activation was bilateral. These results are consistent with previous studies that have reported left-lateralised activation in most regions during the processing of familiar words (e.g., Binder et al., 2009; Vigneau et al., 2006). Unlike familiar words, novel words seem to be less hemisphere specific, which might be due to the extra effort involved in the processing of newly learned lexical representations.

The contrast novel words versus familiar words mainly produced activation in bilateral occipital cortex, left temporal occipital fusiform cortex (VWFA), left inferior frontal cortex (IFC), and motor cortex (see Table 6.8). In the ROI analysis, the factorial ANOVA conducted on signal change in the pars opercularis region of the IFC showed significantly higher activation for novel than familiar words. Activation was also higher in the left in comparison with the right pars opercularis. The hemisphere x familiarity interaction showed that the effect of familiarity was present in both hemispheres but it was stronger in the left than in the right. The hemisphere effect was also significant in both familiar and novel words with a slightly stronger effect for novel words. In summary, activation in IFC seems to be stronger in the left hemisphere, with novel words showing higher activation than familiar words across both hemispheres. Differences between novel and familiar words are especially enhanced in the left IFC, and novel words seem to show greater differences in activation across hemispheres.

Previous studies of visual lexical decision and reading have reported heightened activation for pseudowords in comparison with familiar words in left inferior frontral gyrus and precentral gyrus (e.g., Binder et al., 2003; Mechelli, Gorno-Tempini, & Price, 2003), in the visual word form area (e.g., Bruno, Zumberge, Manis, Zhong-Lin Lu, & Goldman, 2001; Mechelli et al., 2003), and different locations within the occipital cortex (e.g., see Mechelli et al., 2003, for review). This suggests that the current results are highly consistent with previous studies even though a different task was used (semantic categorization). Areas that are more activated for pseudowords than words are generally involved in a type of processing that is not semantic. This is the case of the VWFA, which responds uniquely to words or word-like pseudowords with enhanced activity for pseudowords (e.g., Cohen, Lehericy, Chochon, Lerner, Rivaud, & Dehaene, 2002), so it seems to be involved in orthographic rather than semantic processing. However, the role of inferior prefrontal cortex in less clear because it has been linked to phonological processing (e.g., Myers, Blumstein, Walsh, & Eliassen, 2009) and semantic control (e.g., Badre et al., 2005; Whitney et al., 2010), as reviewed earlier. In a contrast of novel versus familiar words, the difference in activation might well be attributed to both phonological and semantic processing due to the fact that novel words in this study also have a semantic representation, which differs from studies using unfamiliar nonwords. However, it is impossible to claim a phonological or semantic control role based uniquely on the contrasts novel versus familiar. The role of inferior frontal cortex will be further assessed in subsequent comparisons assessing semantic richness.

Semantic richness effect

It was predicted that a semantic richness effect would be observed in two different brain networks: one including semantic representation (anterior temporal pole, posterior middle temporal gyrus, and angular gyrus) and episodic memory (precuneus and cingulate gyrus) areas, and the other semantic control areas (prefrontal cortex). The whole-brain analyses included separate semantic richness contrasts for familiar and novel words.

In line with predictions, comparisons between rich and poor semantics represented in the contrasts familiar rich versus familiar poor and novel rich versus novel poor revealed a consistent pattern of results. The cluster-corrected general linear model analysis for the contrast novel rich versus novel poor showed significant activation in all the areas where a semantic richness effect was expected, except in the anterior temporal poles (see Table 6.11). However, in the contrast familiar rich versus familiar poor none of the clusters identified survived cluster correction for multiple comparisons, even though activation was in the direction of the predictions. In line with the whole-brain analysis, the ROI analysis showed that high-NSF words (rich) showed increased signal change in vAG and the precuneus in comparison with low-NSF words (poor). Moreover, no significant semantic richness x familiarity interaction was found, which confirms that familiar words also showed a reliable effect of semantic richness in vAG and the precuneus, even though this did not reach significance in the whole-brain analysis.

In the anterior temporal poles, uncorrected peaks of activation for both contrasts (familiar and novel) were found within the ventro-rostral region. Significant activation for the contrast novel rich versus novel poor was found in areas that are normally activated together with the anterior temporal pole such as posterior middle temporal gyrus (pMTG) and left fusiform gyrus. It was discussed earlier that signal from the anterior temporal poles is hard to pick up using standard fMRI procedures (e.g., Visser et al., 2010; Binney, Embleton, Jefferies, Parker, Lambon Ralph, 2010), and this is confirmed in Binder et al. (2009)'s review which did not explicitly list foci of activation for this region when averaging activation from over 100 studies. In line with the current data, they did find activation in fusiform gyrus and posterior and anterior MTG. Even though the current experiment did not use distortion-corrected fMRI, activation in the anterior temporal poles was found for the contrast familiar versus novel, so it cannot be claimed that the lack of effect here was only due to failure of the current protocol to pick up the signal. However, the comparison novel rich versus novel poor contained only half of the trials (novel words only) than in the comparison familiar versus novel, so if the signal identified in the temporal lobes was weaker than the actual signal, this might have affected contrasts with fewer trials due to lack of statistical power. This is supported by evidence from a study conducted by Binney et al. (2010), which used spatial remapping correction in order to transform the originally distorted data into distortion-corrected data (e.g., see Visser et al., 2010b, for details). The study compared activation for a semantic task and a numerical task. They found a number of clusters within the anterior temporal lobes including a ventral cluster peaking in the anterior fusiform gyrus, and another in the anterior superior temporal gyrus. The posterior MTG was also activated with spreading activation reaching the posterior portion of the temporal pole. These findings suggest that activation in the temporal pole might fail to reach significance using conventional fMRI analyses, but might do so if distortion-corrected techniques are used. An alternative explanation might be that the current data lacked statistical power for activation in the anterior temporal pole to reach significance after the data were cluster-corrected for multiple comparisons.

In line with predictions, the posterior MTG showed increased activation for novel rich compared to novel poor. A sizeable left-lateralised cluster was found, with activation also extending towards the anterior portion and into the posterior STG. This supports the evidence that suggests this area is involved in semantic representation (e.g., Binder et al., 2009; Binney et al., 2010; Visser et al., 2010a), and shows increased activation when novel words are associated with high number of semantic features. It is important not to confuse this area with a much more posterior MTG location reported in previous studies (e.g., Whitney et al., 2011). The posterior MTG in these studies corresponds to an area located near the junction with the occipital cortex and the angular gyrus (MNI = -56 -50 3, in Whitney et al., 2011), and has been recently linked to semantic control (e.g., Whitney et al., 2011a; 2011b). In the current study, the highest peak of activation was found at MNI -60 -18 -16, so it was much more anterior than in Whitney et al.'s study. In order to avoid this confusion, some researchers refer to the area within the peaks in the current study as ventrolateral temporal cortex and only used posterior MTG for the much more posterior portion of MTG (e.g., Whitney et al., 2011b), which according to the Harvard-Oxford Cortical Structure Atlas [Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library; www.fmrib.ox.ac.uk/fsl], corresponds to the temporooccipital MTG. There is evidence suggesting that posterior middle temporal areas implicated in semantic control are dissociable from ventrolateral temporal cortex, which have been associated with semantic representation (e.g., Sharp et al., 2004; Pobric et al., 2007; Lambon Ralph et al., 2009; Whitney et al., 2011b). This fits well with the current findings since novel words with rich semantics are assumed to have a more well-established representation than novel words with poor semantics which is reflected in significantly more activation in left ventrolateral MTG.

The third semantic representation region that was predicted to show increased activation for rich versus poor was the angular gyrus (AG). The comparison novel rich versus novel poor showed bilateral activation centred around the ventrolateral region and extending further into the superior division of the lateral occipital cortex. Roughly the same location showed increased activation for familiar rich versus familiar poor, but did not reach significance in the whole-brain analysis. This was further explored in the ROI analysis which assessed signal change in ventrolateral angular gyrus (vAG). Results showed higher signal change for trials corresponding to high-NSF words than low-NSF words bilaterally, and no semantic richness x familiarity interaction was found, which suggests that the effect of semantic richness in the vAG region was present in both familiar and novel words.

Previous functional neuroimaging studies have suggested that semantic activation tends to be left-lateralised, but the AG is one of the areas that generally shows bilateral activation along with posterior cingulate gyrus, and the precuneus (Binder et al., 2009), which is consistent with the current finding. As reviewed earlier, the role of this region in semantic processing is well-established since it is the most consistently activated region in studies of semantic contrasts (e.g., for meta-analysis reviews, see Cabeza & Nyberg, 2000; Vigneau et al., 2006; Binder et al., 2009). Seguier et al. (2010), in a study that involved a subdivision of the AG into medial, ventrolateral and dorsal, found that ventrolateral AG (vAG) responded uniquely to semantic tasks and was associated with later stages of conceptual identification, whereas dorsal AG was more involved in the search for a semantic representation. This suggests that within AG, the ventrolateral region might be the strongest candidate for a role in semantic representation while dAG might support a bottom-up network which has been linked to meaning retrieval (Whitney et al., 2009).

Along with semantic representation areas, two episodic memory areas were also predicted to show heightened response for rich versus poor. A significant bilateral cluster was found for the comparison novel rich versus novel poor, with the highest peak of activation in the precuneus and extending further into the posterior cingulate gyrus, and the lingual gyrus. The same comparison in familiar words showed a similar pattern of activation but again did not survive cluster correction for multiple comparisons. The ROI for this cluster was centred in the precuneus near the junction with the posterior cingulate gyrus. The results of the ANOVA extended the findings of the whole-brain analysis, showing a reliable effect of semantic richness with higher signal change for high-NSF words than low-NSF and no effect of hemisphere or familiarity x semantic richness interaction. This suggests that activation in the precuneus and posterior cingulate region is bilateral, and that familiar and novel words show a significant effect of semantic richness with increased activation for high-NSF words.

The above results are in line with predictions and suggest that episodic memory areas might work in concert with semantic representation areas, particularly during the processing of stimuli with rich semantics. This is consistent with previous studies of episodic memory retrieval which have shown increased activation in these areas for stimuli previously encoded in deep processing tasks (e.g., abstract-concrete) versus shallow processing tasks (e.g., uppercase-lowercase) (e.g., Yonelinas, 2002; Shannon & Buckner, 2004). In the current experiment, novel words learned with many semantic features were encoded under conditions of deeper semantic analyses in comparison with novel words with few features. It can then be suggested that high-NSF novel words might create stronger memory traces than low-NSF novel words, which is reflected in higher activation in the precuneus and posterior cingulate gyrus during retrieval. This proposal is further supported by the significant increased activation also found in left hippocampus for novel rich versus novel poor. The left hippocampus is a key area in the encoding of episodic memories for novel words (e.g., Gaskell & Ellis, 2009; Breitenstein et al., 2005; Davis & Gaskell, 2009), and has dense reciprocal connections with the precuneus and posterior cingulate gyrus (e.g., Vogt, Finch, & Olson, 1992). This suggests that the precuneus and posterior cingulate gyrus might receive feedback from the hippocampus during the encoding of new memories and during retrieval of episodic information. Thus, if encoding involves higher semantic analysis, more activation in the episodic network is expected during retrieval.

All the areas discussed above showed increased activation when comparing rich versus poor semantics and were hypothesised to be involved in either semantic representation or episodic memory. However, semantic control areas were expected to show a reverse pattern of activation because of their engagement during effortful semantic retrieval and apparent lack of involvement in conceptual representation. Hence, increased activation for the comparison poor versus rich was expected in inferior prefrontal cortex. This was confirmed in the whole-brain analysis for novel poor versus novel rich, with a significant cluster in the left inferior frontal gyrus (pars opercularis), extending further into precentral and middle frontal gyrus. Peaks of activation were also found in the right hemisphere but did not reach significance. The contrast familiar poor versus familiar rich showed a similar pattern of results, but clusters did not reach significance. These results were further assessed in the ROI analysis. The ANOVA conducted on signal change in the pars opercularis revealed higher signal change for trials of low-NSF words in comparison with trials of high-NSF words. Unlike other ROIs, the effect in pars opercularis was only found in the left hemisphere. No familiarity x semantic richness interaction was found, which suggests that the effect was the same across familiar and novel words. As reviewed earlier, the role of left IFC in regulating activation during semantic retrieval is well-established (e.g., Thompson-Schill et al., 1997; Rodríguez-Ferreiro et al., 2010; Wagner et al., 2001; Badre et al., 2005). The current findings support this assumption since more activation was found when participants categorized words associated with low number of features. The processing of low-NSF words might involve more effort than that of high-NSF words increasing the demand on the semantic control system and consequently producing more activation in IFC.

As mentioned in the Methods section, the coordinates (MNI -44 4 26) for the pars opercularis in the ROI analysis were taken from a study conducted by Pexman et al. (2007) which found increased activation for words with low number of semantic associates in comparison with words with high number of associates. Pexman and colleagues linked this activation to more effortful and extensive lexical and semantic processing during the categorization of words with low number of semantic associates. Another study conducted by Badre et al. (2005), which presented participants with a target and a congruent or an incongruent feature, showed increased activation for incongruent versus congruent trials in left mid- and posterior ventrolateral inferior prefrontal cortex. Peaks of activation in the posterior region were centred at MNI -45 9 27, which correspond roughly to the same location of the mask for the pars opercularis in the current experiment. Badre et al. concluded that different control mechanisms operate in the ventrolateral inferior prefrontal cortex which contribute to guiding access to semantic knowledge that is not retrieved automatically. Other studies also suggest a role of IFC when the association between two stimuli is weak (Norman & Shallice, 1986; Miller & Cohen, 2001). This has been linked to a greater need for top-down mechanisms to guide control access during retrieval of a semantic representation or the association between two concepts. In the current study, low-NSF words might need more conscious effort during semantic categorization than high-NSF words, which increases the demand for involvement of top-down mechanisms in order to guide the association of the target concept and its corresponding category.

6.5.3.3 Summary and conclusion

The results of the current investigation have shown that behavioural findings regarding effects of familiarity and semantic richness on semantic categorization are represented in dissociable brain regions. The current experiment is the first to explore the neural correlates of semantic richness for familiar and novel words using a featural approach. The fMRI analysis showed that high-NSF words compared to low-NSF words activate semantic representation areas (MTG, and AG) and episodic memory areas (posterior cingulate gyrus and precuneus), whereas low-NSF words showed increased activation in semantic control areas, particularly in IFC (pars opercularis). It can be proposed that high-NSF words have much richer semantic representations than low-NSF words which allow more fluent processing during retrieval. Thus, rich semantic representations are reflected in more widespread activation in brain regions associated with conceptual representation, whereas poor semantic representations are reflected in greater activation in semantic control areas as processing becomes less fluent.

The findings of the current investigation support models of distributed feature representations, which assume that concepts are made up of features represented across different modalities (e.g., McNorgan, Kotack, Meehan, & McRae, 2007). These models also assume a hierarchical organization of conceptual knowledge that includes one or different ‘convergence zones’ where features from one modality (e.g., action) or several modalities (e.g., shape, colour) are bind together in order to build conceptual representations (e.g., Damasio, 1989; McNorgan et al., 2011). This implies that as more features bind together, concepts acquire more abstract and richer semantic representations, which is consistent with the current findings. Feature representation models that incorporate more than one convergence zone for the role of semantic integration are usually called ‘deep’ models as opposed to ‘shallow’ models, which include no convergence zones or only one as the distributed-plus-hub model proposed by Patterson et al. (2007). At a neural level, Patterson et al. (2007)’s model proposes the anterior temporal lobe as the only convergence zone where features from different modalities represented in different brain regions (e.g., motor cortex, IFC) converge in order to build up concepts. The current investigation suggests that more than one neural areas might have a role in conceptual representation (e.g., angular gyrus, verntrolateral middle temporal gyrus), as these regions seem to show increased activation during the processing of words with rich semantic representations.

Chapter 7 – Thesis summary and conclusions

The behavioural experiments in this thesis have investigated the influence of context variability, feature variability, and number of semantic features on word learning using a variety of tasks which included naming, recognition memory, semantic decision/categorization, and cued recall/word production. Chapter 1 presented a theoretical account of previous word learning studies including a wide range of paradigms, and a review of semantic memory focused primarily on the featural approach. Chapter 2 investigated effects of context variability and feature variability on word learning in a second language. Chapter 3 investigated feature variability across speakers of English as a first and a second language. Chapter 4 examined semantic richness effects during the processing of newly learned words, and effects of consolidation over time. Chapter 5 assessed whether consolidation over time was due simply to the passing of time or the effect of previous test instances. Finally, Chapter 6 explored the neural correlates of semantic richness for familiar and novel words and drew a line between conceptual representation and semantic control areas of the brain.

7.1 Summary of main findings

Chapter 2 of this thesis investigated word learning in English as a second language. The chapter had two main aims. Experiment 1 investigated to what extend context variability (the number of different sentence contexts novel words appeared in during training) would affect the acquisition of novel words. It predicted that novel words appearing in 12 different sentences during training would show an advantage in naming and semantic decision in comparison with novel words that were only presented in 2 different sentences. Results confirmed these predictions showing a significant context variability effect with an advantage for novel words learned with high context variability. This finding confirmed that participants could easily learn new vocabulary from sentence contexts, consistent with a number of previous studies (e.g., Jenkins et al., 1984; Nagy et al., 1987; Hirsh & Nation, 1992; Lawson & Hogben, 1996). Furthermore, it supported previous findings showing that context variability can benefit word learning (e.g., Dempster 1987; Bjork, 1979;

Cain, 2007; Bolger et al., 2008). More importantly, it showed that context variability can affect word naming (in a second language), with faster RTs for novel words learned with high context variability. A possible confound in this study was *attention* since novel words experienced in many contexts might force individuals to increase attentional resources in comparison with novel words learned in only 2 contexts. In order to avoid this confound, Experiment 2 explored feature variability (the number and type of different features words are presented with during training). This experiment predicted that participants would learn better when presented with core semantic features (as those found in dictionary definitions) plus contextual features than contextual features alone. This was based on a series of studies that had shown people learn better when presented with dictionary definitions and contexts than contexts alone (Beck et al., 1987; Fischer, 1994; Nist & Olejnik, 1995; Bolger et al., 2008). Results showed that feature variability affected semantic decision and cued recall, with better performance for words learned with core semantic features plus contextual features than contextual features alone. However, no feature variability effects were found in word naming and recognition memory. Results in semantic decision and cued recall tasks supported previous studies that had found an advantage for words learned in contexts plus definitions versus contexts alone (e.g., Nist & Olejnik, 1995; Bolger et al., 2008). The fact that the gains in meaning were not reflected in the word naming task was interpreted as lack of semantic involvement in reading aloud novel words with regular spelling, consistent with previous studies of reading aloud (e.g., McKay et al., 2008; Nation et al., 2007).

Chapter 3 mainly explored cognitive differences between monolinguals and bilinguals regarding the acquisition and processing of novel words. This was investigated due to strong evidence supporting the idea that monolinguals and bilinguals differ in word processing and that competition of lexical representations in the nontarget language (for bilinguals) seems to interfere with the processing of target words (e.g., De Groot et al., 2000; Dijkstra et al., 1998; Poulisse & Bongaerts; 1994; Poulisse, 1999). It was predicted that differences between monolinguals and bilinguals would be more noticeable in naming and cued recall tasks than in recognition memory and semantic decision. Results largely confirmed the predictions showing no overall differences in performance between the groups in semantic decision and recognition memory, however, naming and particularly cued recall showed better performance in the L1 group. This suggests that L1 and L2 speakers learned new words equally well regarding meaning, but differ when performing tasks that demand direct recognition (naming) and production (cued recall) of the novel words. This might be explained by interference of lexical representations from the nontarget language becoming activated as L2 speakers attempt to recognize or produce novel words. This is consistent with the Bilingual Interactive Activation Model (BIA) of Dijkstra et al. (1998), which assumes that lexical candidates from the target and the nontarget language become activated during bilingual word processing. The difference in performance here was attributed to a higher number of lexical representations activated in the bilingual group, which would delay word recognition and production in comparison with monolinguals. Another important finding worth considering was the fact that only the L1 group showed feature variability effects on word recognition, with faster RTs for words learned with core plus contextual features. This was further investigated in Chapter 4 where effects of semantics on recognition memory were assessed over a longer period of time.

Chapter 4 explored the learning and consolidation of novel words over time. This research aim was motivated by the fact that a number of word learning studies have reported that novel words need time for consolidation in long-term memory (e.g., Gaskell & Dumay, 2003; Dumay et al., 2004; Dumay & Gaskell, 2007; Leach & Samuel, 2007; Davis et al., 2008). Consolidation has been investigated using explicit and implicit tasks with results showing improvement in performance when words are tested one day after training (e.g., Davis et al. (2008) or even several days after initial training (e.g., Gaskell & Dumay, 2007). Experiments in this Chapter wanted to test the assumption that improvement in performance over time for explicit tasks (e.g., recognition memory) would depend on the level of processing with which words were encoded during training (e.g., Craik & Lockhart, 1972). Thus, if encoding in a perceptual fashion (without meaning), participants' performance was expected to be lower than if encoded with semantics. Furthermore, if words were encoded with rich semantics, performance was expected to be better than during learning with poor semantics. In Experiment 4, results in the recognition memory task were consistent with the predictions and showed no difference in performance on day 3 for conditions with rich, poor, or no meaning. However, this difference was significant on day 8, with performance improving significantly for the rich semantics condition, but not for the other two conditions. This finding was consistent with the levels-of-processing literature (e.g., Craik & Lockheart, 1972; Craik, 2002), which suggests stimuli are better encoded when undergoing deep semantic processing. Results in semantic categorization and word production showed an advantage for the rich semantics conditions on day 3 and day 8 and performance improved equally over time. Performance in the reading aloud task did not show any effect of conditions or day. Experiment 5, which only tested the two conditions with semantics failed to find a difference between the conditions regarding recognition memory, but findings in semantic categorization and word production again showed better performance for the rich semantics condition. In summary, Chapter 4 showed that semantics affected recognition memory, but only a week after training. Conditions with rich and poor semantics did not seem to differ; however, conditions with rich semantics seemed to show more improvement over time (Experiment 4), even though this was not replicated in Experiment 5. Word naming doesn't seem to be affected by semantics or the passing of time. Finally, semantic categorization and word production showed a reliable effect of semantic richness with better performance for words learned with rich semantics. Since participants were tested on the same words on day 3 and then retested on day 8, it was noticed that the improvement over time in most tasks could probably be due to testing and not simply to the passing of time.

Chapter 5 then examined whether the effects of consolidation over time in explicit tasks observed in the two experiments of Chapter 4 was due simply to the passing of time or to the effect of the first test. There is consistent evidence that suggests a test instance can benefit learning significantly (e.g., McDaniel & Manson, 1985; Tulving, 1967; Wheeler et al., 2003; Roediger & Karpicke, 2006). However, in preceding word learning studies this has been somehow ignored (e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). In order to tease apart the effects of the passing of time and those of the test instance, two more experiments were conducted. In Experiment 6, a group of participants were asked to learn words with rich semantics and no semantics and they were tested on day 2 on all novel words, and then retested on day 8 on all same words. In Experiment 7, another group of participants learned the same words in the same conditions (rich semantics, no semantics), but were tested on half of the words on day 2 and half of the words on day 8. Results of these experiments showed that the improvement over time seen in previous experiments was mainly due to the first testing session and not simply due to the passing of time. While participants in Experiment 6 showed improvement over time in all tasks, in Experiment 7 there was a subsequent drop in performance that affected all tasks. These results were consistent with the test-enhanced learning literature which suggests a test instance can benefit performance on subsequent stimulus retrieval (e.g., Roediger & Karpicke, 2006).

All five previous chapters reported behavioural experiments in which participants learned novel words in different conditions manipulating context variability, feature variability, and number of semantic features. Since the findings of these experiments had consistently shown a semantic richness effect, particularly in semantic categorization, Chapter 6 aimed to explore these behavioural findings at a neural level using functional magnetic resonance imaging (fMRI). The Chapter comprised two experiments. The main purpose of Experiment 8 was to collect semantic features from native speakers of British English for a hundred familiar words. Previous feature production norms had only been collected in North America (e.g., McRae et al., 2005; Vinson & Vigliocco, 2008), so it was thought that these might not reflect the conceptual knowledge of British speakers. As a result of Experiment 8, 40 familiar words with high (18 on average) and low (9.6 on average) number of features were selected in order to be used as stimuli for Experiment 9. These 40 words showed a high correlation regarding the number of features in the study by McRae et al. (2005) conducted in North America and the results of Experiment 8. However, some concepts showed great disparity in the number of features listed by North American and British participants.

Experiment 9 was a combined behavioural and fMRI Experiment aiming at identifying the neural correlates of familiar and novel words with high and low number of semantic features. Previous findings regarding the number-of-features effect in familiar words (e.g., Pexman et al., 2002; 2003; Grondin et al., 2006) and novel words in the current thesis were expected to be replicated. At a neural level, high-NSF words were expected to activate brain regions associated with semantic representation (e.g., angular gyrus) and episodic memory (e.g., the precuneus) (e.g., Binder et al., 2009; Vigneau et al., 2006), whereas low-NSF words were predicted to show heightened response in areas associated with semantic control (e.g., inferior frontal gyrus) (Badre et al., 2005; Whitney et al., 2011b). Behavioural results replicated previous findings showing a processing advantage for high-NSF words in comparison with low-NSF words in the semantic categorization task performed in the scanner. This confirmed the assumption that high-NSF words have richer semantic representations than low-NSF words, which generates stronger activation during semantic categorization, speeding up RTs (e.g., Hino & Luper, 1996; Pecher, 2001; Pexman et al., 2003). The neuroimaging data largely confirmed the predictions, with high-NSF words showing increased bilateral activation mainly in the angular gyrus and the precuneus, while low-NSF words activated the left pars opercularis. These results were interpreted as evidence for the existence of two different semantic systems (a semantic representation and a semantic control system), consistent with proposals from previous studies (e.g., Jefferies & Lambon Ralph, 2006; Whitney et al., 2011b). Furthermore, they showed, for the first time, dissociation in the neural representation of high-NSF and low-NSF familiar and novel words. Regarding the familiarity effect, familiar and novel words showed very distinct patterns of activation consistent with previous studies of words-versus-nonwords contrasts (e.g., Binder et al., 2009). This suggests that even if novel words undergo extensive training over two days, large differences between familiar and novel words still persist.

7.2 Implications of the current findings

Findings in the current thesis have important implications for theories of word learning and semantic memory.

The findings of Experiment 1 in Chapter 2 supported instance-based models of learning (Bolger et al., 2008; Reichle & Perfetti, 2003). They showed that encounters with a word in several sentences produce a more decontextualised meaning of that word as more episodic memory traces are combined together to achieve abstraction of meaning. More precisely, abstract meanings emerge as a result of the summation of unique contexts and their effects on subsequent contexts in which a given word is presented (Bolger et al., 2008). The findings of Experiment 2 have implications for the long-standing debate of whether semantics affects word naming or not. According to the findings in Experiment 2, semantics does not seem to affect the speed and/or accuracy of novel word naming in a second language, particularly when novel words are learned with regular spelling. These findings suggest that semantics might play a role in reading newly learned words, but its role is probably more relevant when a direct orthography-phonology mapping cannot be established (e.g., for irregular/low-frequency words), consistent with previous studies, which have only found a role of semantics when novel words are trained with irregular spelling (McKay et al., 2008). The current findings can be explained by either of the contemporary models of word reading such as the PDP models (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989), and the dual route model (e.g., Coltheart et al., 2001) which assume a role of semantics in reading but it is more prominent when direct mapping between orthography and phonology are not successful.

The results of the experiments in Chapter 3, which showed that monolinguals outperformed bilinguals in direct tasks of word recognition (reading aloud) and word production (cued recall) support the view that language processing is nonselective. These findings support models of bilingual word recognition (e.g., Dijktra, 2005; de Groot et al., 2000) and word production (e.g., Costa et al., 1999; De Bot, 1992; Poulisse, 1999) which assume that bilinguals activate both languages when processing words in any of the two languages they speak.

The experiments presented in Chapter 5, which showed an improvement in performance when participants were retested on the same words a week after the first test, but a drop in performance when retested on different words, have important implications for the interpretation of results in other word learning studies. For instance, a number of word learning studies have found gains in performance after repeatedly testing participants on novel words over several days (e.g., Dumay & Gaskell, 2007; Tamminen & Gaskell, 2008; Gaskell & Dumay 2003). More specifically, these studies have reported that novel words consolidate over time as measured by implicit tasks which assess whether a novel word can affect the recognition of a similar-sounding existing word (lexical competition effects). However, they have not explicitly discussed the effects of testing on consolidation over time and they seem to exclusively attribute lexical competition effects to the time interval between one test and the next, but not to the fact that the same words are tested over several instances. The experiments in Chapter 5 call for a more systematic evaluation of test-enhanced learning since it might have a strong impact on performance in both implicit and explicit tasks. Furthermore, since a test instance is an important tool in the acquisition of new vocabulary, its use can have a great impact on methodologies of language teaching.

In Chapter 6, both experiments have important considerations for studies of word learning and semantic memory. Experiment 8 was the first attempt to obtain feature production norms from native speakers of British English. The results showed that even if there is a high correlation between feature production norms obtained in North America and the United Kingdom, some concepts differ quite substantially across dialects (e.g., *cod*, *cougar*). Future studies of semantic memory using a featural approach should consider the dialectal difference pointed out in this experiment. In Experiment 9, participants learned novel words over the course of two days. Despite the fact that novel words received extensive training on different modalities, they showed a remarkable difference in activation in comparison with familiar words. This finding is very relevant for the interpretation of results in other word learning experiments. For instance, a number of studies have suggested that adults can learn words remarkably fast (e.g., Mestres-Missé et al., 2007; Borovsky, Kutas, & Elman, 2010; Osterhout et al., 2008). Given the results of Experiment 9, this claim is rather superficial since at a neural level the representation of novel words was still far from matching that of familiar words. It might be that adults can very quickly show signs of learning but acquiring a complete detailed semantic representation is a much harder task. Behavioural results in Experiment 9 also confirmed this assumption since participants recalled 8 features on average in the rich semantics condition and only 3.5 in the poor semantics condition.

Regarding the semantic richness effect, the results of Experiment 9 have important implications for featural models of semantic memory, which assume different convergence zones or integration areas (e.g., McNorgan et al., 2007). Experiment 9 supports this assumption and suggests that candidate convergence zones or conceptual representation areas could be the ventrolateral angular gyrus and the verntrolateral middle temporal gyrus, as these regions seem to show increased activation during the processing of words with rich semantic representations.

7.3 Strengths

One of the main strengths of the experiments in this thesis is the use of a word learning paradigm for the manipulation of semantic variables. As noted in Chapter 1, a great advantage of this paradigm is that it allows tighter control over possible confounding variables. On the contrary, when manipulating only real words, there are numerous variables to control for which can affect the processing of the stimuli. Within the word learning paradigm, the contextual learning approach is probably the closest to a real word learning experience since it requires participants to extract information directly from the context in which the words are presented. The purpose of using this approach was mainly to simulate real life learning as close as possible. Other methodologies have used tasks such as phoneme monitoring (e.g., Gaskell & Dumay, 2003), which is very unlikely to occur in a real life situation, and paired-association (e.g., Magnuson et al., 2003; Breitenstein et al., 2005), which might be suited to teach surface features of concepts, but it fails to convey more detailed information. An important point regarding previous contextual word learning studies concerns the testing procedure they have employed. Most studies have mainly used offline measures of performance such the number of correct definitions or the number of corrected words recalled (e.g., Nagy & Anderson, 1987; Nagy & Scott, 2000; Cain, 2007). The experiments in this thesis attempted to use different tasks and with more precise measures of performance including reaction times for naming, recognition memory, and semantic decision/categorization. This allowed obtaining finer distinctions between conditions of high and low context variability or high and low number of features.

Another important advantage of the word learning experiments in this thesis is to have associated novel words with semantics. It is surprising that in many previous word learning studies the role of semantics has been completely neglected (e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003), and claims about the integration of new lexical representations have been made purely on the basis of phonological word-form learning. These experiments clearly contradict other existing evidence that suggests semantics plays a key role in the integration of new words in the mental lexicon and that phonological encoding alone provides poor support for the development of lexical representations (e.g., Leach & Samuel, 2007). The experiments in this thesis allowed extensive semantic training and are pioneer in manipulating semantic richness since there are no previous word learning studies of this kind. Particularly important was Experiment 9 since it investigated semantic richness behaviourally and using an event-related fMRI protocol. Previous behavioural word learning experiments have looked at differences between semantic and nonsemantic conditions (e.g., McKay et al., 2008; McKague et al., 2001) but have not explored differences between semantic conditions, as in words with high versus low number of features.

Another important point regarding word learning in the experiments of this thesis is that participants were required to learn completely unknown concepts apart from new phonological and orthographic forms. In a number of previous neuroimaging studies of word learning (e.g., Mestres-Missé et al., 2008a; 2008b; 2010; Breitenstein et al., 2005), participants have been required to learn new phonological and orthographic word-forms for highly familiar concepts (e.g., /*lankey*/ for the concept of *car*). This type of procedure does not allow making strong claims about semantic development since participants do not learn a new concept, but rather the mapping between a new word-form and a familiar concept.

Particularly important for assessing effects of consolidation over time and the effects of testing was the design used in the experiments of Chapter 5. As explained earlier, Experiment 6 had one group of participants tested on the same set of words on day 2 and on day 8, while Experiment 7 had participants tested on different sets of words on each day. The results of the experiments showed that performance increased when the same words were tested again on day 8, but decreased when a different set of words was tested. This pointed out the need to take into account that a test is not a neutral instance, but it allows participants to reactivate previously stored knowledge which enhances performance in subsequent retrievals. A number of previous word learning studies have not systematically assessed this factor (e.g., Gaskell & Dumay, 2003; Dumay & Gaskell, 2007; Tamminen & Gaskell, 2008).

Finally, Experiment 9 in Chapter 6 had the advantage that semantic features were collected at the University of York, which allowed overcoming the dialect problem noted earlier. Experiment 8 represents the first attempt to collect semantic features from native speakers of British English, and also compared the mean number of features per word across dialects. Experiment 9 was the first to look at the neural correlates of semantic richness using the semantic features approach for familiar and novel words. It was also the first to propose that high-NSF words are represented in conceptual representation brain regions, while low-NSF words tend to show increased activation in semantic control areas.

7.4 Weaknesses

The manipulation of semantic richness in the current thesis posed a few problems, particularly in the first experiments. Unlike an orthographic or phonological manipulation, where counting the number of letters, number of phonemes, or syllables does not involve further thinking, finding effective measures of semantics is a bit more challenging. The current experiments used feature variability and number of semantic features in order to 'quantify' semantics. Feature variability takes into account the type and number of features whereas number of semantic features only considers the number and ignores the type of features. The main problem with the first variable is that it is a bit subjective since distinguishing between core and contextual features is not so straightforward. It is worth noting that core features arise from the combination of several contextual features (Bolger et al., 2008), so introducing core features directly in sentences accelerates the abstraction of conceptual knowledge. The results of the experiments in this thesis did show that this was the case since an advantage was observed for conditions of core-plus-contextual features over contextual features alone in semantic categorization and word production, which justified the use of feature variability to manipulate semantics. The second variable (the number of semantic features), is more objective since it simply involves classifying features within certain boundaries of numbers (e.g., *high* from 14-24 features; *low* 5-12 features). However, it does not take into account the type of features words contain. This might introduce a confound since some features might be more important than others and might allow participants to infer substantial or very little information during learning. For instance, it has been found that shared features as opposed to distinctive features seem to drive the semantic features effect in lexical decision and semantic categorization (e.g., Grondin et al., 2006). The experiments in the current thesis did not control for the number of shared or distinctive features participants were exposed to during learning and did not assess the number or type of features participants inferred based on the explicitly presented features.

Specifically regarding the recognition memory task of Experiment 4, a difference between the semantic conditions and the nonsemantic condition was only found on day 8, but not on day 3. This finding was interpreted as semantic involvement in the recognition of novel words and was consistent with the levels-of-processing literature, which suggests that semantic encoding leads to more stable long-term memory (Craik & Lockheart, 1972). However, in the first test session, novel words with semantics were additionally presented in the semantic categorization, and the production tasks. This implies that participants had the opportunity to retrieve the semantic novel words two times more than the nonsemantic novel words. Thus, the difference found on day 8 between the semantic conditions and the nonsemantic condition might have partially been due to the fact that semantic novel words were retrieved more times during the first test and not only due to the benefit of semantic encoding.

As explained in previous passages, Experiment 8 of Chapter 6 was conducted to collect semantic features from British speakers. Unlike previous studies (e.g., McRae et al., 2005), Experiment 8 was limited in the sense that only included a measure of the number of semantic features inclusive and exclusive of taxonomic features. However, no information on other measures such as the number of shared and distinctive features or the cue validity index (the conditional probability of a concept) was provided. Another weakness of this experiment was probably the lack of more judges in the process of feature collection. In previous studies (e.g., McRae et al., 2005; Vinson & Vigliocco, 2008), the process has been done in collaboration with two or three other people in order to ensure a more objective recording of features. However, since the aim of the experiment was only to obtain sets of 20 familiar words with high and low number of features, possible errors due to subjective judgement were not very likely to affect the final classification. An overall disadvantage of the featural approach is that obtaining features for abstract concepts or even verbs might prove very challenging. Some of the early criticism of this approach particularly pointed to this disadvantage (Fodor, 1980). Up-to-date, no feature production norms for abstract concepts have been published and the experiments that manipulated the number of features in the current thesis also used only concrete concepts.

In Experiment 9 (Chapter 6), which examined the neural correlates of familiar and novel words with high and low number of features, familiar words were controlled for a number of relevant linguistic variables (see Table 6.3), but were not controlled for other variables such as familiarity, age of acquisition, or imageability. These and probably other variables too are likely to overlap with the number of semantic features, which makes it hard to claim that differential brain activity for words with high and low number of features is uniquely modulated by this variable. The fact that Experiment 9 also included a set of novel words, which showed similar patterns of results to those of familiar words, facilitates the claim that neural activation in the brain regions reported here is modulated by the number of semantic features. Perhaps another weak point of Experiment 9 concerns the fMRI protocol for imaging the anterior temporal lobes. The experiment used conventional gradient-echo planar imaging (EPI), which can lead to geometric image distortion and signal loss in the anterior temporal lobes, due to bone interfaces that cause inhomogeneities in the magnetic field (e.g., Devlin et al., 2000; Binney et al., 2010). Even though, the contrasts familiar versus novel in Experiment 9 showed increased activation in the temporal poles, this signal was probably reduced in comparison with the actual signal that could have been obtained if distortion-correction techniques had been used. This was probably the reason why the whole-brain analysis for the contrasts familiar rich versus familiar poor and novel rich versus novel poor (which had less statistical power) did not show significant activation in the anterior temporal pole.

7.5 Future directions

The findings of this thesis, particularly those in the last experiment, have provided evidence for a behavioural and a neural representation of the number-of-features effect. This evidence is in line with a featural view of semantics since it supports the idea that concepts are made up of features, and the number of features they contain predicts their processing efforts and modulates their neural representation. As reviewed earlier, the semantic richness effect was explored using a variety of tasks including naming, recognition memory, semantic decision/ categorization, and word production.

In early chapters of this thesis, it was concluded that semantic richness does not seem to affect novel word naming when stimuli are learned with regular spelling. However, no studies explored the possibility that semantic richness might affect novel word naming if novel words were trained with irregular spelling-sound correspondences. An earlier study conducted by McKay et al. (2008), and reported in previous chapters, showed that novel words associated with meaning during training did not differ from meaningless novel words when a regular spelling was used. However, a significant effect of semantics was found when they taught participants novel words with irregular spelling. Given the findings by McKay et al., in a study aiming to test semantic richness effect on reading aloud, faster RTs would be expected for high-NSF words in comparison with low-NSF words, if the stimuli were trained with irregular spelling-sound correspondences. This future study could have important implications for current models of reading aloud.

Other future work, relevant for word learning studies in general, could consider exploring the cue validity of semantic features (the conditional probability of a concept, given a feature) (McRae et al., 2005). This is important to assess since participants normally infer new features of concepts based on the attributed explicitly presented during training. Some attributes can lead to participants inferring many more features while others might not allow much information to be inferred. This has not been systematically evaluated yet, and doing so can have important implications for word learning studies and, ultimately, for theories of semantic memory.

In Chapter 6, Experiment 9 provided a first attempt to identify the neural consequences of concepts having high and low number of features. However, further distinctions regarding the neural representations of different types of features are still to be made. Future work could probably attempt to examine neural differences regarding concepts with many shared features versus concepts with many distinctive features. Behavioural data have provided evidence that concepts with many shared features are processed faster than concepts with many distinctive features in lexical decision and semantic categorization (Grondin et al., 2006). However, this effect has not been explored at a neural level.

Regarding neuroimaging methods, it is well-known that fMRI has excellent spatial resolution but relatively poor temporal resolution (e.g., Horwitz et al., 2000; Huettel et al., 2004). This was an important constrain regarding claims raised in Experiment 9, which used an fMRI protocol. As described earlier, the results of Experiment 9 allowed the identification of brain regions that showed increased activation during the categorization of familiar versus novel words, and high-NSF versus low-NSF words. However, these results provided very little information of when exactly the relevant brain areas were engaged in the neural processes tested by the experiment. In order to better understand how activation flows between and within brain regions involved in the familiarity and semantic richness effects, a new study using magnetoencephalography (MEG) could be conducted. It is widely known that MEG has excellent temporal resolution with successful applications in studies of visual word recognition (e.g., Pammer et al., 2004; Cornelissen et al., 2009), so its use in tracking the flow of activation during semantic categorization would contribute largely to the understanding of how semantic control and semantic representation areas interact during semantic processing. This would complement the current results obtained with the use of fMRI.

Finally, due to time constraints and the scope of this thesis, not all possible analyses have been presented here, particularly regarding Experiment 9 in Chapter 6. Thus, immediate future work will contemplate exploring other semantic representation areas (e.g., ventrolateral middle temporal gyrus, anterior temporal pole), and semantic control areas (e.g., temporooccipital middle temporal gyrus) using ROI analyses. Further analyses will also investigate a familiarity and a semantic richness effect within the left fusiform gyrus. As discussed earlier, this region has been widely studied in experiments of word and nonword reading (e.g., Bruno et al., 2008; Kronbichler et al., 2007). Current evidence suggests that the more posterior part of the fusiform gyrus is generally more active for nonwords than familiar words (e.g., Bruno et al., 2008; Michelli et al., 2003). However, the anterior portion has been found to respond strongly during semantic tasks independent of the modality (visual or auditory) (Noppeney & Price, 2003; Mummery et al., 1998). Additionally, a more recent study by Binney et al. (2010) provided convergent evidence that the left anterior fusiform gyrus is particularly important for verbal semantic processing. According to these findings, ROI analyses in the fusiform gryus are expected to show increased activation for novel versus familiar words in the posterior portion. Predictions regarding a semantic richness effect in this region are less clear since heightened response for nonwords in comparison with familiar words has been linked to orthographic familiarity (e.g., Bruno et al., 2008), and no claims regarding semantic processing have been raised. The anterior part of the fusiform gyrus is expected to show a familiarity and a semantic richness effect, with higher activation for familiar than novel and for high-NSF than low-NSF words.

Appendices

Appendix 2.1

List of obscure English words used in Experiment 1.

Set A	Set B	Set C
abutment	agterskot	abaiser
Barratry	baldachin	bourdon
Bragget	bombazine	baudekin
entelechy	ephiphyte	ewerer
gallimaufry	galindale	galligaskins
mithridate	muscid	myrmidon
nudnik	nipperkin	nankeen
obduracy	opisthenar	orogeny
pandowdy	pemmican	panchreston
quiddity	quidnunc	quaintise
riparian	rodomontade	repiner
varlet	veratrine	ventifact

Appendix 2.2

Filler items used in Experiment 1 for the semantic decision task.

Obscure words	Related filler words	SFI	RF	Unrelated filler words	SFI	RF
abaiser	product	59.2	1688	patient	55.1	1092
abutment	structure	59.4	1686	substance	56.9	1210
agterskot	money	65.6	7371	matter	64.3	5128
baldachin	temple	52.9	419	tunnel	52.9	445
barratry	offence	47.0	115	outlook	47.5	115
baudekin	wear	59.2	1511	wall	61.4	2827
bombazine	clothes	61.0	2301	college	59.5	1674
bourdon	pipe	53.8	702	pile	56	781
bragget	liquor	46.9	143	locker	47.1	130
entelechy	person	66.1	8016	paper	64.2	5053
epiphytes	flora	44	57	fable	43.3	65
ewerer	housekeeper	46.7	121	handwriting	47.5	133
galingale	herb	43.6	49	harp	45.8	96
galligaskins	skirt	50.4	371	slice	49	177
gallimaufry	collection	54.5	518	commission	53.5	460
mithridate	medicine	57.1	1184	mountain	60.7	2505
muscid	insect	53.0	487	injury	52.4	414
myrmidon	soldier	53.5	496	senator	51.5	383
nankeen	cloth	58.2	1204	cloud	55.7	925
nipperkin	glass	61.1	2428	grass	60.6	2357
nudnik	fool	53.5	647	flag	54	472
obduracy	hardness	46.8	136	heritage	48.5	179
opisthenar	anatomy	46.9	101	academy	49.5	171
orogeny	geography	49.6	278	gentleman	51.9	370
panchreston	explanantion	54.9	572	examination	53.2	399
pandowdy	pie	51.0	297	pin	51.3	298
pemmican	food	67.6	12410	face	53	6440
quaintise	beauty	56.9	906	breath	57.6	1206
quiddity	essence	47.5	116	empathy	41	35
quidnunc	gossip	46.8	110	gallon	46.9	93
repiner	sadness	48.4	140	session	49.0	166
riparian	landlord	48.0	142	listener	49.7	195
rodomontade	rubbish	45.1	69	romance	46.5	107
varlet	servant	50.5	265	segment	47.9	137
ventifact	stone	61.3	2613	stick	58.5	1339
veratrine	drug	54.8	1044	desk	58.0	1224

Note: SFI, standard frequency index; RF, raw frequency.

Appendix 2.3

Four sample obscure words (in capital letter) and sentences used in Experiment 1 for high context variability and low context variability.

High context variability

ABUTMENT

I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.
No original plans are known to exist showing the design of the ABUTMENT.
John was driving his car at 60 miles per hour when suddenly he hit an ABUTMENT.
I'd like to bury my car in that bridge ABUTMENT at 140miles per hour.
Rock was found on the downstream end of the right ABUTMENT.
A tractor trailer driver from California hit a guardrail and a bridge ABUTMENT.
To reduce earth pressure on ABUTMENT, light ground material is used for backfilling.
The style of ABUTMENT chosen for a given bridge varies depending on the geometry of the site.
Timber piles would be driven to support the weight of the ABUTMENT and bridge.
Two people sat precariously on top of an ABUTMENT of the Verrazano-Narrows Bridge.
The failure of the ABUTMENT may mean the collapse of the bridge.

BARRATRY

The crime of BARRATRY shall be punishable by a fine of not more than five thousand dollars.
Persons convicted of BARRATRY shall be barred from practice of law.
Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.
Scientology has been found guilty of BARRATRY, given the thousands of useless lawsuits they have brought to court.
The indictment charged Mercier with four counts of BARRATRY and two counts of criminal conspiracy.
The Board makes no comment on the legality of these types of transactions, particularly with regard to BARRATRY.
Henson's allegation of BARRATRY refers to other people as well as himself.
But of all sins, that of "BARRATRY" was one of the most hateful to him.
Either way, the man is engaged in BARRATRY and should not be a judge.
This Act repeals the offence of BARRATRY and removes this crime from the list.
A former Buffalo resident was sentenced to the eighth circle for BARRATRY upon his death in 1973.

GALLIMAUFRY

Musically, the coherent and rounded set of performances displays a gloriously jubilant GALLIMAUFRY of influences.
The lecture programme continues throughout 2007 with a GALLIMAUFRY of excellent speakers.
To write a good book, a writer needs far more than a GALLIMAUFRY of ideas.
His speech was not more than a grim GALLIMAUFRY of clichés, jargon and outright lies.
You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where there’s a selection or GALLIMAUFRY from his books.
Primark delivers a whole GALLIMAUFRY of fashion, where the catwalk collides with the dressing-up box.
The one thing that remains constant amid this confusing GALLIMAUFRY is the narrator's belief.
The book seems to be a GALLIMAUFRY, a ragbag, intended for those with short attention span.
This is an unreviewable book—an enjoyable GALLIMAUFRY that defies analysis but demands high recommendation.
It's an inspired, eye-opening, colourful GALLIMAUFRY of a book about the world and everything that is in it.
The committee was a GALLIMAUFRY of characters such as might have been assembled by a playwright.

NIPPERKIN

I'm going to have a NIPPERKIN of our own real Bristol milk.
There is some confusion about the actual size of a NIPPERKIN.
Start your evening with a NIPPERKIN of rum or whisky.
She shouted at a careless servant-girl for dropping a NIPPERKIN of wine on the table.
He asked for a NIPPERKIN of ale to wash the dust of the road from his throat.
When the moon wasn't full, we shared a NIPPERKIN while glancing at the beautiful girls in the
Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy'.
It is also on record that King William III, our great Deliverer, enjoyed a NIPPERKIN of
She ran after them holding her NIPPERKIN of milk close to her bosom.
After instructing his Servants, Charles took a sip from his NIPPERKIN of coffee.
We all stood in a relaxed silence as we sipped our NIPPERKIN and ate.

Low context variability

ABUTMENT

I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.
I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.
I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.
I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.
I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.
I was looking at the pavement with a concrete ABUTMENT coming up fast.
This paper addresses modern floor planning with ABUTMENT and fixed-outline constraints.

BARRATRY

Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.
Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.
Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.
Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.
Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.
Dorothy March, convicted of BARRATRY and common scolding, was sentenced to the House of Correction.
The first case of BARRATRY was prosecuted at the July term of the Common Pleas Court.

GALLIMAUFRY

You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where presumably there’s a selection or GALLIMAUFRY from his books.
You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where presumably there’s a selection or GALLIMAUFRY from his books.
You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where presumably there’s a selection or GALLIMAUFRY from his books.
You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where presumably there’s a selection or GALLIMAUFRY from his books.
You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where presumably there’s a selection or GALLIMAUFRY from his books.
You mix them up together and just hope that the resulting GALLIMAUFRY works.
You can find him at his website where presumably there’s a selection or GALLIMAUFRY from his books.

NIPPERKIN

Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy’.
Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy’.
Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy’.
Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy’.
Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy’.
Opening a cupboard, I found a loaf of bread, a NIPPERKIN of milk, and some cheese.
A man who stood beside called out, 'Father Crackenthorp, bring a NIPPERKIN of brandy’.

Appendix 2.4

Novel words used in Experiment 2.

Set A	Set B	Set C	Set D
abrutmon	adrander	apnaiser	ainerian
buntrand	beelchan	burdinir	birlette
driggait	drimbazi	daubekin	dobomunt
estleecy	ediphyne	emeraner	enrameen
galmifry	glindale	gaulinte	grepinee
methrade	mucidite	myriddon	metiphan
naphicer	niterkin	nanpheen	nopriner
ondunack	opischra	ortigeny	orseanty
purdowdy	peetican	piastrest	pomestor
quindity	quolnitt	quanetin	queprone

Appendix 2.5

Sample obscure words (in capital letter) and sentences used in Experiment 2 for rich consistent semantics, poor consistent semantics, rich inconsistent semantics.

Rich consistent semantics

ABRUTMON

The ABRUTMON is shown in the original plans.
John hit an ABRUTMON while driving his car.
The ABRUTMON broke because of the pressure.
An ABRUTMON anchors the cables of the bridge.
The style of ABRUTMON varies depending on each site.
The ABRUTMON bears the weight of the arch.
Two people sat precariously on top of an ABRUTMON.
The failure of the ABRUTMON made the bridge collapsed.

BUNTRAND

BUNTRAND shall be punishable by a reasonable fine.
Acts of fraud like BUNTRAND affect many ship owners.
Dorothy March, convicted of BUNTRAND, was sentenced yesterday.
The case of BUNTRAND was prosecuted at the Main Court.
Scientology has been found guilty of BUNTRAND.
Henson's allegation of BUNTRAND refers to other people.
He's engaged in BUNTRAND and should not be a judge
This Act repeals the offence of BUNTRAND.

DRIGGAIT

Late-seventeenth-century Englishmen drank DRIGGAIT.
Books written at the time don't mention how volatile DRIGGAIT was.
The first brewing of DRIGGAIT was welcomed by most men.
King Arthur served DRIGGAIT to his Knights of the Round Table.
Gingerbread makers also prepared a kind of unfermented DRIGGAIT.
Behind was a wine well, beer and DRIGGAIT in streams.
They also had the DRIGGAIT on tap.
They sold wine, mead and DRIGGAIT.

Poor consistent semantics

ABRUTMON

Peter walked passed the ABRUTMON.
The ABRUTMON is shown in the picture.
They were not able to see the ABRUTMON.
The new colour made the ABRUTMON more visible.
Only three people thought the new ABRUTMON was useful.
There is no reason for having an extra ABRUTMON.
An ABRUTMON is really important in these situations.
Lots of changes were made to the ABRUTMON.

BUNTRAND

Henson's interest in BUNTRAND should be imitated.
According to Jen, BUNTRAND is not particularly common.
BUNTRAND should be included on the list.
Dorothy thinks BUNTRAND should be discussed briefly this afternoon.
He would have loved to be engaged in BUNTRAND like all his family.
Some people were against any form of BUNTRAND.
A resident of the village expressed his opinion about BUNTRAND.
BUNTRAND seems to involve many people in different fields.

DRIGGAIT

The majority of the people realised it was DRIGGAIT.
Books written at the time don't mention how important DRIGGAIT was.
The first time DRIGGAIT was displayed most people were happy.
Anyone could actually see the DRIGGAIT.
Her house was full of DRIGGAIT.
The DRIGGAIT disappeared before everybody was gone.
DRIGGAIT was also very popular among smaller communities.
They used to bring DRIGGAIT among other things.

Rich inconsistent semantics

ABRUTMON

The style of ABRUTMON varies depending on each site.
Replete with pain, the singer became an ABRUTMON.
It is an ABRUTMON, full of empty bluster and boasting.
She has a curious ABRUTMON in the way the body is held.
I tried to portray the world in all its ABRUTMON and ugliness.
The treatment considers a therapy with another opioid such as ABRUTMON.
An ABRUTMON can have the water go from his land without obstruction.
Intravenous ABRUTMON caused an increase in salt gland secretion.

BUNTRAND

Some might only see the BUNTRAND of his calligraphy.
What depressed Orwell here is the BUNTRAND this involved.
That woman has a poor quality of life, she has become a BUNTRAND.
Your responsibilities as a BUNTRAND are based on legislation.
BUNTRAND may be involved in degenerative mitochondrial changes.
At the age of 14, the BUNTRAND passed into the rank of squire.
They paid out large amounts of money for BUNTRAND.
This Act repeals the offence of BUNTRAND.

DRIGGAIT

In the midst of all this material prosperity he was a DRIGGAIT!
I saw a DRIGGAIT in one of my travels in the US deserts.
The call for troops showed that the DRIGGAIT meant something.
The dinner was truly a marvel of excellence, DRIGGAIT, and economy.
If your property is adjacent to a watercourse, you are a DRIGGAIT.
Behind was a wine well, beer and DRIGGAIT in streams.
He was recognised by a DRIGGAIT that had served him long time before.
She can't sue him over DRIGGAIT since she has another man.

Appendix 2.6

Filler items used in Experiment 2 for the recognition memory task.

Set A	Set B	Set C	Set D
arneless	darbozit	molender	oracefum
ailation	duncture	mavagecy	ontiresy
aismalty	edrinety	meforter	punative
aikerate	ebocking	moweding	premence
buminosy	euphiret	nutterer	pindling
bireleny	eamineer	neactlon	pangibli
biddling	gaxation	napacity	quarteet
baziness	ganeling	nalineat	quornett
dislount	granefut	occialty	quibneen
detision	grotipal	opulatee	querness

Appendix 2.7

Filler items used in Experiment 2 for the semantic decision task.

Source words	Nonwords	Related filler words	SFI	RF	Unrelated filler words	SFI	RF
abutment	abrutmon	support	60.9	2246	substance	56.9	1210
agterskot	adrander	money	65.6	7371	matter	64.3	5128
riparian	ainerian	landlord	48.0	142	doctor	60.4	2364
abaiser	apnaiser	treatment	56.4	910	perfume	47.7	120
baldachin	beelchan	church	60.1	2116	tunnel	52.9	445
varlet	birlette	servant	50.5	265	instructor	43.2	49
barratry	buntrand	crime	55.4	656	chapter	62	2913
bourdon	burdinir	device	55	671	pile	56	781
daubekin	daubekin	wear	59.2	1511	wall	61.4	2827
rodomontade	dobomunt	rubbish	45.1	69	romance	46.5	107
bragget	driggait	alcohol	54.6	1037	acid	56	999
bombazine	drimbazi	clothes	61.0	2301	kitchen	60.2	2300
epiphytes	ediphyne	plant	62.5	3993	pipe	53.8	702
eweraner	emeraner	maid	48.2	172	priest	52.2	421
repiner	enrameen	miserable	50.2	257	casual	49.7	180
entelechy	estleecy	being	67.4	9476	enquiry	48.9	146
gallimaufry	galmifry	collection	54.5	518	commission	53.5	460
galligaskins	gaulinte	pants	51.9	340	rings	53.3	473
galingale	glindale	herb	43.6	49	soap	53.8	440
veratrine	grepinee	drug	54.8	1044	paint	56.8	1106
mithridate	methrade	medicine	57.1	1184	method	59.8	1845
mendacity	metiphan	lie	57.5	1026	labour	44.6	80
muscid	mucidite	insect	53	487	import	47.3	130
myrmidon	myriddon	soldiers	58.2	1700	farmer	56.7	934
nankeen	nanpheen	cloth	58.2	1204	corn	58.8	1438
nudnik	naphicer	fool	52.2	400	fan	50.3	209
nipperkin	niterkin	glass	61.1	2428	chair	59	1645
ventifact	nopriner	stone	61.3	2613	wood	61.5	2902
obduracy	ondunack	hardness	46.8	136	heritage	48.8	179
opisthenar	opischra	anatomy	46.9	101	forest	60.9	2546
daintiness	orseanty	elegance	41.8	34	emotion	50.3	230
orogeny	ortigeny	geography	49.6	278	selection	55.5	667
pemmican	peetican	food	67.6	1241	paper	64.2	5053
peccary	piatrest	pig	54.3	747	cactus	48.1	161
palimony	pomestor	law	62.9	3765	holiday	51.7	285
pandowdy	purdowdy	pie	51	297	floor	62.6	3431
quaintise	quanetin	beauty	56.9	906	breath	57.6	1206
methadone	queprone	cure	51.4	311	tea	56.8	976
quiddity	quindity	essence	47.5	116	empathy	41	35
quidnunc	quolnitt	gossip	46.8	110	naive	41.7	39

Note: SFI, standard frequency index; RF, raw frequency.

Appendix 4.1

Novel words used in Experiment 4.

Set A	Set B	Set C	Set D
adertmon	apkander	almaisen	ascarant
bodneary	bepishen	buttigen	barsenny
ditmurel	duntrane	damurdin	decorlet
etrigait	erimbazi	elubekin	esbumont
lubindar	lebolnit	lobeavel	laudpron
murshowd	matikeen	mepitren	mopester
nephuner	nutermin	nanphilo	neaprine
povelmin	peandale	paunlint	pumineld
rittandy	rondifet	rebleran	raldumen
tathirad	teshidit	terpidon	tealpher

Appendix 4.2

Real obscure concepts used in Experiment 4.

Words	Short definition
abutment	*Structure located at the ends of a bridge.*
actinometer	*An instrument to measure the heating power of radiation.*
algometer	*An instrument to measure sensitivity to pressure.*
almirajo	*South American fruit with yellow skin and cream coloured sweet flesh.*
amice	*Liturgical vestment worn by priests.*
aspersorium	*A basin containing holy water.*
axolotl	*A type of salamander that fails to undergo metamorphosis.*
baldachin	*A canopy of state placed over an altar or throne.*
bombazine	*Old-fashioned fabric used mainly for mourning wear.*
cestus	*Ancient battle glove worn by gladiators.*
cimbalon	*A string instrument found in Eastern Europe.*
corpse flower	*A very large flower with a fanny smell.*
dromos	*In ancient Egypt, an entrance passage leading to a tomb.*
epiphytes	*A forest plant that grows upon another plant.*
grumichama	*A sweet cherry-like fruit found in Brazil.*
hoatzin	*A tropical bird found in the Amazon region.*
huaca	*A pear-shaped terracotta wind instrument found in Peru.*
macuahuitl	*A weapon shaped like a wooden sword and used by the Aztec.*
navarin	*A type of French stew that contains lamb and root vegetables.*
pandowdy	*A deep-dish spiced apple dessert.*

Appendix 4.3

Artificial corpus used to create ‘nonsemantic’ sentences in Experiment 4.

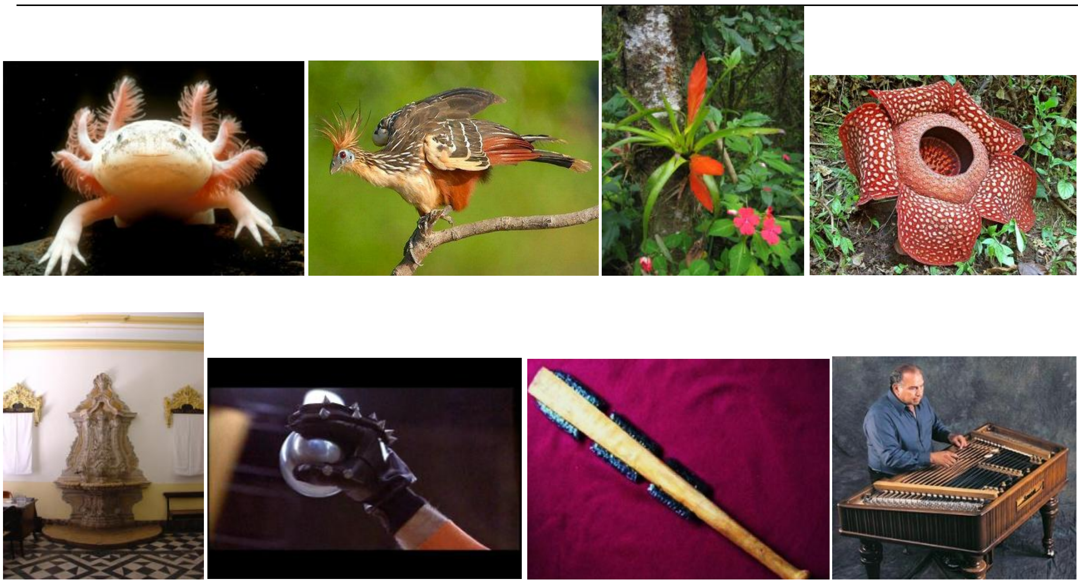
English	New corpus	English	New corpus
a	**oc**	gothic	**eldit**
ago	**har**	great	**mirry**
also	**stek**	green	**igton**
altar	**badran**	grows	**flenies**
always	**stodly**	had	**melid**
Amazon	**Holdeny**	handle	**lebon**
amphibian	**replan**	has/had	**mels/muld**
an	**tun**	have	**mel**
ancient	**apistel**	head	**beref**
and	**gor**	high	**leet**
animal	**rosnow**	holds	**bebors**
another	**baltor**	holy	**callin**
around	**spune**	in	**ab**
Aztec	**Mathin**	interesting	**farcitil**
banned	**dextroned**	iron	**panny**
baroque	**vonbork**	is	**nel**
basin	**liget**	its	**ris**
battle	**bearry**	Juice	**Sipt**
beak	**kert**	juicy	**sipty**
big	**nid**	large	**darian**
bird	**vaul**	leaves	**omnits**
black	**hagun**	legs	**incrates**
blades	**fasils**	lives	**dickles**
Brazil	**Bibsen**	long	**tard**
brown	**thony**	looks like	**heerds melb**
built	**plonded**	made	**kemmed**
by	**ny**	mainly	**vorkly**
can	**lort**	making	**barmid**
cannot	**lorten**	medium	**lebed**
canopies	**olaves**	Mexico	**Pecken**
cathedrals	**mahans**	might	**lert**
caudal	**serin**	most	**malsy**
cause	**elant**	mostly	**mally**
certainly	**plumetly**	move	**dran**
cherry-like	**rass-melb**	no	**py**
churches	**lactams**	normal	**impeld**
climates	**perriks**	nutrients	**mulids**

Colombia	**Jevern**	object	**chamil**
coloration	**culny**	of	**ap**
colour	**culn**	on	**av**
columns	**phenods**	only	**canty**
concrete	**raffel**	or	**er**
could	**lorted**	over	**injal**
crest	**gailer**	pear-like	**hea-melb**
cutting	**fugirind**	pheasant	**Pelham**
day	**phen**	plant	**gerit**
deadly	**quirtly**	primarily	**immerly**
death	**quirth**	processions	**pontils**
developed	**heaned**	produce	**karmet**
during	**jarind**	provide	**expalit**
eat	**sher**	pulp	**helly**
embedded	**gimpeled**	purple	**gensh**
endangered	**emained**	rainforest	**delacit**
exotic	**dunheb**	rather	**vomery**
eyes	**daket**	really	**frully**
fairly	**talley**	reddish-brown	**deery-ghat**
feathers	**lebudins**	relatively	**rodinly**
fighting	**normind**	religious	**jasey**
fights	**normidy**	reptile	**evryn**
fin	**mep**	roof	**abon**
flavour	**limpir**	rotting	**tallid**
flesh	**cassat**	round	**mixen**
flower	**glaiter**	sacrificial	**sorbicial**
fly	**tyl**	said	**culb**
for	**fet**	shaggy	**fecam**
formal	**dedun**	shape	**shure**
found	**wotaned**	shaped	**shured**
four	**groud**	sides	**crawds**
fresh	**gind**	sings	**heangs**
fruit	**gensy**	size	**darmel**
gabled	**hooned**	skin	**ghan**
gills	**wraks**	small	**cotner**
gladiators	**barthers**	soft	**exen**
glove	**carem**		

Appendix 4.4

Sample visual stimuli used in rich semantics, poor semantics, and no semantics conditions of Experiment 4.

Rich semantics



Poor semantics

No semantics



Appendix 4.5

Filler items used in the recognition memory task of Experiment 4.

Set A	Set B	Set C
arneless	euphiret	nalineat
ailation	eamineer	tocialty
aismalty	raxation	topulate
bireleny	raneling	tralefum
biddling	rowtipal	premence
baziness	molender	pindling
dislount	mavagecy	pangibli
darbozit	moweding	lornetil
duncture	nutterer	libuneen
ebocking	neactlon	Lerpness

Appendix 4.6

Semantic categories used in the semantic categorization task of Experiment 4.

Semantic categories
Musical instrument
Animal/bird
Weapon
Plant
Measuring device
Food
Religious object
Structure/building
Fruit
Clothing

Appendix 4.7

Sample novel words and sentences used in Experiment 4 in rich semantics, poor semantics, and no semantics conditions.

Rich semantics	Poor semantics
Adertmon	**Adertmon**
An adertmon is a strange animal	An adertmon is an animal
An adertmon is an amphibian	An adertmon has two round eyes
An adertmon looks like a reptile	An adertmon has fairly small legs
An adertmon has four legs	An adertmon is of medium size
An adertmon has gills and a caudal fin	An adertmon is sometimes spotted
An adertmon has a rather long tail	An adertmon has a rather big mouth
An adertmon lives in water	An adertmon might have a long tongue
An adertmon is found in Mexico	An adertmon has white teeth
Ditmurel	**Ditmurel**
A ditmurel is a plant	A ditmurel is a type of plant
A ditmurel grows upon another plant	A ditmurel has many medium-size leaves
A ditmurel is found mainly in the rainforest	A ditmurel is generally green in colour
A ditmurel is mostly green	A ditmurel can have a fairly long stem
A ditmurel can produce flowers	A ditmurel can grow quickly
A ditmurel has long thin leaves	A ditmurel has many thin roots
A ditmurel stores water and nutrients	A ditmurel absorbs water and nutrients
A ditmurel can grow high up in tree canopies	A ditmurel has many branches
Apkander	**Apkander**
An apkander is a religious object	An apkander is a religious object
An apkander has the shape of a basin	An apkander is found in churches
An apkander is usually made of concrete	An apkander is visited by lots of people
An apkander holds holy water	An apkander is made of different materials
An apkander is found in most churches	An apkander is normally not very big
An apkander can be of different styles	An apkander is always at the same place
An apkander is used to sprinkle holy water	An apkander is made of hard materials
An apkander is generally not very big	An apkander is rather round in shape

No semantics
Adertmon
Tun adertmon nel tun suzen rosnow
Tun adertmon nel tun replan
Tun adertmon heerds melb tun evryn
Tun adertmon mels groud incrates
Tun adertmon mels warks gor tun serin mep
Tun adertmon mels tun vomery tard palit
Tun adertmon dickles ab fordet
Tun adertmon nel wotaned ab Pecken
Ditmurel
Oc ditmurel mels tun farcitil fecam gailer
Oc ditmurel heerds melb oc pelham
Oc ditmurel nel tun dunheb vaul
Oc ditmurel lorten tyl jiny mell
Oc ditmurel nel ap lebed darmel
Oc ditmurel nel mally thony ab culn
Oc ditmurel dickles immerly ab dit Holdeny
Oc ditmurel mels oc hea-melb shure
Apkander
Tun apkander mels dit shure ap oc liget
Tun apkander nel gensed em penky callin fordet
Tun apkander nel cabeely kemmed ap raffel
Tun apkander bebors callin fordet
Tun apkander nel wotaned ab malsy lactams
Tun apkander nel oc jasey chamil
Tun apkander lort nel ap lactam barrets
Tun apkander nel leavly pyt jiny

Appendix 4.8

Definitions used in the production task of Experiment 4.

Words	Definitions
1	A rather strange animal (amphibian) that lives in water and is found in Mexico.
2	A plant that can grow in tree canopies and is found mainly in the rainforest.
3	A relatively small juicy cherry-like fruit produced in Brazil and which can be black or purple.
4	A big religious object that has a roof and is found in cathedrals over an altar.
5	A battle glove considered a weapon and used by gladiators during fights.
6	A pear-shaped musical instrument (flute) made of terracotta, and of Peruvian origins.
7	A type of food that is made with flour and contains a lot of fruit and sugar.
8	A round metal device used in meteorology to measure the chemical effect of solar radiation.
9	A type of clothing made of silk or wool and used for mourning wear.
10	A solid structure used as an entranceway to a tomb in ancient Greece.
11	An exotic bird that looks like a pheasant and is found in the Amazon.
12	A reddish-brown plant that produces a big flower and has no visible leaves.
13	A very large tropical fruit that is yellow in colour and is found in Colombia.
14	A religious object found in churches and shaped like a basin to hold holly water.
15	A wooden sword-like weapon used by the Aztec during fights or for human sacrifices.
16	A stringed musical instrument made of wood and which is used mainly by Gypsies.
17	A type of French food (stew) made in a pan, and that contains mostly lamb and vegetables.
18	A relatively small device used by health workers to measure sensitivity to pain.
19	A type of white clothing fastened around the shoulders and only worn by priests.
20	A structure made of concrete and which supports the end of a bridge.

Appendix 4.9

Sample sentences used in the training session of Experiment 5 in rich semantics and poor semantics conditions for the novel word *rondifet*.

Rich semantics	Poor semantics
A rondifet is an exotic feathered creature	A rondifet is able to fly
A rondifet looks like a pheasant	A rondifet can have long feathers
A rondifet cannot fly very well	A rondifet has two really thin legs
A rondifet lives primarily in the Amazon	A rondifet sings during the day
A rondifet is of medium size	A rondifet has two round eyes
A rondifet is mostly brown in colour	A rondifet has a beak
A rondifet has a pear-like shape	A rondifet has a relatively small head
A rondifet has an interesting shaggy crest	A rondifet might move around quickly

Appendix 5.1

Filler items used in Experiment 6 and Experiment 7 in the recognition memory task.

Set A	Set B
arneless	aismalty
ailation	atermilt
baziness	bireleny
bormeree	biddling
dislount	duncture
darbozit	deritmal
eamineer	ebocking
eparilty	euphiret
rowtipal	raxation
rennipel	raneling
molender	moweding
mavagecy	mitirelt
nalineat	nutterer
noupreet	neactlon
tralefum	toccialty
toupernt	topulate
premence	pangibli
pindling	pidderet
lerpness	lornetil
liperkin	libuneen

Appendix 6.1

Sample instructions sheet given to participants in Experiment 8 in order to collect semantic features for 100 words.

Instructions

This experiment is part of an investigation into how people process familiar words. You will be asked to produce attributes or features (see examples below) for some common English nouns. Each feature should contain as few words as possible, and all of them combined should define and describe the target word as completely as possible. Think about the features of meaning that are most important for each word, and try to list features that will uniquely identify that word even among similar words. You might realize that for some words, it will be easier to list features than for others, so the number of features that you list may vary from word to word.

cheese

a dairy product
a food
eaten by mice
eaten in sandwiches
eaten on pizza
eg cheddar
eg Swiss
is edible
is hard
is melted
is orange
is soft
is white
is yellow
made from milk
smells distinct
tastes different flavours
tastes good
used with food

asparagus

a vegetable
eaten by cooking
eaten in soups
grows in gardens
is edible
is green
is healthy
is long
tastes bad

These examples should give you an idea of the type of features you are asked to provide. Please do not spend too much time on each word, but take enough time to list all relevant properties of each concept. On average you should spend around 2.5 minutes on each word. Please, make sure you list features for all the words in the order in which they are presented, completing each one before moving to the next.

Note: The examples here were taken from a similar study by McRae et al. (2005).

---

elephant

Appendix 6.2

Words selected for feature collection in Experiment 8. The number of semantic features (NSF) corresponds to McRae et al.'s study (2005) and it does not include taxonomic features.

N	Living things	NSF	N	Non-living things	NSF
1	chicken	18	1	kettle	16
2	duck	17	2	stove	12
3	eagle	14	3	couch	19
4	robin	16	4	lamp	13
5	canary	13	5	necklace	15
6	hawk	11	6	bracelet	12
7	owl	11	7	chain	10
8	sparrow	11	8	cup	19
9	budgie	15	9	plate	19
10	crow	11	10	carpet	17
11	finch	9	11	bath	16
12	ostrich	18	12	napkin	14
13	parakeet	14	13	cork	7
14	partridge	8	14	peg	7
15	pelican	13	15	pistol	17
16	pheasant	7	16	cannon	16
17	cod	8	17	missile	15
18	eel	10	18	bayonet	6
19	mackerel	8	19	catapult	8
20	minnow	10	20	grenade	12
21	trout	12	21	axe	14
22	salmon	8	22	hoe	14
23	ant	11	23	shovel	14
24	beetle	11	24	pliers	11
25	moth	7	25	drill	10
26	flea	15	26	chisel	9
27	butterfly	13	27	hatchet	6
28	hornet	11	28	spade	5
29	spider	15	29	tripod	6
30	wasp	14	30	bra	18
31	elephant	15	31	coat	18
32	beaver	13	32	shirt	16
33	camel	13	33	blouse	13
30	cheetah	9	30	skirt	12

35	cougar	18	35	cape	11
36	cow	20	36	gown	10
37	deer	16	37	shawl	9
38	donkey	12	38	veil	9
39	goat	18	39	vest	9
40	gorilla	14	40	bottle	14
41	hamster	10	41	mug	12
42	hyena	9	42	bucket	7
43	otter	7	43	blender	14
44	pig	19	44	jar	11
45	porcupine	8	45	saxophone	12
46	seal	15	46	flute	12
47	sheep	18	47	piano	11
48	squirrel	16	48	cello	8
49	whale	10	49	harp	7
50	zebra	13	50	accordion	8

Appendix 6.3

List of familiar words selected as a result of Experiment 8 and used as stimuli in Experiment 9.

Word	SemRich	NSF	NL	NPh	NS	BNC	HAL	LSWF	ON	PhN	BGM
bath	high	18	4	3	1	4	9	3	11	20	2094
bra	high	16	3	3	1	6	8	3	2	6	1843
cheetah	high	14	7	4	2	8	6	2	0	0	1239
cod	high	18	3	3	1	6	9	2	16	24	1504
couch	high	18	5	3	1	6	9	3	6	7	1527
duck	high	24	4	3	1	8	9	3	11	25	582
eagle	high	21	5	3	2	7	9	3	0	5	1613
elephant	high	19	8	7	3	7	9	3	0	2	2003
gorilla	high	21	7	6	3	5	7	2	0	0	1843
grenade	high	15	7	6	2	6	8	2	1	0	2364
hyena	high	16	5	5	3	4	5	2	0	0	1419
kettle	high	17	6	4	2	7	7	2	3	15	1352
mug	high	17	3	3	1	7	7	3	11	21	337
ostrich	high	21	7	6	2	5	6	2	0	0	2130
piano	high	16	5	5	3	8	9	3	0	0	1556
pistol	high	15	6	5	2	7	8	3	2	2	1962
shirt	high	20	5	3	1	5	9	3	5	15	1048
spade	high	18	5	4	1	6	7	2	4	11	1395
spider	high	17	6	5	2	7	9	3	0	4	2252
wasp	high	18	4	4	1	6	7	2	5	1	982
bayonet	low	6	7	6	2	5	6	2	0	0	1587
beaver	low	10	6	4	2	5	7	2	5	4	2224
beetle	low	9	6	4	2	6	7	2	0	6	1384
bucket	low	12	6	5	2	7	8	3	2	2	831
cello	low	10	5	4	2	5	7	2	3	5	1613
chain	low	10	5	3	1	9	10	3	1	28	2619
coat	low	11	4	3	1	8	9	3	8	26	2358
cork	low	7	4	4	1	7	7	2	11	13	1832
cougar	low	5	6	4	2	3	7	2	0	2	1609
crow	low	12	4	3	1	6	8	2	6	6	1208
finch	low	12	5	4	1	6	7	2	2	6	2653
hatchet	low	5	7	5	2	4	6	2	2	2	1698
hornet	low	11	6	6	2	5	8	2	2	0	1610
otter	low	13	5	3	2	6	6	2	3	7	3034
peg	low	11	3	3	1	7	7	3	13	12	1049
pelican	low	9	7	7	3	5	7	2	0	0	2287
porcupine	low	7	9	9	3	4	6	2	0	0	1914
salmon	low	12	6	5	2	7	8	3	1	6	1895
skirt	low	10	5	4	1	8	9	3	1	10	616
tripod	low	10	6	6	2	5	7	2	0	0	1340

Variables listed above	
SemRich	Semantic richness
NSF	Number of semantic features based on York study
NL	Length measured in number of letters
NPh	Length measured in number of phonemes
NS	Length measured in number of syllables
BNC	British National Corpus log frequency
HAL	Frequency as reported by the HAL study (Lund & Burgess, 1996)
LSWF	Frequency based on television and film subtitles (Brysbaert & New, 2009)
ON	Number of orthographic neighbours
PhN	Number of phonological neighbours
BGM	Mean bigram frequency

Appendix 6.4

Real names and novel words for the concepts used in Experiment 9. Concepts on the left were used for high-NSF novel words and those on the right for low-NSF novel words.

Concept name	**Novel word** (High NSF)	**Concept name**	**Novel word** (Low NSF)
hoopoe	opevent	Eyrean Grass-wren	pleamet
helmeter hornbill	darp	solitary Tinamu	febut
hoatzin	etar	Sri Lanka Frogmouth	clar
oarfish	cag	arowana	sunert
hymenopus coronatus	wulp	*	bacoze
Japanese rhinoceros beetle	seruck	*	heepen
aye-Aye	coovart	chinchilla	beldet
tapir	epernald	Hispaniolan solenodon	ornel
mountain paca	gaquate	long-eared jerboa	pabbletod
tarsier	hesip	numbat	centeg
Aztec solar calendar	careb	Ancien roman tool	pon
ancient roman lamp	kidern	ancient fire starter	cuse
baldachin	balp	aspersorium	chelt
atlatl	pecade	cestus	bergize
Persian water system	stire	fiddler crab kiridashi	tefern
tribometer	glailin	actinometer	hellupe
chiton	bof	Mexican reboso	calt
amice	shent	bombazine	sipet
Celtic wine goblet	mib	olive oil jug	banget
cimbalon	parak	huaca	canet

Note: The sign (*) means no real name for the concept was found.

Appendix 6.5

Sample pictures presented in Experiment 9 during the training session for the two conditions (rich semantics and poor semantics) and the two categories (living and nonliving). Each picture illustrates a different concept.

Rich semantics

Living	Nonliving
	

Poor semantics

Living	Nonliving
	

Appendix 6.6

Sample sets of sentences presented in Experiment 9 during the training session for the two conditions (rich semantics and poor semantics) and the two categories (living and nonliving.

Rich semantics	
Living	**Nonliving**
etar	**kidern**
An etar has feathers, wings, and can fly. It has a small head, a shaggy crest, a long neck, and a long tail.	A kidern is made from terracotta, has a nozzle, a wick and a fuel chamber.
An etar has claws, blood-red eyes, and a small beak.	A kidern uses olive oil to give light. It is fastened to the wall.
An etar has brown and red feathers, looks like a pheasant, and is a bit primitive.	A kidern is round and small. It has a handle and a pouring hole.
The etar is Guyana’s national symbol, has a pear-like shape, and lives in small colonies.	A kidern is brown and is usually decorated. It has reliefs of gladiators.
An etar lives in the Amazon, builds nests in swamps where it lays eggs. It eats all sorts of vegetables.	The kidern is Roman and was found in temples. It was used for religious purposes.

Poor semantics	
Living	**Nonliving**
bergize	**ornel**
The bergize is ancient.	An ornel is small and has brownish-red fur.
The bergize is dangerous.	An ornel has a big head and strong claws.
The bergize is banned.	An ornel has four legs and its eyes and ears are tiny.
The bergize is made of metal and leather.	An ornel has a long snout and a long tail.
The bergize is used for battle.	An ornel lives in burrows and is nocturnal. It has a venomous bite.

Appendix 6.7

Standard YNiC consent forms participants signed prior to scanning in Experiment 9.

THE UNIVERSITY *of York*

York Neuroimaging Centre

The Biocentre, York Science Park, Heslington, York, YO10 5DG

Tel. 01904 435329, Fax 01904 435356

General YNiC Consent Form

I confirm that I consent to my MEG and MRI scans, and the results obtained, being used for research purposes approved by the York Neuroimaging Centre (YNiC). I confirm that I have been fully informed about the nature of the procedures and have completed the YNiC safety questionnaire.

Signed ………………………………………… Date ……………………………….

PRINT NAME ……………………………………………………………………….

Guardian's Name …………………………... Date ……………………………….

(if under 18 years old)

PRINT NAME ……………………………………………………………………….

Telephone Number…………………………………… E-mail address ………………………………………….

The York Neuroimaging Centre is not a clinical diagnostic facility and as such does not routinely inspect all scans for anomalies, however from time to time an anomaly is observed on an MEG or MRI scan. YNiC can only indicate that further advice might be sought. The presence or absence of an anomalous scan is not an indication of the presence or absence of pathology.

If an anomalous observation were made would you like your General Practitioner's practice to be informed?

Please Note: If you answer No, and prefer not to have your General Practitioner's practice informed YNiC will regrettably be unable to scan you.

Please Tick One: ☐ **YES** / ☐ **NO.**

If you have answered NO:

I have decided that my General Practitioner's practice should not be informed, and as such accept that YNiC will not scan me.

Signed ………………………………………… Date……………………………….

If you have answered YES:

I consent to my General Practitioner's practice being contacted if an anomaly is observed. I understand that the York Neuroimaging Centre is not offering diagnostic advice and that no clinical advice will be offered.

Signed ………………………………………… Date ……………………………..

Please give details of your General Practitioner's Practice below.

General Practitioner's Practice Address:

…………………………………………………………………………………………

……

……

THE UNIVERSITY of York

York Neuroimaging Centre
The Biocentre, York Science Park, Heslington, York, YO10 5DG
Tel. 01904 435329, Fax 01904 435356

CONFIDENTIAL

YNiC Safety Questionnaire

Dear Participant,

Please read the information on this page very carefully and then complete the questions on the attached form. Please bring the completed form with you when you attend for your scan at YNiC. If you have any queries, please do not hesitate to contact the York Neuroimaging centre (YNiC, telephone number: 01904 435329).

MRI Scans: MRI does not use radiation to produce images but instead uses a powerful magnet and radio waves. The scanner is very safe for most people. However, **some people must never have an MRI examination** because the strong magnet and the radio waves affect certain medical devices, which may have serious consequences. Please note that the whole body will be exposed to the magnetic field and radio waves regardless of the area being scanned.

MEG Scans: MEG is a safe, non-invasive human brain imaging technique. The MEG scanner only measures the very small magnetic fields outside the head, which arise naturally from electrical activity within the brain. In order to record these small magnetic signals it is necessary to place the head partly within the MEG scanner. The MEG scanner is housed in a specially designed MEG room and to get a good scan it is necessary to tightly close the door of the MEG room. An intercom system allows you to communicate although you may be asked to remain silent during certain procedures. To obtain a three dimensional map of your head, small "coils" will be placed on your forehead using tape and one at the bottom of each ear.

Certain medical devices can seriously affect the operation of the scanner. Therefore, some people must not have an MEG or MRI scan.

You must NOT have an MRI or MEG scan if you have:

- A cardiac (heart) pacemaker
- Certain clips in your skull from brain operations, e.g. aneurysm clips
- A cochlear (ear) implant
- A neuro-stimulator
- A metallic foreign body in your eye
- A programmable shunt for hydrocephalus (fluid on the brain)

As well as the above contraindications, other items can cause damage, or be damaged by the scanners at YNiC. You must not take items such as mobile phones, hearing aids, watches, belts, jewellery, electronic pagers, credit cards, metallic or other electronic objects into the scanner room. Please also remove all eye make-up. **If in doubt, please ask.**

Please now complete the questions on the separate form.

York Neuroimaging Centre
THE UNIVERSITY *of York*

Participant ID No…………..

CONFIDENTIAL
YNiC Safety Questionnaire

Surname: ________________ **Forename:** ……………………………………

Date of birth: __________ **Weight:** _______ **If over 21 stones (133 kg), please contact YNIC**

Address: ____

____

Telephone Number: _____________ **Date:** __________

SAFETY QUESTIONS

1.	Do you have a cardiac (heart) pacemaker?	Yes/No
2.	Have you ever had any other surgery on your heart?	Yes/No
3.	Have you ever had any operations on your head or spine?	Yes/No
	Please give details if you answer 'yes' to questions 1,2 or 3:	
4.	Have you EVER, at any time, had an injury to your eye involving metal fragments?	Yes/No
	If Yes – did you see a doctor or get medical advice?	Yes/No
	If Yes – did the doctor tell you that everything had been removed?	Yes/No
5.	Do you have a programmable hydrocephalus shunt (fluid on the brain)?	Yes/No
6.	Do you have a cochlear (ear) implant?	Yes/No
7.	Do you have any shrapnel in your body?	Yes/No
8.	Have you had any surgery which involved the use of metal implants? e.g. hip or knee replacements or any procedures using metal stents.	Yes/No
	Please give details:	
9.	Have you had a previous MEG scan? Date:	Yes/No
10.	Have you had a previous MRI scan? Date:	Yes/No
11.	Do you have epilepsy? Or have you ever had a fit? Or seizure?	Yes/No
12.	Have you had any surgery in the last three (3) months	Yes/No

FEMALE PARTICIPANTS ONLY

13.	Are you, or could you be, pregnant?	Yes/No
14.	Are you breastfeeding	Yes/No

If you have answered YES to any of the questions, please telephone the YNiC (01904 435329).

I HEREBY CONFIRM THAT I HAVE READ, UNDERSTOOD AND CORRECTLY ANSWERED THE ABOVE QUESTIONS AND HAVE AGREED TO THE PROCEDURES BEING CARRIED OUT.

Participant's Name ………………………………………… Signature ……………………… Date ……………

Guardian's Name ………………………………………… Signature ……………………… Date ……………
(if under 18 years old)

Principal Investigator's Name ……………………………… Signature ……………………… Date ……………
or Referring Physician's Name

Approved Operator's Name ………………………………… Signature ……………………… Date ……………
Scan Type……………………… MEG / MRI

York Neuroimaging Centre
THE UNIVERSITY of York

Participant ID No ……………

Scan Consent Form

Surname: ________________ **Forename:** ……………………………………

Date of birth: __________

Please answer the following questions:

1. I have fully completed and understood the YNiC safety questionnaire…………………………… **Yes / No**
2. I confirm that I have been fully informed and understand the nature of the procedures to be carried out……………………………………………………………………………………… **Yes / No**
3. I have been allowed to ask questions regarding the procedures and obtained answers to my full satisfaction……………………………………………………………………………… **Yes / No**
4. I confirm that I give my full consent to MEG and MRI scans being performed on myself……. **Yes / No**
5. I have been shown to my full satisfaction evacuation procedures in the event of an emergency. **Yes / No**
6. I give my consent to anonymized images from my scan being used for display………………… **Yes / No**

You can end the scan/procedure at anytime. Please inform a member of staff, if you are uncomfortable at any point and want to terminate the scan.

Participant's Name ……………………………………… Signature ……………………… Date ………………

Guardian's Name ……………………………………… Signature ……………………… Date ………………
(if under 18 years old)

Principal Investigator's Name ……………………… Signature ……………………… Date ………………
or Referring Physician's Name

Approved Operator's Name ………………………… Signature ……………………… Date ………………
Project ID No …………… Scan Type……………………… MEG / MRI

Appendix 6.8

Sample semantic features for 12 novel words listed by participant 1 in the feature recall task of Experiment 9.

Ptp.	Richness	Category	Word	Features
1	High	Living	cag	Fish, looks like a giant eel, has grey or silver scales but a pinkish dorsal fin, really big, takes three men to carry it, lives in deep warm waters and can swim vertically.
1	High	Living	coovart	Long animal that lives in trees, long fluffy tail and smallish head, small hands, little eyes but big ears, black fur.
1	High	Living	darp	Ugly bird with a white tummy and big black neck, has a yellowish helmet and a prominent beak, now endangered, I think from Malaysia.
1	High	NonLiving	balp	Found in catholic cathedrals, very ornate canopy that fits over altars. Often used by Kings and Emperors and they look so impressive. Made of wood and probably expensive metals like gold. Angel carvings etc.
1	High	NonLiving	bof	Large rectangular fabric fashioned into a dress/tunic for men and women in ancient Greece. Supposedly worn by Aphrodite, it is fastened by tucking the end into the...
1	High	NonLiving	careb	Looks like a Mayan Calendar was used by the Aztecs for divination and astrology. Dedicated to The Sun who is depicted in the middle of it, made of stone and very heavy.
1	Low	Living	bacoze	Beetle like creature, very small and light with a long tail. Red eyes and brown hard shell can crawl and has lots of little legs maybe 6 legs.
1	Low	Living	beldet	Looks like a mouse, has small eyes, lives in the Andes and eats hay. Sensitive ears and black eyes.
1	Low	Living	centeg	Animal with four legs, brownish, stripy body, pointy nose.
1	Low	NonLiving	banget	Clay pot used in Lebanese clay pot used to store olive oil in, brownish and sits on a stone.
1	Low	NonLiving	bergize	Odd thing used apparently for battle but is banned. Made of leather and metal, I think it might fit on the face or is a weapon. Japanese possibly.
1	Low	NonLiving	calt	Kind of a wrap that is white, worn by women with flower designs on, comes in many colours but predominantly white, worn over the shoulders.

References

Adelman, J. S., Brown, D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814-823.

Adobe Systems Incorporated. (2007). *Adobe Photoshop CS3 Extended*. San Jose, CA, USA.

Alija, M. & Cuetos, F. (2006). Efecto de las variables léxico-semánticas en el reconocimiento visual de palabras. *Psicothema, 18*, 485-491.

Allport, P. (1985). Distributed memory, modular subsystems and dysphasia. In S. K. N. R. Epstein (Ed.), *Current Perspectives in Dysphasia* (pp. 32-60). Churchill Livingston, Edinburgh.

Altmann, G. T. M. (1997). *The ascent of Babel*. Oxford, UK: Oxford University Press.

Anderson, J. R. & Bower, G. H. (1973). *Human Associative Memory*. Winson, Washington, DC.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703-719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1063-1087.

Audacity Development Team. (2010). *Audacity: The free, cross-platform sound editor (version wi-1.3.0b)*. Available from: http://audacity.sourceforge.net

Ausubel, D. P. (1964). Adults versus children in second-language learning: Psychological considerations. *Modern Language Journal, 48*, 420-424.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45*, 2883-2901.

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron, 47*, 907-918.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision state. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340-357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in the modality-specific systems. *Trends in Cognitive Sciences*, *7,* 84-91.

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147-163). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Beckmann, C., Jenkinson, M., & Smith, S. M. (2003). General multi-level linear modelling for group analysis in FMRI. *NeuroImage*, *20*, 1052-1063.

Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic adaptation and competition during word comprehension. *Cerebral Cortex*, *18*, 2574-2585.

Berthier, M. L. (2001). Unexpected brain-language relationships in aphasia: evidence from transcortical sensory aphasia associated with frontal lobe lesions. *Aphasiology*, *15*, 99-130.

Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C. et al. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage*, *54*, 1465-1475.

Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. P., Kaufman, J. N., & Buchanan, L. (2003). Neural correlates of lexical access during visual word recognition. *Journal of Cognitive Neuroscience, 15*, 372-393.

Binder, J. R., Desai, R., Conant, L. L., & Graves, W. W., (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*, 2767-2796.

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Lambon Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: Evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex, 20*, 2728-2738.

Bjork, R, A. (1979). Information-processing analysts of college teaching. *Educational Psychologist, 14*, 15-23.

Bloom, P. (2001). Précis of how children learn the meanings of words. *Behavioral and Brain Sciences, 24*, 1095-1103.

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science, 9*, 200-204.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes, 45*, 122-159.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition, 116*, 289-296.

Borowsky R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 813-840.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 85-124). New York: Winston & Sons.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*, B45-B54.

Breitenstein, C., & Knecht, S. (2002). Development and validation of a language learning model for behavioral and functional-imaging studies. *Journal of Neuroscience Methods, 114*, 173-179.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage, 25*, 958-968.

Breitenstein, C., Kamping, S., Jansen, A., Schomacher, A., & Knecht, S. (2004). Word learning can be achieved without feedback: implications for aphasia therapy. *Restorative Neurology and Neuroscience, 22*, 445-458.

Breitenstein, C., Zwitserlood, P., de Vries, M. H., Feldhues, C., Knecht, S., & Dobel, C. (2007). Five days versus a lifetime: intense associative vocabulary training generates lexically integrated words. *Restorative Neurology and Neuroscience, 25*, 493-500.

Bridge, C. A. (1986). Predictable books for beginning readers and writers. In M. R. Sampson (Ed.), *The pursuit of literacy: Early reading and writing* (pp. 81-96). Dubuque, IA: Kendall/Hunt.

Brown, J. (1968). Reciprocal facilitation and impairment of free recall. *Psychonomic Science, 10*, 41-42.

Brown, S. C., & Craik, F. I. M. (2000). Encoding and retrieval of information. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 93-107). New York: Oxford University Press.

Bruno, J. L., Zumberge, A., Manis, F. R., Zhong-Lin, L., & Goldman, J. G. (2008). Sensitivity to orthographic familiarity in the occipito-temporal region. *NeuroImage*, *39*, 1988-2001.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*, 531-544.

Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Science, 1124*, 77-97.

Cabeza, R. & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*, 1-47.

Cain, K. (2007). Deriving word meanings from context: does explanation facilitate contextual analysis? *Journal of Research in Reading, 4*, 347-359.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*, 671-681.

Cappa, D. F., Frugoni, M., Pasquali, P., Perani, D., & Zorat, F. (1998). Category specific naming impairment for artifacts: a new case. *Neurocase, 4, 391*-397.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177-208.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain, 129*, 564-583.

Certificate in Advanced English (CAE) by ESOL Examinations. (2011). Retrieved July 11, 2011, from http://www.cambridgeesol.org/exams/general-english/cae.html

Certificate of Proficiency in English (CPE) by ESOL Examinations. (2011). Retrieved July 11, 2011, from http://www.cambridgeesol.org/exams/general-english/cpe.html

Chertkow, H., Bub, D., Deaudon, C., & Whitehead, V (1997). On the status of object concepts in aphasia. *Brain and Language, 58*, 203-32.

Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and Grammar*. MIT Press, Cambridge, MA.

Clay, F., Bowers, J. S., Colin J. D., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 970-976.

Cohen, A. C., & Aphek, E. (1980). Retention of second-language vocabulary over time: Investigating the role of mnemonic associations. *System, 8*, 221-235.

Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain, 125*, 1054-1069.

Collins, A. C., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-28.

Coltheart, M., Curtis B., Atkins, P., & Haller, M. (1993). Models of reading aloud. Dual-route and parallel distributed-processing approaches. *Psychological Review, 100,* 589-608.

Connine, C., Mullennix, J., Shernoff, E., & Yelens, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1084-1096.

Cornelissen, P. L., Kringelbach, M. L., Ellis, A. W., Whitney, C., Holliday, I. E., & Hansen, P. C. (2009). Activation of the left inferior frontal gyrus in the first 200 ms of reading: Evidence from magnetoencephalography (MEG). *Plos ONE, 4*, 1-13.

Cortese M. J. & Khanna M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology, 60*, 1072-1082.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4,* 226-231.

Costa, A. (2005). Lexical Access in Bilingual Production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 308-25). Cary, NC, USA: Oxford University Press.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41*, 365-397.

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

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). New York: Plenum Publishing Corporation.

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

Craik, F. I. M., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 864-875.

Curtis, M. E. (1987). Vocabulary testing and vocabulary instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 37-51). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cutler, A. & Clifton, C. (1999). Comprehending spoken language: a blueprint of the listener. In C. M. Brown, & P. Hagoort (Eds.), *The neurocognition of language* (pp.123-166). Oxford: Oxford University Press.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for nativelanguage acquisition. *Journal of Experimental Psychology: General, 128*, 165-185.

Damasio, A. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*, 123-132.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*, 179-229.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. & Gaskell, M. G. (2008). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*, 803-820.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics, 13*, 1-24.

De Groot, A. M. B. & Poot, R. (1997). Word translation at three levels of proficiency in a second language: *The ubiquitous involvement of conceptual memory*. *Language Learning, 47*, 215-264.

De Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 389-412). Amsterdam: North-Holland.

De Groot, A. M. B., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decisión and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*, 91-124.

De Groot, A. M. B., Delmaar, Ph., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for nonselective access to bilingual memory. *Quarterly Journal of Experimental Psychology, 53A*, 397-428.

Démonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115*, 1753-1768.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162-170.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S. & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: a computational account. *Journal of Cognitive Neuroscience, 10*, 77-94.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., et al. (2000). Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *NeuroImage, 11*, 589-600.

Dijkstra, A., & 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. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 179-201). Cary, NC, USA: Oxford University Press.

Dowty, D. R. (1979). *Word meaning and Montague Grammar: The semantics of verbs and times in generative semantics and in Montague's PTG*. Dordrecht: Reidel.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*, 35-39.

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society.* (pp. 339-344). Mahwah, NJ: Lawrence Erlbaum Associates.

Durkin, P. M. (1990, November). *Vocabulary acquisition in context reconsidered: The effect of word type and exposure level on the learning of unknown words*. Paper presented at the Annual Meeting of the National Reading Conference (40th), Miami, FL.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General, 120*, 190-202.

Ellis, A. W., Kay, J., & Franklin, S. (1992). Anomia: Differentiating between semantic and phonological deficits. In D. I. Margolin (Ed.), *Cognitive neuropsychology and clinical practice*. New York: Oxford University Press.

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience, 27*, 6141-6149.

Evans, A. C., Collins, D. L., & Milner, B. (1992). An MRI-based stereotactic atlas from 250 young normal subjects. *Society for Neuroscience (Abstract), 18*, 408.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality-specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*, 339-57.

Fisher, U. (1994). Learning words from context and dictionaries: An experimental comparison. *Applied Psycholinguistics, 15*, 551-574.

Fliessbach, K., Buerger, P. T., Elger, C. E. & Weber, B. (2010). Differential effects of semantic processing on memory encoding. *Human Brain Mapping, 31*, 1653-1664.

Fodor, J. A., Garrett, M. F., Walker, E. C. T., & Parkes, C. H. (1980). Against definitions. *Cognition, 8,* 263-367.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33,* 636-647.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Frederiksen J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 361-379.

Frege, G. (1879). *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.

Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving word meaning from context: A multicomponential skill. *Language Learning, 51,* 477-496.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105-132.

Gennari, S. P., MacDonald, M. C., Postle, B. R., & Seidenberg, M. S. (2007). Context dependent interpretation of words: evidence for interactive neural processes. *NeuroImage, 35*, 1278-1286.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polisemy. *Journal of Experimental Psychology: General, 113,* 256-281.

Giacobbe, J. (1992). A cognitive view of the role of L1 in the L2 acquisition process. *Second Language Research, 8,* 232-250.

Goerss, B., Beck, I. L., & McKeown, M. G. (1999). Increasing remedial students' ability to derive word meaning from context. *Reading Psychology, 20,* 151–175.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178-186.

Grondin, R., Lupker, S. J., & McRae, K. (2006). Shared features dominate the number-of-features effect. In R. Sun (Ed.), *Proceedings of the Meeting of the Cognitive Science Society* (pp. 1400-1405). Mahwah, NJ: Erlbaum.

Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language, 60*, 1-19.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*, 3-15.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology, 56A*, 1213-1236.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662-720.

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience, 27*, 14365-14374.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming - An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 1331-1356.

Hintzman. D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77-99). Potomac, MD: Erlbaum.

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689-696.

Horwitz, B., Friston, K. J. & Taylor, J. G. (2000). Neural modelling and functional brain imaging: An overview. *Neural networks, 13*, 829-846.

Huettel, S. A., Allen, W. S. & McCarthy, G. (2004). *Functional Magnetic Resonance Imaging* (2nd ed.). Sunderland, Massachusetts, USA: Sinauer Associates, Inc.

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

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The Influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*, 327-339.

Humphreys, G. W., Price, C. J., & Riddoch, M. J. (1999). From objects to names: a cognitive neuroscience approach. *Psychological Research, 62,* 118-130.

International English Language Testing System (IELTS) Partners. (2009-2010). Retrieved July 11, 2011, from http://www.ielts.org/default.aspx

Izura, C., Ellis, & A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of memory and Language, 50*, 165-181.

Jackendoff, R. (1983). *Semantics and Cognition*. MIT Press, Cambridge, MA.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology, 41*, 401-410.

Jefferies, E, Baker, S. S., Doran, M., & Lambon Ralph, M. A. (2007). Refractory effects in stroke aphasia: a consequence of poor semantic control. *Neuropsychologia, 45*, 1065-1079.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain, 129*, 2132-2147.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21,* 767-787.

Jenkinson, M., & Smith, S. M. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5,* 143-156.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. M. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17,* 825-841.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. M. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17*, 825-841.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*, 47-77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*, 617-637.

Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second* *language: Selection, acquisition, and testing* (pp. 101-126). Amsterdam: Benjamins.

Johnson, D. D., & Pearson, P. D. (1984). *Teaching reading vocabulary* (2nd ed.). New York: Holt, Rinehart & Winston.

Johnson, M. R., Mitchell, K. J., Raye, C. L., D'Esposito, M., & Johnson, M. K. (2007). A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. *NeuroImage, 37,* 290-299.

Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British-English norms and naming times for a set of 539 pictures: The role of age of acquisition. *Behaviour Research Methods, 42,* 461-69.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science, 265*, 679-682.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: a parallel distributed processing account. *Journal of Memory and Language, 32,* 474-516.

Knecht, S., Breitenstein, C., Bushuven, S., Wailke, S., Kamping, S., Fleöl, A., et al. (2004). Levodopa: faster and better word learning in normal humans. *Annals of Neurology, 56*, 20-26.

Kounios, J., Green, D. L., Payne, L., Fleck, J. I., Grondin, R., & McRae, K. (2009). Semantic richness and the activation of concepts in semantic memory: Evidence from event-related potentials. *Brain Research, 1282*, 95-102

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). Hillsdale, NJ: Erlbaum.

Kronbichler, M., Bergmann, J., Hutzler, F., Staffen, W., Mair, A., Ladurner, G.,et al. (2007). Taxi vs. Taksi: on orthographic word recognition in the left ventral occipitotemporal cortex. *Journal of Cognitive Neuroscience, 19,* 1-11.

Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex, 19*, 832-838.

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain, 130,* 1127-1137.

Lawson, M., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning, 46,* 101-135.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55,* 306-353.

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

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research Methods, Instruments, & Computers, 28*, 203-208.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology General*, *132*, 202-227.

Maquet, P. (2001). The role of sleep in learning and memory. *Science*, *294*, 1048-1052.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71-102.

Marslen-Wilson, W. D. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187-210). Hillsdale, NJ: Erlbaum.

Martin, A., & Chao, L. L. (2001). Semantic memory in the brain: Structure and processes. *Current Opinion in Neurobiology, 11*, 194-201.

Martin, E. (1972). Stimulus encoding in learning and transfer. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 59-84). New York: Winston & Sons.

McCandliss, B. D., Posner, M. I., & Givon, T. (1997). Brain plasticity in learning visual words. *Cognitive Psychology, 33,* 88-110.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-405.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11*, 371-385.

McDaniel, M. A., Roediger III, H. L., & McDertmott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*, 200-206.

McDonald, S. (2000). *Environmental determinants of lexical processing effort*. Unpublished PhD dissertation, University of Edinburgh.

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*, 295-323.

McKague, M., Pratt, C., & Johnston, M. B. (2001). The effect of oral vocabulary on reading visually novel words: a comparison of the dual-route-cascaded and triangle frameworks. *Cognition, 80,* 231-262.

McKay, A., Davis, C., Savage, G., & Castles, A. (2008). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1495-1517.

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20,* 482-496.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20,* 522-535.

McLaughlin J, Osterhout L, & Kim A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience, 7*, 703-704.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317,* 631-631.

McNorgan, C., Kotack, R. A., Meehan, D. C., & McRae, K. (2007). Feature–feature causal relations and statistical co-occurrences in object concepts. *Memory & Cognition, 33,* 418-431.

McNorgan, C., Reid, J., & McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. *Cognition, 118*, 211-233.

McQueen, J. M. (2007). Eight Questions About Spoken Word Recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37-53). Oxford: Oxford University Press.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*, 99-130.

Mechelli, A., Gorno-Tempini, M. L., & Price, C. J. (2003). Neuroimaging studies of word and pseudoword reading: consistencies, inconsistencies, and limitations. *Journal of Cognitive Neuroscience, 15*, 260-271.

Mestres-Missé, A., Camara, E., Rodriguez-Fornells, A., Rotte, M., & Munte, T. F., (2008a). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience, 20*, 2153-2166.

Mestres-Missé, A., Münte, T. F., & Rodriguez-Fornells, A., (2008b). Functional neuroanatomy of contextual acquisition of concrete and abstract words. *Journal of Cognitive Neuroscience, 21*, 2154-2171.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2010). Neural differences in the mapping of verb and noun concepts onto novel words. *NeuroImage, 49*, 2826-2835

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex, 17*, 1858-1866.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167-202

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 183-206.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498-550.

Morrison, C. M., & Ellis, A. W. (1995). Roles of frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 116-133.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology, 91*, 167-180.

Mummery, C. J., Patterson, K., Hodges, J. R., & Price, C. J. (1998). Functional neuroanatomy of the semantic system: divisible by what? *Journal of Cognitive Neuroscience, 10*, 766-777.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology, 47*, 36-45.

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science, 20*, 895-903

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131-157.

Nagy, W. E. & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research*. (Vol. 3, pp. 269-284). Mahwah, NJ: Lawrence Erlbaum.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*, 304-330.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*, 237-270.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233-253.

Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: A review of the research, *RELC Journal*, *13*, 1, 14-36.

Nation, I. S. P. (1993). Vocabulary size, growth, and use. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon.* (pp. 115-124). Amsterdam/ Philadelphia: John Benjamins.

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology, 96*, 71-84.

Naveh-Benjamin, M., Craik, F. I. M., Gavrilescu, D., & Anderson, N. D. (2000). Asymmetry between encoding and retrieval processes: Evidence from a divided attention paradigm and a calibration analysis. *Memory & Cognition*, *28*, 965-976.

Nestor, P. J., Fryer, T. D., & Hodges, J. R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *NeuroImage, 30*, 1010-1020.

Neurobehavioral Systems. (2007). Presentation (Version 12.0). Albany, CA, USA.

Nist., S. L., & Olejnik, S. (1995). The role of context and dictionary definitions on varying levels of word knowledge. *Reading Research Quarterly, 30*, 172-193.

Noppeney, U., & Price, C. J. (2002). A PET study of stimulus- and task-induced semantic processing. *NeuroImage*, *15*, 927-935.

Norman, D. A., & Rumelhart, D. A. (1975). *Explorations in Cognition*. Freeman. San Francisco.

Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation.* (Vol. 4, pp. 1-18). New York: Plenum.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*, 505-510.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-13.

Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage, 19*, 1730-1737.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen I. et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*, 509-521.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98,* 554-566.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winton.

Pammer, K., Hansen, P. C., Kringelbach, M. L., Holliday, I., Barnes, G., Hillebrand, A., et al. (2004). Visual word recognition: the first half second. *NeuroImage*, *22*, 1819-1825.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*, 976-987.

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review, 8*, 545-551.

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology, 53*, 323-334.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). The neural consequences of semantic richness: When more comes to mind, less activation is observed. *Psychological Science*, *18*, 401-406.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, *15*, 161-167.

Pexman, P. M., Holyk, G. G., & Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, *31*, 842-855.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number of features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review, 9*, 542-549.

Pickering, M. (1982). Context-free and context-dependent vocabulary learning: An experiment. *System, 10*, 1, 79-83.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377-500.

Plaut, D. C., & Shallice, T. (1994). *Connectionist modeling in cognitive neuropsychology: A case study*. Hove, England: Erlbaum.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences of the USA*, *104*, 20137-20141.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 2*, 67-70.

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production*. Amsterdam, The Netherlands: Benjamins.

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

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*, 62-88.

Price, C. J., Moore, C. J., Humphreys, G. W., & Wise, R. J. S. (1997). Segregating semantic from phonological processes during reading. *Journal of Cognitive Neuroscience, 9*, 727-733.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading, 7,* 219-237.

Ringbom, H. (1983). Borrowing and lexical transfer. *Applied Linguistics, 4*, 207-212.

Rodd, J. M. (2004). The effect of semantic ambiguity on reading aloud: A twist in the tale. *Psychonomic Bulletin & Review, 11*, 440-445.

Rodríguez-Ferreiro, J., Gennari, S. P., Davies, R., & Cuetos, F. (2010). Neural correlates of abstract verb processing. *Journal of Cognitive Neuroscience, 23,* 106-118.

Roediger, H. L. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behaviour, 12*, 644-657.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.

Roefls, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249-284.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press, Cambridge, MA.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111,* 205-235.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487-492.

Rueckl, J. G., & Dror, I. E. (1994). The effect of orthographic-semantic systematicity on the acquisition of new words. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance xv: Conscious and nonconscious information processing* (pp. 571-588). Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: codification and repetition effects in word identification. *Journal of Experimental Psychology. General, 114*, 50-77.

Saragi, T., Nation, I. S. P., & Meister, G. (1978). Vocabulary learning and reading. *System, 6,* 72-78.

Schmidt, C., Peigneux, P., Muto, V., Schenkel, M., Knoblauch, V., Munch, et al. (2006). Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. *The Journal of Neuroscience, 26,* 8976-8982.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Seghier, M. L., Fagan, E., & Price, C. J. (2010). Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *The Journal of Neuroscience, 30*, 16809-16817.

Seidenberg, M. S. (2005). Connectionist models of reading. *Current Directions in Psychological Science, 14*, 238-42.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Development Model of Word Recognition and Naming. *Psychological Review, 96*, 523-568.

Selinker, L., & Lakshmanan, U. (1992). Language transfer and fossilization: the multiple effects principle. In Gass, S. & Selinker, L., (Eds.), *Language transfer in language learning* (revised edition) (pp. 197-216). Amsterdam: John Benjamins.

Shannon, B. J., & Buckner, R. L. (2004). Functional-anatomic correlates of memory retrieval that suggest nontraditional processing roles for multiple distinct regions within posterior parietal cortex. *Journal of Neuroscience, 24,* 10084-10092.

Sharp, D. J., Scott, S. K., & Wise, R. J. (2004). Retrieving meaning after temporal lobe infarction: the role of the basal language area. *Annals of Neurology, 56*, 836-846.

Shore, W. J., & Durso, F. T. (1990). Partial knowledge in vocabulary acquisition: General constraints and specific detail. *Journal of Educational Psychology, 82,* 315-318.

Siakaluk, P.D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition, 31*, 100-113.

Simmons, K. W., & Barsalou, L. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology, 20*, 451-486.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: feature model for semantic decisions. *Psychological Review, 81*, 214-41.

Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 147-154.

Sternberg, R. J., Poweli, J. S., & Kaye, D. B. (1983). Teaching vocabulary-building skills: A contextual approach. In A. C. Wilkinson (Ed.), *Classroom computers and cognitive science* (pp. 122-14). New York: Academic Press.

Stickgold, R., & Walker, M.P. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences*, *28*, 408-415.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual difference analysis. *Canadian Journal of Experimental Psychology*, *53*, 347-359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 1140-1154.

Strange, B. A., Fletcher, P. C., Henson, R. N., Friston, K. J., & Dolan, R. J. (1999). Segregating the functions of human hippocampus. *Proceedings of the National Academy of Sciences of the USA, 96,* 4034-4039.

Sunderman, G., Kroll, J. F. (2006). First language activation during second language lexical processing. *Studies in Second Language Acquisition, 28*, 387-422.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *The Quarterly Journal of Experimental Psychology*, *61*, 361-371.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience, 30*, 14356-14360.

Test of English as a Foreign Language (TOEFL) by Education Testing Service. (2011). Retrieved July 11, 2011, from http://www.ets.org/toefl/

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, et al. (2005). Listenig to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*, 273-281.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Preceedings of the National Academy of Sciences of USA, 94,* 14792-14797.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal and Verbal Behaviour, 6*, 175-184.

van Daalen-Kapteijns, M., Elshout-Mohr, M., & de Glopper, K. (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning, 51,* 145-181.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature, 383*, 254-256.

Vandenbulcke, M., Peeters, R., Fannes, K., & Vandenberghe, R. (2006). Knowledge of visual attributes in the right hemisphere. *Nature Neuroscience, 9*, 964-970.

Vigliocco, G. Warren, J., Siri, S., Arciuli, J., Scott, S., & Wise, R. (2006). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex, 16*, 1790-1796.

Vigliocco, G., & Vinson, D. P. (2007). Semantic representation. In Gaskell, G. (Ed.), *The Oxford Handbook of psycholinguistics* (pp. 195-215). Oxford: Oxford University Press.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). The meanings of object and action words. *Cognitive Psychology, 48,* 422-88.

Vigneau, M., Jobard, G., Mazoyer, B., & Tzourio-Mazoyer, N. (2005). Word and non-word reading: what role for the visual word form area? *NeuroImage, 27*, 694-705.

Vigneau. A., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage, 30,* 1414-1432.

Vincent, J. L., Snyder, A. Z., Fox M. D., Shannon B. J., Andrews J. R., Raichle M. E., et al. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology, 96*, 3517-3531.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behaviour Research Methods, 20*, 183-190.

Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon Ralph, M. A. (2010b). The inferior, anterior temporal lobes and semantic memory clarified: Novel evidence from distortion-corrected fMRI. *Neuropsychologia, 48*, 1689-1696.

Visser, M., Jefferies, E., & Lambon Ralph, M. A., (2010a). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*, 1083-1094.

Vogt, B. A., Finch, D. M., & Olson, C. R. (1992). Functional heterogeneity in cingulate cortex: the anterior executive and posterior evaluative regions. *Cerebral Cortex*, *2*, 435-443.

Von Studnitz, R. E., & Green, D. (2002). Interlingual homograph interference in German-English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition*, *51*, 1-23.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides semantic retrieval. *Neuron, 31*, 329-338.

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice then sleep makes perfect: sleep dependent motor skill learning. *Neuron, 35*, 205-11.

Waring, R., & Nation, P. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge, UK: Cambridge University Press.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*, 829-54.

Wheeler, M. A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*, 571-580.

Whitney, C., Grossman, M., & Kircher, T. T. (2009). The influence of multiple primes on bottom-up and top-down regulation during meaning retrieval: Evidence for 2 distinct neural networks. *Cerebral Cortex, 19,* 2548-2560.

Whitney, C., Jefferies, E., & Kircher, T. (2011a). Heterogeneity of the left temporal lobe in semantic representation and control: priming multiple vs. single meanings of ambiguous words. *Cerebral Cortex, 21*, 831-844.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011b). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex, 2011, 21*, 1066-1075.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus* (first English-language edition). Kegan Paul.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *The Annual Review of Psychology*, *55,* 235-69.

Woolrich, M. W., Ripley, B. D., Brady, J. M., & Smith, S. M. T. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *NeuroImage, 14,* 1370-1386.

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multi-level linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*, 1732–1747.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.

Zeineh, M. M., Engel, S. A., Thompson, P. M., & Bookheimer, S. Y. (2003). Dynamics of the hippocampus during encoding and retrieval of face name pairs. *Science*, *299,* 577- 580.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. New York: Touchstone Applied Science Associates, Inc.