IDIOMS IN THE BILINGUAL MENTAL LEXICON

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

SECOND LANGUAGE STUDIES

MAY 2016

By

Jung-Ts Lin

Dissertation Committee:

Thom Hudson, Chairperson
James D. Brown
Theres Grüter
Seongah Im
Amy J. Schafer
ACKNOWLEDGEMENTS

This dissertation could not have been finished without the support I received from faculty members, organizations, friends, and family.

First, I would like to thank my advisor and the chair of my committee, Dr. Thom Hudson. He has always been very supportive of my research interests, helped me refine my dissertation topic, and generously spent time discussing my research with me. There were a couple of times when I did not know how to make my dissertation topic feasible. He listened carefully to my thoughts, and then asked me questions. Through answering his questions, I gradually developed my research questions. I am also grateful for his feedback on the results of my data analysis. From his feedback, I learned how to present and discuss results in a way that helps readers easily understand. This is a skill that I will continue to use. His generosity and patience encouraged me to finish this dissertation.

I would also like to thank Dr. Theres Grüter and Dr. Amy J. Schafer. They both helped me refine the method I used to conduct my study. I was new to the field of psycholinguistics, so I had tremendous difficulty designing my experiment at the very beginning. They were both very patient and enthusiastic in discussing it and sharing their brilliant ideas with me. They taught me how to design and develop a rigorous experiment, from preparing the materials and conducting pilot studies to deciding on the procedures. Without their guidance and help, I would have had difficulty realizing this research. Although I still have a lot to learn about experimental design, they sparked an interest in this kind of experiment that I will carry with me.

My sincere gratitude also goes to Dr. Seongah Im and Dr. James D. Brown. They both helped me analyze the data. They were very patient about listening to my research design, examining my data, and discussing possible solutions with me. They
taught me how to apply statistical theories to authentic data, and to look at research questions from different perspectives. Their depth of knowledge has broadened my understanding of statistics.

I also want to thank my supervisor, Dr. Yang Zhang, at the Mānoa Institutional Research Office. As my mentor at work, she has always been very supportive of my doctoral studies. It is not easy to balance work and writing a dissertation. During the journey to finish my dissertation, I sometimes felt tired and frustrated, but she always cheered me up and helped me accomplish my goal without jeopardizing my performance at work. Her enthusiasm for life inspires me to be enthusiastic and optimistic toward life even when it is sometimes hard.

I am truly grateful to my family in Taiwan, Ying-Chieh Lin, Hsiu-Chuan Kuo, and Jung-Ying Lin. I would like to thank my friends in Hawai‘i, especially Ruobing Chi, Huizhong Wang, Dawei Wang, Yue Tang, Yanxia Li, Yang Lei, XiaoXi Xu, XuHua Cheng, and too many others to mention here. Without your company, I cannot imagine that I would have come so far on this journey. Your love and support is deeply rooted in my heart and will never fade. From all of you, I have learned how love and support can influence and even change a person. In different ways, each of you has become a model, inspiring me to try to help and encourage other people as you have done for me.

I am thankful to my participants both in Taiwan and in Hawai‘i. I especially want to thank the Linguistics Beyond the Classroom Project at the University of Hawai‘i at Mānoa. Completing this dissertation would have been impossible without your contribution. I am also thankful to my editor, Laurie Durand, who helped me realize my thoughts in polished texts, and made my dissertation more reader friendly. Last but not least, I would like to express my sincere appreciation to the Ministry of
Education in Taiwan. Without its financial support, I do not think this journey would even have started.

This journey has been bumpy, but the views along the way have never been boring. I enjoyed it greatly, and I will not forget the wonderful people and interesting things that I have met with along the way.
This study investigates how native and non-native speakers of English understand English idioms by investigating how context and other factors affect idiom processing and how idioms are represented in speakers’ mental lexicons. The study utilizes Abel’s (2003) dual idiom representation model (DIRM), which differs from previous idiom processing models in that it is based on second language (L2) data as well as first language (L1) data.

In order to address the questions of how L1 and L2 speakers process and conceive of idioms, the study conducted an experiment in which L1 English speakers and L2 English speakers completed a self-paced reading (SPR) task. Both groups also completed a semantic decomposability survey and an idiom familiarity survey, and the L2 group additionally completed a metaphoric equivalence survey.

The participants were 36 native speakers of Chinese and 40 native speakers of English, all either undergraduate or graduate students at the University of Hawai‘i at Mānoa (UHM). For the study materials, 24 English idioms from a list of idioms utilized by Abel (2003) were selected on the basis of word frequency, currency, degree of familiarity, semantic decomposability, and metaphoric equivalence. The SPR task used all of these idioms, and included 48 passages (12 priming contexts, 12 neutral contexts, and 24 fillers), which participants read on a computer screen. Their reaction times (RT) were recorded by the computer. The surveys asked the participants to rate the 24 idioms in terms of familiarity, semantic decomposability, and metaphoric equivalence (L2 group only) on a 5-point Likert scale.

The research provides several important results. First, both the L1 and the L2 language users processed the idioms faster in priming contexts than in neutral contexts. The dissertation discusses this finding in terms of the graded salience
hypothesis (Giora, 1997) of idiom processing. Second, both the L1 and the L2 language users tended to conceive of the idioms as semantically decomposable rather than semantically nondecomposable, and this tendency was stronger in the L1 group; this finding differs from previous studies’ reports. Drawing on the dual idiom representation model, the dissertation provides possible reasons for this difference. The analysis of the L1 group’s idiom processing found an interaction effect between familiarity and semantic decomposability: As the degrees of familiarity and semantic decomposability increased, the RTs in the neutral contexts decreased greatly, but the RTs in the priming contexts did not. In the L2 participants’ idiom processing, as the degrees of metaphoric equivalence between L1 and L2 increased, the RTs in the neutral contexts decreased greatly, while the RTs in the priming contexts did not. However, as the degrees of semantic decomposability increased, the RTs in the priming contexts decreased greatly, while the RTs in the neutral contexts did not.

This study contributes to research on second language acquisition and idiom processing by using the DIRM to investigate both L1 and L2 language users’ idiom comprehension. In addition, it provides insights into how to select idioms and create reliable rating surveys for idiom research. Furthermore, its results offer L2 language teachers some pedagogical implications by providing a better understanding of how L2 learners conceive of and process L2 idioms.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................ iii
ABSTRACT............................................................................................................................. vi
TABLE OF CONTENTS.......................................................................................................... viii
LIST OF TABLES.................................................................................................................... xii
LIST OF FIGURES ............................................................................................................... xiv

CHAPTER 1 INTRODUCTION .............................................................................................. 1

CHAPTER 2 LITERATURE REVIEW ..................................................................................... 9

2.1 Context and familiarity in idiom processing.............................................................. 10
2.2 The graded salience hypothesis (GSH) and the mixed results of previous research ......................................................................................................................... 13
2.3 Language exposure and idiom comprehension.......................................................... 18
2.4 L1 idiom processing models and their limitations..................................................... 21
  2.4.1 The idiom list hypothesis ....................................................................................... 21
  2.4.2 The lexical representation hypothesis ................................................................. 22
  2.4.3 The configuration model ..................................................................................... 25
2.5 The dual idiom representation model ...................................................................... 29
2.6 The bilingual lexical representation of idioms......................................................... 37
2.7 Conceptual metaphors............................................................................................... 42
  2.7.1 Crosslinguistic influence on conceptual metaphors ............................................. 46
  2.7.2 The modified hierarchical model and crosslinguistic conceptual metaphors ........ 48
  2.7.3 Crosslinguistic conceptual metaphors and L2 idiom comprehension.................. 50
2.8 Chapter summary ........................................................................................................... 51
2.9 The study: NS and NNS idiom processing .................................................................... 53
  2.9.1 Sentence-by-sentence self-paced reading ................................................................. 53
  2.9.2 Research questions and predictions ....................................................................... 54

CHAPTER 3 METHODS ...................................................................................................... 62
3.1 Research instruments ................................................................................................. 63
3.2 Participants .................................................................................................................. 67
  3.2.1 Establishing sample sizes ...................................................................................... 67
  3.2.2 Characteristics of the participants .......................................................................... 68
  3.2.3 Language proficiency test ..................................................................................... 68
3.3 Materials ...................................................................................................................... 69
  3.3.1 Process of idiom selection ...................................................................................... 69
    3.3.1.1 Validity of idiom ratings ................................................................................... 70
    3.3.1.2 Reliability of the idiom rating surveys ............................................................. 74
  3.3.2 Familiarity, semantic decomposability, and metaphoric equivalence surveys .......... 74
  3.3.3 Idiom contexts ........................................................................................................ 78
3.4 Procedures .................................................................................................................... 80
3.5 Pilot Studies .................................................................................................................. 80
  3.5.1 Pilot study one: Selection of idioms ......................................................................... 80
  3.5.2 Pilot study two: Naturalness of contexts ................................................................. 81
  3.5.3 Pilot study three: Development of the metaphoric equivalence survey ... 81
  3.5.4 Pilot study four: Procedures of the experiment ..................................................... 81
3.6 Data collection .............................................................................................................. 82
CHAPTER 4 RESULTS AND DISCUSSION

4.1 Comprehension questions and general language proficiency test ..........84
  4.1.1 Data analysis .............................................................................84
  4.1.2 Results ......................................................................................85

4.2 Influence of context and language on NS and NNS idiom processing (RQ1 & RQ2) .................................................................................88
  4.2.1 Data analysis .............................................................................88
  4.2.2 Results ......................................................................................88
  4.2.3 Discussion .................................................................................95

4.3 Idiom representation in the NS and NNS mental lexicon (RQ3) ..........97
  4.3.1 Data analysis of previous studies ...........................................97
  4.3.2 Data analysis of the present study ...........................................98
  4.3.3 Results ......................................................................................99
  4.3.4 Discussion ................................................................................112

4.4 Factors influencing NS and NNS idiom processing (RQ4 & RQ5) ......118
  4.4.1 Data analysis .............................................................................118
    4.4.1.1 Model construction in a comparable previous study ..........119
    4.4.1.2 Model construction in the present study ...............................120
  4.4.2 Results ......................................................................................124
    4.4.2.1 The language group model ...............................................124
    4.4.2.4 English-speaking group: Two-predictor model .................125
    4.4.2.5 English-speaking group: Three-predictor model ...............126
    4.4.2.6 Chinese-speaking group: Three-predictor model ...............127
    4.4.2.7 Chinese-speaking group: Six-predictor model ....................129
  4.4.3 Discussion ................................................................................129
CHAPTER 5 CONCLUSION

5.1 Summary of tasks and findings

5.2 Implications of the findings

5.3 Limitations and future research questions

5.4 Significance and contribution of this study

APPENDIX A Idioms

APPENDIX B Language background survey and modified C-test

APPENDIX C Familiarity and semantic decomposability surveys

APPENDIX D Idiom contexts

APPENDIX E Distribution of familiarity and semantic decomposability ratings

APPENDIX F Normality and linearity check

APPENDIX G Distribution of reaction time before log transformation

REFERENCES

APPENDIX A Idioms

APPENDIX B Language background survey and modified C-test

APPENDIX C Familiarity and semantic decomposability surveys

APPENDIX D Idiom contexts

APPENDIX E Distribution of familiarity and semantic decomposability ratings

APPENDIX F Normality and linearity check

APPENDIX G Distribution of reaction time before log transformation

REFERENCES
LIST OF TABLES

Table 1. Factor Analysis ...........................................................................................................72
Table 2. Summary of Data Analysis ............................................................................................84
Table 3. Summary of General Language Proficiency Test Results ............................................88
Table 4. Results of the Three-Predictor Model.........................................................................93
Table 5. Results of the Three-Predictor Model.........................................................................94
Table 6. Results of Regression Analysis for Effects of Language Group on Semantic
Decomposability Ratings ............................................................................................................99
Table 7. NS and NNS Ratings of Idioms as Semantically Decomposable and
Semantically Nondecomposable ...............................................................................................101
Table 8. Summary of Descriptive Statistics of Percentages by Semantic
Decomposability and Language Group .........................................................................................102
Table 9. NS and NNS: Mean Scores for Six Idioms ..................................................................107
Table 10. NNS: Correlations between Familiarity, Semantic Decomposability, and
Metaphoric Equivalence ............................................................................................................108
Table 11. NS: Correlations between Familiarity and Semantic Decomposability ....108
Table 12. NS and NNS: Mean Scores for the Other 17 Idioms .....................................................110
Table 13. Comparison of the Current Study and the Previous Studies.................................114
Table 14. Summary of Independent Samples t-Test of Mean Familiarity Rating Scores
.....................................................................................................................................................117
Table 15. Language Group Predictor Model ............................................................................125
Table 16. English-speaking Group: Three-predictor Model .......................................................127
Table 17. Chinese-speaking Group: Three-predictor Model ......................................................128
Table 18. Summary of Significant Predictors in the Two Language Groups ..............130
Table 19. Effect of Semantic Decomposability on the Reaction Time Difference in the English-Speaking Group

.......................................................... 131
LIST OF FIGURES

Figure 1. Revised hierarchical model ................................................................. 38
Figure 2. The distributed feature model ............................................................. 39
Figure 3. Modified hierarchical model ............................................................... 40
Figure 4. The cognitive semantics view of idiom processing ............................... 43
Figure 5. Crosslinguistic difference in metaphoric referents .............................. 49
Figure 6. Hypothetical graph of an interaction effect between language group and context type ................................................................. 56
Figure 7. Hypothetical graph of an interaction effect between language group and context type ................................................................. 57
Figure 8. Hypothetical graph of no interaction effect between language group and context type ................................................................. 58
Figure 9. Screenshots from the self-paced reading task ....................................... 66
Figure 10. A screenshot from the familiarity rating survey .................................. 75
Figure 11. A screenshot from the semantic decomposability rating survey .......... 77
Figure 12. A screenshot from the metaphoric equivalence rating survey .......... 78
Figure 13. Two outliers in the language proficiency test .................................... 86
Figure 14. Distribution of accuracy rates on the language proficiency test .......... 87
Figure 15. Distribution of RTs in priming contexts ............................................ 89
Figure 16. Distribution of RTs in neutral contexts ............................................. 90
Figure 17. Language group and context ............................................................ 92
Figure 18. Ratio of two types of semantic decomposability ............................... 103
Figure 19. Percentages in the semantically decomposable category across languages ................................................................. 104
Figure 20. Percentages in the semantically nondecomposable category across languages
CHAPTER 1
INTRODUCTION

Language can be interpreted not only literally but also figuratively. For example, the phrase “comparing apples and oranges” can literally mean comparing the two kinds of fruit. However, in certain contexts, it can figuratively mean that two things are incomparable. This figurative interpretation is a conventionalized use in English-speaking society. Conventionalized use is one of the properties of formulaic expressions (Yorio, 1980) because language is constructed not only by arranging individual words in a grammatical fashion, but also by using series of words in a conventional fashion (Sinclair, 1991). Wray and Perkins (2000) held that formulaic expressions are stored and retrieved as a whole from memory, and are resistant to grammatical analysis. Grammatical analysis refers to a process of understanding a literal expression (e.g., please give me a cup of coffee) by identifying the meaning of each word and then connecting the meaning of all of the words in the expression to produce the semantic meaning of the whole. In contrast, the meaning of a formulaic expression cannot be obtained from the meaning of its constituents; for example, an idiom can be viewed as equivalent to a single word—an “idiom word” (Bobrow & Bell, 1973)—in that its meaning is semantically nondecomposable.

Semantic nondecomposability is an important feature in early first language (L1) idiom processing models, namely the idiom list hypothesis (ILH; Bobrow & Bell, 1973), the lexical representation hypothesis (LRH; Swinney & Cutler, 1979), and the configuration model (CM; Cacciari & Tabossi, 1988). Because the meaning of an idiom can be ambiguous (i.e., between a figurative meaning and a literal meaning),
these three models focus on the order in which native speakers of English (NS) retrieve the figurative meaning and the literal meaning of an idiom. However, the three models make different arguments regarding when idioms are recognized and retrieved as a whole.

Because these three models are based on native speaker data, they assume that people are familiar with idioms. This assumption is acceptable for L1 adult speakers because they have already been exposed to a great deal of the language in question (Ellis, Simpson-Vlach, & Maynard, 2008), and are considered to be advanced language users. However, the same assumption is not applicable to all second language (L2) learners, classroom-taught learners in particular, due to the limited amount of their exposure to the target language (Danesi, 1992). Nevertheless, chances are that L2 learners are still able to figure out the meanings of idioms even if they are unfamiliar with them (Abel, 2003). In this sense, the idiom list hypothesis, the lexical representation hypothesis, and the configuration model fail to explain how L2 users comprehend idioms because they do not address the influence of degree of familiarity on idiom comprehension.

Giora’s (1997) graded salience hypothesis argues that familiarity is crucial to idiom processing, and that the salient meaning of an idiom can be either its figurative meaning or its literal meaning depending on degree of familiarity. The graded salience hypothesis is more powerful than the previous three models in explaining the retrieval of figurative meanings for L1 and L2 language users. However, some studies have found that for nonnative speakers (NNS), the literal meaning is always the salient meaning (Cieślicka, 2006; Kecskés, 2000; Siyanova, Conklin, & Schmitt, 2011). Others have found that NNSs can activate figurative meanings as fast as they
do literal meanings, indicating that the literal meaning is not always dominant (Conklin & Schmitt, 2008; Schmitt & Underwood, 2004).

One problem with the existing research on this topic is that the researchers did not measure and compare L1 and L2 language users’ familiarity with idioms. Therefore, we do not know what the salient meaning of an idiom is in the L1 and L2 language users’ mental lexicon. According to a later model, the dual idiom representation model (DIRM; Abel, 2003), both semantic decomposability and familiarity play important roles in idiom processing. The model is based on a study that found that NSs tended to conceive idioms as semantically nondecomposable, while NNSs had a tendency to conceive idioms as semantically decomposable. In other words, according to the dual idiom representation model, NSs and NNSs conceive the semantic decomposability of idioms differently, which affects whether they retrieve an idiom as a whole or by its constituents, which in turn affects the speed of processing. The speed of processing is a function of processing ease or processing difficulty (Jegerski, 2014). Another problem of the existing studies (Cieślicka, 2006; Conklin & Schmitt, 2008; Kecskés, 2000; Schmitt & Underwood, 2004; Siyanova et al., 2011) is that not all of them take semantic decomposability into consideration.

A preliminary question in the field is whether L2 learners are unable to understand idioms if they are unfamiliar with the idioms’ figurative meanings. According to the dual idiom representation model, even if L2 learners are unfamiliar with an idiom, they still have a chance of figuring out its meaning by making use of its semantic decomposability and the conceptual metaphor it draws on. “Conceptual metaphor” refers to the metaphor used in an idiom. For example, the conceptual metaphor for blow one’s stack is “anger is heated fluid in a container” (Gibbs,
Bogdanovich, Sykes, & Barr, 1997). Kövecses (2003) and Yu (1995) showed that the conceptual metaphor of an idiom can be either the same or different across languages. The similarity or difference of conceptual metaphors between languages will affect learners’ idiom comprehension. Gibbs and his colleagues conducted a series of studies (Gibbs et al., 1997; Gibbs & O’Brien, 1990) in which they found that people accessed conceptual metaphors when comprehending idioms, but not when comprehending literal meanings of idioms or nonidiomatic phrases. However, little research has investigated the influence of L1–L2 conceptual metaphor equivalence on NNS idiom processing.

The present study will use the dual idiom representation model to investigate the factors influencing L1 and L2 language users’ activation of the figurative meanings of idioms. The dual idiom representation model embraces more aspects of idiom comprehension than the other earlier models, and these aspects have been found to be important (Gibbs et al., 1997; Gibbs & Nayak, 1989; Gibbs & O’Brien, 1990; Giora, 1997). To be specific, the dual idiom representation model can account for both L1 and L2 idiom comprehension in terms of familiarity, semantic decomposability, and conceptual metaphor. Therefore, using this model can give us a more precise picture of how NSs and NNSs comprehend idioms. This study will investigate which of the three factors (i.e., familiarity, semantic decomposability, and conceptual metaphor) affect L1 and L2 language users’ idiom processing. Are there any interaction effects among these factors in idiom processing? Do NSs tend to conceive idioms as semantically nondecomposable, while NNSs tend to conceive idioms as semantically decomposable? What role does familiarity play in processing idioms in priming contexts and neutral contexts for NNSs and NSs?
To investigate these questions, the study conducted an experiment using a self-paced reading (SPR) task with native speakers of English and Chinese-speaking English learners as participants. The participants read 48 passages on a computer, progressing through the passages by pressing the spacebar at their own pace. The 48 passages included 12 priming contexts, 12 neutral contexts, and 24 fillers. In the priming contexts, the participants read one sentence priming the figurative meaning of an idiom before they encountered the idiom. In the neutral context, the participants did not read priming sentences before they encountered the idioms. The fillers did not contain any idioms. The reaction times (RT) for the sentences that contained idioms were recorded for analysis. It is assumed that the RTs in the priming contexts should be faster than those in the neutral contexts due to the priming of the figurative meanings. In other words, the participants should experience greater processing ease in the priming contexts than in the neutral contexts.

The participants took part in two sessions with an interval of one week. They encountered each idiom twice, once in each of the two contexts. If they encountered an idiom in the priming context in session 1, then they would encounter the same idiom in the neutral context in session 2, and vice versa. The RT differences between the two context types were used as the dependent variable in the analysis of the data. The reason to use the RT difference is that, otherwise, the different lengths of the idioms could become a confounding factor in the data analysis.

L1 adults are considered to be advanced language users; the interaction between degrees of familiarity and semantic decomposability is very likely to play an important role in explaining any RT differences between priming and neutral contexts for the NS participants. In contrast, L2 adult learners, classroom-taught foreign language learners in particular, are less advanced language users, so they are likely to
be less familiar with idioms. Consequently, the NNS participants might tend to rely on semantic decomposability and metaphorical equivalence between L1 and L2 to interpret the figurative meanings of the idioms. No study, to my knowledge, has examined whether there is any interaction effect between the factors of familiarity, metaphorical equivalence, and semantic decomposability in second language idiom processing. The importance of this question is that according to the dual idiom representation model, when people are unfamiliar with idioms, they will resort to semantic decomposability and conceptual metaphor to facilitate their comprehension. Familiarity is expected to be a significant factor for both either NSs or NNSs; in addition, I am interested in knowing whether it plays a role in NS idiom processing when interacting with semantic decomposability, or in NNS idiom processing when interacting with semantic decomposability and conceptual metaphor equivalence.

In addition to investigating idiom processing, the current study also investigates how idioms are stored in the NS and NNS mental lexicon. The same participants who completed the SPR task rated the degree of semantic decomposability of each idiom on a 5-point Likert scale. An idiom given one or two points by a participant was counted once as semantically nondecomposable, while an idiom given two or three points by a participant was counted once as semantically decomposable. The percentage of each idiom in each category was calculated separately for the two language groups, and the mean percentages were then calculated for each category for each language group. A threshold of 75% agreement (see Chapter 2 for the rationale for this threshold) was applied to investigate whether an idiom was conceived of as semantically decomposable or nondecomposable by NS and NNS language users. Because NSs are more familiar with idioms, compared with L2 learners, it is very likely that they tend to store idioms as a whole in the mental lexicon, which means
that they conceive idioms to be semantically nondecomposable. In contrast, NNSs tend to store idioms as their literal constituents in the mental lexicon, which means that they conceive idioms to be semantically decomposable.

Furthermore, familiarity is expected to have a main effect on the processing of idioms in the two types of contexts. According to the graded salience hypothesis, the salient meaning of an idiom will be activated initially regardless of the context, and the context will play a role as people encounter more information during the processing. Therefore, both NSs and NNSs are expected to process familiar idioms with ease. This is because the salient meanings of familiar idioms are the figurative meanings. However, they will both experience processing difficulty when encountering unfamiliar idioms. This is because the salient meaning (i.e., literal meaning) is inconsistent with the contextually appropriate meaning at first glance. The participants will initially activate the salient but contextually inappropriate meaning. As they encounter more contextual information, they will have to abandon the initially activated meaning, and then activate a contextually appropriate meaning.

This study aims to make a unique contribution to the research on idiom comprehension by using improved methodology and presenting a powerful theoretical model. Chapter 1 has briefly explained the importance of the topic, what has been done in the field, and the gaps in the previous studies. Chapter 2 reviews the theoretical background and empirical research, particularly on the role of context in idiom processing and idiom representation in the mental lexicon of both L1 and L2 language users, and presents five research questions. Chapter 3 describes the materials, the methods, and the procedures of the experiment. Chapter 4 presents the data analysis, the results, and discussion for each research question. Chapter 5
discusses earlier claims and this study’s findings, and how the findings can be useful in L2 idiom teaching and learning.
CHAPTER 2

LITERATURE REVIEW

This chapter will discuss the influence of context and language exposure on idiom processing and idiom representation. Context can help language users select the contextually appropriate meanings of idioms, namely figurative meanings rather than literal meanings. However, Peleg, Giora, and Fein (2001) found that context does not play as important a role as familiarity in the initial processing of idioms. Familiarity with formulaic expressions is related to amount of language exposure (Ellis et al., 2008). In addition, the chapter will discuss four idiom processing models in detail, namely the idiom list hypothesis (ILH; Bobrow & Bell, 1973), the lexical representation hypothesis (LRH; Swinney & Cutler, 1979), the configuration model (CM; Cacciari & Tabossi, 1988), and the dual idiom representation model (DIRM; Abel, 2003). The first three models were based on L1 data, so some problems emerge when they are applied to L2 idiom processing. In addition, each of the three models has its own limitations. However, the dual idiom representation model considers the role of the mental lexicon in idiom comprehension by both native speakers and bilinguals. Previous studies have investigated some of the factors in the dual idiom representation model, such as familiarity and semantic decomposability (Conklin & Schmitt, 2008; Siyanova et al., 2011; Titone & Connine, 1994) and metaphor (Boers, 2000a, 2000b; Littlemore, 2003; Littlemore, Chen, Koester, & Barnden, 2011). However, none of these studies has used the dual idiom representation model as a
framework to investigate bilinguals’ idiom comprehension. Therefore, the present study will use this model to investigate NS and NNS on-line idiom processing.

2.1 Context and familiarity in idiom processing

Idioms are considered to be ambiguous because they have both literal and figurative meanings. For example, “kick the bucket” can literally mean kicking a bucket with a foot, or figuratively mean “to die.” In a priming context, the literal meaning is considered to be contextually inappropriate, while the figurative meaning is considered to be contextually appropriate. On the contrary, in a literal context, the figurative meaning is considered to be contextually inappropriate, while the literal meaning is considered to be contextually appropriate. In other words, context plays a role in disambiguation.

Two viewpoints on how context affects access to contextually appropriate and contextually inappropriate meanings of idioms have spurred debates. According to the direct access view (Gibbs, 1994), lexical access and contextual information interact within one mechanism, and context plays a crucial role in initial processing. Therefore, contextually inappropriate meanings will be abandoned as early as possible. In the reordered access model (Duffy, Morris, & Rayner, 1988) as well, lexical access and contextual information interact within one mechanism, and context plays a crucial role in initial processing. This model argues that when people encounter words with multiple meanings, meaning dominance affects the relative activation of the multiple meanings. Thus, the only difference between the direct access view and the reordered access model is that the former expects contextually inappropriate meanings to be inhibited, while the latter does not. On the contrary: In the reordered access model, contextually appropriate meanings will be elevated and contextually inappropriate meanings will passively decay. Because there is no
inhibition of contextually inappropriate meanings, if the inappropriate meanings are the dominant meanings, people will take more time to process the ambiguous words, contra the direct access view.

Although the direct access view and the reordered access model differ in how they treat contextually inappropriate meanings, in the present study the two models explain the role of context in figurative language comprehension processes similarly. In both the priming context and the neutral context, a figurative interpretation of an idiom is contextually appropriate. In addition, the figurative interpretation, not the literal interpretation, is the dominant meaning of an idiom. Therefore, when people encounter the idioms in both types of context, the dominant, figurative meanings will be activated as early as possible. In other words, the processing times should not be significantly different in the priming and the neutral contexts.

Both of these viewpoints focus on the influence of context on idiom processing, and both argue that literal meanings will be abandoned and figurative meanings will be maintained. However, neither viewpoint can explain why Cieślicka’s (2006) participants were still able to activate the literal meanings of idioms after they had read the idioms’ last words. Moreover, the direct access view fails to explain why Cieślicka’s and Matlock and Heredia’s (2002) studies found that figurative meanings were activated no faster than literal meanings if literal meanings were abandoned as soon as possible. These two problems suggest that (a) language users might not “abandon” literal meanings during idiom processing and (b) there may be factors other than context that affect the processing speed of figurative and literal meanings.

Giora’s graded salience hypothesis (GSH; 1997) argued that the activation of the figurative or literal meaning of an idiom depends on the saliency of the meaning
in the mental lexicon. If an idiom is very frequent or familiar to language users, its salient meaning is the figurative meaning; otherwise, its salient meaning is the literal meaning. Furthermore, Peleg et al. (2001) investigated the influence of saliency and context on NSs’ processing of novel, one-word metaphors. They gave 60 undergraduate and graduate students in the United States a series of ambiguous sentences to read. When they were reading, a word string occurred either in the initial position or in the final position, and the participants had to decide whether it was a word or a nonword. The word string was either related to the salient (but contextually inappropriate) meaning of the ambiguous word in the sentence or related to the nonsalient (but contextually appropriate) meaning of the ambiguous word, or it was unrelated to the ambiguous word. The study found that saliency played an important role at the beginning of sentence processing because the salient, contextually inappropriate meanings were processed as fast as the nonsalient, contextually appropriate meanings, and that the importance of context outweighed saliency at the end of sentence processing because the salient, contextually inappropriate meanings were processed slower than the nonsalient, contextually appropriate meanings in the final position of the sentence. The finding that the salient, contextually inappropriate meanings were processed as fast as the nonsalient, contextually appropriate meanings in the initial processing suggests that lexical access and contextual information are two separate mechanisms, because contextually appropriate and inappropriate meanings are both initially activated, regardless of contextual compatibility. Moreover, Peleg et al.’s (2001) study supports my claim that there is a factor other than context that influences the processing speed of figurative and literal meanings, which is familiarity.
2.2 The graded salience hypothesis (GSH) and the mixed results of previous research

The graded salience hypothesis takes into account the influence of familiarity in the processing of ambiguous words. This strengthens the model’s power in explaining both L1 speakers’ and L2 learners’ idiom processing. For this reason, the graded salience hypothesis may help explain the inconsistencies in the results of research on L1 and L2 idiom processing that have been reported in the literature.

Many studies have found that the literal meanings rather than the figurative meanings of idioms are the salient meanings in the bilingual mental lexicon (Cieślicka, 2006; Kecskes, 2000). These findings support the literal salience hypothesis (LSH) (Giora, 1997, 1999, 2002, 2003). For example, Cieślicka (2006) asked 43 undergraduate students in Poland to take part in a crossmodal lexical priming experiment. They were all native speakers of Polish. They listened to English sentences and saw target words appear on the computer screen at two time points. One was in the final position of the idiom, and the other was in the penultimate position. The target words were related to either the figurative meaning of the idiom or the literal meaning of the idiom. The results showed that the NNS participants were more easily primed for the target words that were related to the literal meanings of the idiom constituents than for the target words that were related figuratively to the metaphorical interpretation of the idioms. Moreover, Kecskes (2000) investigated how NSs and NNSs understood situation-bound utterances (SBU), such as “get out of here” and “what can I do for you.” An SBU is a pragmatic unit whose meaning comes from both linguistic and extralinguistic factors. According to the cognitive-pragmatic view, the meaning of an SBU is shaped by metaphor and conceptual knowledge of the world, which is similar to other formulaic expressions, such as idioms. The study found that NNSs tended to rely on the literal meanings of the SBUs, but NSs could
retrieve functional meanings of SBUs directly without accessing the literal meanings. In other words, the literal salience hypothesis contends that for NNSs, literal meanings of ambiguous words are always salient.

However, other research does not support the literal salience hypothesis. Conklin and Schmitt (2008) asked native speakers of English and L2 learners to read idioms in figurative and literal contexts in self-paced reading. In the priming contexts, figurative meanings were the appropriate interpretations. In the literal contexts, literal meanings were the appropriate interpretations. In addition, the participants read control passages that did not contain idioms. Conklin and Schmitt found that both the native speakers of English and the L2 learners were able to process the figurative meanings and literal meanings of idioms equally faster than they could process the controls. The researchers claimed that NNSs, like NSs, could process formulaic expressions as wholes, which allowed them to process the expressions quickly, regardless of whether their meanings were figurative or literal. However, Siyanova et al. (2011) used eye-tracking to investigate English native speakers’ and L2 learners’ eye fixations and durations on idioms (e.g., first-pass reaction time, mean reaction time, and fixation counts), and found that the native speakers processed both figurative meanings and literal meanings faster than they processed controls, but the L2 learners processed literal meanings faster than figurative meanings and controls. In the priming contexts, figurative meanings were the appropriate interpretations (e.g., “at the end of the day”: figurative meaning, “eventually”). In the literal contexts, literal meanings were the appropriate interpretations (e.g., “at the end of the day”: literal meaning, “in the evening”). A control was a phrase similar to an idiom, but that had nothing to do with the idiom; the controls were created by replacing the last word of an idiom with another word to eliminate the figurative meaning (e.g., “at the end of
the researchers did not mention how they selected the replacement words (for instance, whether they based the choice on word frequency or another factor).

Siyanova et al. (2011) argued that the preceding context they provided in their study disambiguated the meanings of the idioms, which facilitated the NSs’ ability to activate contextually appropriate meanings. Therefore, they activated the figurative meaning in the priming context as fast as they activated the literal meaning in the literal context. However, the NNSs were not able to make use of the context to disambiguate meanings, so their reaction times (RTs) for the figurative meanings were slower than their RTs for the literal meanings.

The graded salience hypothesis (Giora, 1997), which highlights the relationship between familiarity and salient meaning in the bilingual mental lexicon, suggests one possible explanation for the inconsistency between Conklin and Schmitt’s (2008) results and Siyanova et al.’s (2011) results. Conklin and Schmitt did not discuss how familiar the native speakers of English and L2 learners who participated in their study were with the idioms. Siyanova et al. measured the L2 learners’ but not the L1 speakers’ familiarity with the idioms. It is very likely that the L1 speakers were more familiar with the idioms than the L2 learners due to greater prior exposure.

In Conklin and Schmitt’s (2008) study, it is possible that both the NS and NNS participants were familiar with all the idioms, and the salient meanings were the figurative meanings for both groups. If such were the case, then the participants may have activated the figurative meanings in both contexts. In other words, when the contexts were figuratively-biasing, the participants activated the contextually appropriate, figurative meanings. However, when the contexts were literally-biasing, the participants still activated the figurative meanings, which in these contexts were contextually inappropriate. The figurative meanings were still the salient meanings
that were more strongly activated in the mental lexicon than the nonsalient literal meanings. Therefore, although the results indicated that both NSs and NNSs activated figurative meanings and literal meanings equally fast, it is possible that Conklin and Schmitt were comparing the RTs of the figurative meanings themselves, instead of comparing the RTs of figurative meanings with the RTs of literal meanings.

Because Siyanova et al. (2011) measured only the L2 learners’ familiarity with the idioms, we do not know whether the two participant groups’ familiarity was comparable. Nevertheless, an explanation of the results based on familiarity may apply: If the L1 speakers were familiar with the idioms, even if they encountered the idioms in the literal contexts, they still activated the more salient, figurative meanings immediately. Therefore, the RTs in the literal contexts could be the result of activating the figurative meanings instead of the literal meanings. In the case of the L2 learners, it should be noted that their average familiarity rating was only 3.5 out of 5; therefore, the degree of familiarity may not have been sufficient to make the figurative meanings of the idioms the salient meanings. In other words, for the L2 learners, the salient meanings were still the literal meanings of the idioms, and incompatible with the priming contexts, so the L2 learners’ processing of figurative meanings in the priming contexts was slower than their processing of the salient, literal meanings in the literal context. In addition, Siyanova et al. (2011) identified the point of idiom uniqueness (Flores d’Arcais, 1993) for the L1 speakers and used that point to analyze the L2 learners’ RTs. The point of idiom uniqueness is the point where people recognize a phrase that they are reading as an idiom. The study found that for the L2 learners, the processing time before the point of idiom uniqueness was significantly longer in the priming context than in the literal context in terms of the total reading times and fixation counts. According to the graded salience hypothesis,
one possible explanation for this finding is that the salient meanings of the idioms were the literal meanings in the L2 learners’ mental lexicon (which supports the suggestion that the L2 learners in Siyanova et al.’s study were not sufficiently familiar with the idioms). Thus, the priming contexts would cause processing difficulty because the salient meanings of the idioms were contextually inappropriate. However, the processing difficulty was reduced when context came into play as the participants continued to read the sentences. On the contrary, the literal context would facilitate processing because the salient meanings of the idioms were contextually appropriate. Therefore, the initial words of an idiom with a contextually inappropriate meaning in a priming context (i.e., the literal meaning) should take longer to process than those of an idiom with a contextually appropriate meaning in a literal context (i.e., the literal meaning). Furthermore, the differences in the RTs in both types of context were gradually reduced as the participants read the sentences to the end, so that the RTs of the later words in the idioms did not show a significant difference between the two types of context.

In summary, both Conklin and Schmitt’s (2008) study and Siyanova et al.’s (2011) study investigated NNS and NS idiom processing. The former used self-paced reading, while the latter used eye-tracking. The former found that NSs and NNSs activated figurative meanings as fast as literal meanings. The latter found that NSs activated figurative meanings as fast as literal meanings, but NNSs activated literal meanings faster than figurative meanings. Both studies indicated that the preceding context plays an important role in the activation of figurative and literal meanings. However, Conklin and Schmitt (2008) argued that both NSs and NNSs were good at using contextual clues to select appropriate meanings, while Siyanova et al. (2011) argued that NNSs were not as good as NSs at using contextual clues. One problem in
interpreting the cause of this difference in results is that the studies do not take familiarity into consideration. Familiarity, as the graded salience hypothesis (Giora, 1997) recognizes, is crucial to both NS and NNS idiom processing.

Given the importance of familiarity, the present study investigates the influence of familiarity on idiom comprehension.

2.3 Language exposure and idiom comprehension

The studies mentioned in section 2.1 illustrate the importance of familiarity in idiom comprehension. In addition to familiarity, some other studies look at the influence of language exposure, or language proficiency, on idiom comprehension.

Matlock and Heredia (2002) found that familiarity is related to language proficiency. The researchers asked native speakers of English and bilinguals (early bilinguals and late bilinguals) to read sentences on a computer. The sentences included two verb types, namely phrasal verbs and identical verb + preposition. The participants’ reading times were recorded. Results showed that native speakers of English and early bilinguals processed phrasal verb sentences significantly faster than verb + preposition sentences. No verb type effect was found for late bilinguals. Therefore, the researchers concluded that the order of activating literal meanings and figurative meanings of idioms was related to learners’ language proficiency. The late bilinguals had been exposed to the target language less than the native speakers and early bilinguals, and had not mastered the target language, so they usually relied on L1 literal translation. However, when late bilinguals become more proficient, they are able to retrieve figurative meanings without L1 literal translation. In contrast, the native speakers and early bilinguals were advanced language users, so they were able to retrieve the figurative meanings first. The results are consistent with the graded salience hypothesis’s argument that familiarity decides the salient meaning, and the
salient meaning is always activated first. Moreover, familiarity is related to language exposure, so L1 speakers’ degree of familiarity with idioms differs from L2 learners’. This finding further supports my claim that in Siyanova et al.’s (2011) study, the NSs’ degree of familiarity with idioms probably differed from the NNSs’, so it is important to measure the familiarity of different participating language groups in order to interpret their reaction times.

In addition, Ellis et al. (2008) investigated factors affecting L1 speakers’ and L2 learners’ processing of formulaic expressions, and found that L1 speakers tended to rely on the form–meaning–function relationship of the formulaic expressions, while L2 learners tended to rely on the frequency of the formulaic expressions. The discrepancy is fundamentally due to language exposure. The frequency of occurrences of a formulaic expression play an important role in learning. According to the power law of learning (Newell, 1990; Speelman & Kirsner, 2005), the effects of repeated language exposure are more significant at earlier stages of learning than at later stages. For example, the effect of 10 extra exposures on processing speed is more obvious if the learner has been exposed to the target form five times before than if the learner has been exposed to it five hundred times before. This implies that less advanced language users can obtain a processing advantage through encountering increased occurrences of formulaic expressions. Advanced language users (e.g., native speakers), on the other hand, will not obtain such a processing advantage because they do not have to rely on the frequency of occurrences to recognize formulaic expressions; instead, they can rely on the coherence between the form, the meaning, and the function of formulaic expressions.

The link between form, meaning, and function of a formulaic expression is similar to Abel’s (2003) “idiom entry.” An idiom entry is built for an idiom in the
mental lexicon. A familiar idiom is easier than an unfamiliar idiom to store and retrieve in the mental lexicon as an idiom entry, because language users have a form–meaning–function link for a familiar idiom, but not for an unfamiliar one.

Abel (2003) investigated the relation of familiarity and language exposure, and found that adult L1 speakers were more familiar with idioms than L2 learners in a self-rating task. This is because the L1 language users had been exposed to the idioms more often than the L2 learners, resulting in strengthened connections between the form (e.g., “hit the ceiling”) and the function (e.g., describing anger). The strengthened connections are important properties of idiom entries. In contrast, the L2 learners had a tendency to rely on the frequency of the individual words to recognize idioms. However, recognizing high-frequency words does not transfer to recognizing the formulaic expressions that contain the high-frequency words if the language users have not acquired the ability to recognize the coherence of the forms and meanings of the formulaic expressions (Ellis et al., 2008). This explanation may also account for Littlemore et al.’s (2011) finding that L2 learners had difficulty understanding metaphoric expressions even if the expressions were composed of familiar words. The L2 learners in the study were familiar with the high-frequency words, but knowing the meanings of the high-frequency words could not guarantee their understanding of the idioms unless the learners also had the connections of the idioms’ forms and meanings in their mental lexicon or knew the idioms’ L2 metaphoric referents. A metaphoric referent can help language users link the literal meaning of constituents to the figurative meaning of an idiom. Given the importance of language exposure in idiom comprehension, the present study compares L1 speakers’ and L2 learners’ idiom processing.
2.4 L1 idiom processing models and their limitations

The graded salience hypothesis (Giora, 1997) posits that familiarity is crucial to idiom comprehension because it affects the order of activating figurative and literal meanings. Although the three traditional models of L1 idiom processing introduced in Chapter 1 also address the order of activating figurative and literal meanings, they do not take familiarity into account.

2.4.1 The idiom list hypothesis

Bobrow and Bell (1973) conducted two experiments with 414 undergraduate students in the United States as participants. The participants were asked to read a set of four sentences. In one experiment (i.e., the unambiguous condition), the last sentence only had a literal interpretation. In the other experiment (i.e., the ambiguous condition), the last sentence had both a literal and a figurative interpretation. In a post test, the participants had to indicate which meaning, namely the literal or figurative meaning, they perceived first. The results showed that in the unambiguous condition, the participants tended to indicate that they perceived the literal meanings first. In the ambiguous condition, the participants tended to indicate that they perceived the figurative meaning first. The researchers argued that the reason why the participants perceived figurative meanings first in the ambiguous condition was because they activated an “idiom mode” of processing, which was different from the processing mode they activated in the unambiguous condition.

Bobrow and Bell’s (1973) study is specifically about idioms. Clark and Lucy (1975) conducted a study on requests, which also produced some results that may be relevant to idiom processing. They asked 23 participants to read 10 pairs of conversationally conveyed requests. Each pair contained a positive request (e.g., “Can you make the circle blue?”), and a negative request (e.g., “Must you make the circle
blue?”). The participants saw the requests on the left side of a computer screen, and a picture (e.g., a circle that was either pink or blue) on the right. They were asked to judge whether the picture on the right correctly reflected the request on the left by pressing a “yes” button or a “no” button on a keyboard as quickly as possible. The researchers found that their participants interpreted the requests literally first; they did not interpret them nonliterally unless they found that the literal interpretations were inappropriate in the presented contexts. The results seem to support the idiom list hypothesis in that they suggest that a special mode, namely an idiom mode, is activated when people process nonliteral language. This idiom mode remains unactivated when people process literal language.

The idiom list hypothesis assumes that literal meanings are always processed prior to nonliteral meanings, but the model does not clearly state the reason for this assumption. Notably, the research supporting the idiom list hypothesis is based on off-line tasks. These off-line tasks can provide information on postperceptual processes, but not necessarily perceptual processes. Therefore, this model does not contribute to understanding whether literal meanings are processed prior to nonliteral meanings in on-line processing.

2.4.2 The lexical representation hypothesis

Swinney and Cutler (1979) asked 62 undergraduate students in the United States to participate in two experiments measuring RTs for determining whether word strings were composed of acceptable English phrases. The results showed faster RTs for idioms (e.g., “break the ice”) than for controls (e.g., “break the cup”). The controls were created by replacing the first or the last word in the idiom with a word of the same length and of equal or higher frequency. The results supported the lexical representation hypothesis that figurative meanings are processed faster than literal
meanings because figurative meanings have been lexicalized in the mental lexicon, while literal meanings have not.

Lexicalization is an important concept in the lexical representation hypothesis. Lexicalization means that a formulaic expression is not conceived of as individual words but rather as a single large lexical item. In the second experiment, the researchers investigated whether the participants would react differently to idioms at different levels of syntactic frozenness. Syntactic frozenness was operationalized based on four levels of syntactic transformation (Fraser, 1974). Level 1 means no syntactic transformation (e.g., “to let off some steam”). Level 2 means a mild level of syntactic transformation, as in “kick the bucket.” “Kick the bucket” can undergo a simple transformation, such as “John’s kicking the bucket was tragic,” but cannot undergo passivization (e.g., “The bucket was kicked by John.”). Level 3 means an intermediate level of syntactic transformation (e.g., “Throw in the towel” can undergo most transformations), and Level 4 means the maximum level of syntactic transformation (e.g., “lay down the law”). The results showed no main effect of frozenness level on reaction time. The researchers concluded that the idioms were lexicalized, and therefore were retrieved as whole units, so the levels of frozenness did not affect the reaction times. The assumption of the lexical representation hypothesis is that the lexicalized meanings of idioms are always the figurative meanings. However, it fails to explain what happens when someone’s lexicalized meaning of an idiom is not the figurative meaning in the mental lexicon. Levorato (1993) showed that it is likely that the lexicalized meaning is not always the figurative meaning. For instance, for the younger children in the study, the lexicalized meaning of an idiom was the literal meaning.
Levorato (1993) investigated whether the interpretation of figurative language changed as a function of the development of language proficiency. Two age groups of children (i.e., eight years old and ten years old) were asked to read stories that ended with idioms, which had been evaluated by teachers as high or low familiarity idioms. The last word of the idioms was missing. The children had to fill in the blanks to complete the stories. It was found that the older children knew more high familiarity idioms than the younger children. Interestingly, although both age groups failed to complete the stories ending with low familiarity idioms in an idiomatic way, the older children’s answers were closer to the figurative meaning of the idioms than the younger children’s answers. For example, one of the stories in Levorato’s materials was about a boy who transferred to a new school. His classmate wanted to borrow a crayon from him. The boy decided to loan the crayon to his classmate because he wanted to break the ______. The idiomatic answer was “ice.” However, when the younger children were unfamiliar with the idiom, they filled in the blank in a literal way, such as with “crayon.” In contrast, the older children who were unfamiliar with the idiom could still select a figurative completion, such as “shyness.”

The results indicated that: (a) not all idioms were lexicalized for the children because they were still unfamiliar with certain idioms; (b) when the idioms were unfamiliar to them, the older children were more capable of using certain resources to find the metaphoric referents of the idioms, compared with the younger children. Take “break the shyness,” for instance; “shyness” is the metaphoric referent of “ice” in “break the ice,” so when the older children had not learned the conventional way of expressing shyness (i.e., “ice”) in this idiom, they tended to use the metaphoric referent directly. The use of metaphoric referents suggests that the older children resorted to the conceptual level in their mental lexicon to attempt to find a possible,
appropriate word to fit the context when they did not know the correct word at the lexical level; and (c) lack of lexicalized idioms in the mental lexicon did not necessarily cause the children to comprehend idioms literally.

In brief, Levorato’s (1993) study indicated that not all idioms’ lexicalized meanings are figurative meanings in less advanced language users’ mental lexicon. The assumption of lexicalization in the lexical representation hypothesis (i.e., that the lexicalized meanings of idioms are figurative meanings) prevents it from illustrating how idioms are represented in language learners’ mental lexicon: The lexicalized meanings are not always figurative meanings; they can be literal meanings.

2.4.3 The configuration model

Cacciari and Tabossi (1988) asked 102 undergraduate students in Italy to participate in three crossmodal priming experiments. Experiment 1 was about predictable idioms. Experiments 2 and 3 were about unpredictable idioms. In the crossmodal task, the participants listened to sentences that contained either idiomatic strings (e.g., “in seventh heaven”) or literal word strings (e.g., “in seventh position”). After they listened to each sentence, they saw a word on a computer screen. The word could be semantically related to the figurative meaning of an idiom, semantically related to the meaning of the last word in a literal word string, or semantically unrelated to the idiom or the literal word string. In Experiments 1 and 2, the words were displayed immediately after the participants listened to the sentences. In Experiment 3, the words were displayed 300 milliseconds (ms) after the participants listened to the sentences. Before the crossmodal priming tasks, 15 participants, who did not participate in the tasks, were asked to paraphrase a set of familiar idioms. The researchers did not discuss how they evaluated idioms as familiar. Only idioms whose
meanings were paraphrased correctly by at least 95% of the participants were considered to be predictable idioms. Moreover, another 15 participants were asked to complete fragments as sentences in five lists. In List 1, the fragments were the shortest (e.g., “He told…”). In List 5, the fragments were the longest (e.g., “He told him to go to the…”). Idioms were considered to be unpredictable if more than 5% of the participants failed to complete the fragments. The researchers did not explain why they chose the paraphrase task to select predictable idioms. A possible explanation is that predictable idioms are likely to be familiar, and therefore could be paraphrased correctly by most people. The researchers found that (a) when the idioms were predictable, the participants reacted faster to idiomatically related targets than to literally related targets; (b) when the idioms were nonpredictable, the participants reacted faster to literally related targets than to idiomatically related targets; and (c) when the idioms were nonpredictable, and the targets were presented 300 ms after the idioms were heard, the participants reacted faster both to the idiomatically related targets and the literally related targets than to the control targets. The researchers proposed that idioms contain a key word, which functions as the point of idiom uniqueness, and which helps language users recognize the idiom: Before the idiom is recognized, it is processed literally, but once the point of idiom uniqueness is identified, the idiom is processed figuratively.

Drawing on their notion of a point of idiom uniqueness, Cacciari and Tabossi (1988) claimed that there is no specific idiom mode for idiom processing. Instead, the activation of figurative meanings relies on the point of time when the point of idiom uniqueness is identified. However, the researchers admitted that there was no criterion to identify the point of idiom uniqueness. In addition, the point of idiom uniqueness could change in different contexts.
In summary, the configuration model is similar to the lexical representation hypothesis because both claim that there is no idiom mode specific to idiom processing, which means there are no special processing strategies for literal interpretations or for figurative interpretations. However, the lexical representation hypothesis claims that figurative meanings will always be processed first, while the configuration model claims that figurative meanings will only be processed when the point of idiom uniqueness is identified. Moreover, the lexical representation hypothesis claims that idioms are lexicalized, which implies that idioms are conceived of as nondecomposable in the mental lexicon. However, it does not address the processing of nonlexicalized idioms. As for the idiom list hypothesis, literal meanings will always be processed first, and figurative meanings will only be processed when the literal meanings do not fit the contexts. The idiom list hypothesis does not take into account the possibility that figurative meanings are activated initially.

These three models are all based on L1 speakers’ data, and do not consider L2 learners. Therefore, it is problematic to attempt to apply any of these processing models to L2 idiom processing. One of the problems is that L1 speakers’ familiarity differs from that of L2 learners (Matlock & Heredia, 2002), but these three L1 idiom processing models do not take familiarity into consideration. Furthermore, none of these models explains what language users do if they encounter idioms with which they are unfamiliar. As mentioned previously, if language users have not acquired the form–meaning–function connection of an idiom, they may rely on L1 translation or L2 metaphoric referents of the idiom to understand the figurative meaning. The degree of ease of linking literal constituents to L2 metaphoric referents depends on the degree of semantic decomposability of an idiom.
Gibbs and Nayak (1989) asked participants to rate idioms in terms of semantic decomposability, metaphoric transparency, literal well-formedness, and syntactic productivity on 7-point Likert scales. A correlation analysis found that semantic decomposability and metaphoric transparency had a positive correlation. The result indicated that when an idiom is semantically decomposable, the overall figurative meaning is easily derived from its constituent meanings. In other words, the metaphoric referents and the literal meanings of the constituents share the same semantic field (e.g., the metaphoric referent is an extension of the literal meaning of the constituent), which results in high metaphoric transparency. Therefore, people can easily find the metaphoric referent of a semantically decomposable idiom. On the contrary, the metaphoric referent of a semantically nondecomposable idiom does not share the same semantic field with the constituent meanings of the idiom, resulting in low metaphoric transparency. Therefore, people usually have difficulty finding its metaphoric referent. Consequently, people also have difficulty understanding the figurative meaning of a semantically nondecomposable idiom. In brief, the degree of semantic decomposability of an idiom is positively correlated to the ease of finding its metaphoric referents.

In summary, semantic decomposability and conceptual metaphor are two important factors in idiom comprehension, which are not addressed in the three traditional L1 idiom processing models. To address the limitations of these three models (i.e., the idiom list hypothesis, lexical representation hypothesis, and configuration model), an alternative hypothesis is needed, and it must take L2 language learners into consideration. The dual idiom representation model (DIRM; Abel, 2003) provides such a hypothesis.
2.5 The dual idiom representation model

The dual idiom representation model can explain not only how NSs understand idioms, but also how NNSs understand idioms. In addition to this advantage, the dual idiom representation model takes into account certain important features of idiom comprehension, which are not accounted for by the idiom list hypothesis, the lexical representation hypothesis, or the configuration model. These features are reflected in its four central assumptions. First, a decomposable idiom has both an idiom entry and constituent entries in the mental lexicon, while a nondecomposable idiom has an idiom entry only. Previous studies on the priming effects of compounds support this assumption (Sandra, 1990; Zwitserlood, 1994). These studies found that for decomposable compounds, both constituents showed priming effects. For example, the Dutch word *kerkorgel* is a compound (*kerk*: church; *orgel*: organ). Zwitserlood (1994) used *kerkorgel* as a prime, and *priester* (priest) as a target, semantically related to the first constituent *kerk*. In addition, *muziek* (music) was used as a target, semantically related to the second constituent *orgel*. However, for nondecomposable compounds, only the first constituent showed a priming effect. For example, Dutch *klokhuis* is a compound (*klok*: clock; *huis*: house), which means “core of an apple.” Zwitserlood used *klokhuis* as a prime, and *tijd* (time) as a target, semantically related to the first constituent *klok*. In addition, *tuin* (garden) was used as a target, semantically related to the second constituent *huis*.

The results suggested that the decomposable compounds had lexical representations for the constituents and other lexical representations for the compounds as a whole. However, the nondecomposable compounds did not have lexical representations for all the constituents, but only lexical representations for the compounds as a whole. In this sense, the nondecomposable compounds were like
monomorphemic words, having single lexical representations of the whole words. Decomposable idioms, analogous to decomposable compounds, could be represented by the lexical entries of the constituents, while nondecomposable idioms, analogous to nondecomposable compounds, could be represented only by a separate lexical representation.

Second, a high-familiarity, decomposable idiom can develop an idiom entry more easily than a low-familiarity, decomposable idiom. Morphological studies have found that high-frequency morphologically complex words were accessed as a whole despite the existence of constituent entries, while low-frequency morphologically complex words were accessed through their individual morphemes (Caramazza, Laudanna, & Romani, 1988; Frauenfelder & Schreuder, 1992). Abel (2003) distinguished familiarity from frequency, and focused on the frequency effect. According to Abel, frequency refers to the objective occurrence of a linguistic unit that can be statistically measured. Familiarity refers to a subjective measure that exists at the conceptual level. It is usually measured by a rating scale (Abel, 2003; Gibbs & Nayak, 1989; Titone & Connine, 1994). However, as Abel mentioned, high frequency words result in high familiarity, but not vice versa. L2 learners, for example, may be familiar with certain expressions due to repeated instruction in school, but those expressions may be infrequently used by native speakers of the target language. In the same vein, some high-frequency expressions used by native speakers may be unfamiliar to L2 learners due to lack of exposure. Therefore, the present study uses familiarity rather than frequency to better illustrate the lexical representations in the bilingual mental lexicon.

An idiom entry is like a place in the mental lexicon that holds the information about the form, the meaning, and the function of an idiom. In other words, it is a
“form–function composite,” in Nattinger and Decarrico’s (1992) terms. Moreover, Ellis et al. (2008) maintained that NSs perceived formulaic expressions as coherent wholes because they can identify the distinctive meanings of formulaic expressions. That is, formulaic expressions, such as idioms, have strong form–meaning connections. The composite of form, meaning, and function allows an idiom to be processed faster than other idioms that have not formed such a composite. It is much easier to form strong connections between form, meaning, and function with idioms familiar to language users than with idioms unfamiliar to language users because repeated language exposure can strengthen form–meaning connections (Ellis et al., 2008).

Third, when an idiom does not have an idiom entry at the lexical level (e.g., unfamiliar idioms), an alternative is to link the literal meanings of its constituents to the metaphoric referents at the conceptual level. The older children in Levorato’s (1993) study, for example, knew the metaphoric referent “shyness” for “ice” in “break the ice,” but had not learned that “ice” was the conventional way of expressing shyness in this particular idiom. Therefore, what they would need to do in order to learn the idiom would be to make a connection between the literal meaning of “ice” and its metaphoric referent “shyness.” Furthermore, the activation of conceptual metaphors is easier for some idioms (e.g., “miss the boat”) than others (e.g., “kick the bucket”) as a function of semantic decomposability. Gibbs and Nayak (1989) tested the idiom decomposability hypothesis (IDH) and found that people were more able to find the metaphoric referents of constituents when the idioms were judged to be semantically decomposable, and had more difficulty finding the metaphoric referents when the idioms were judged to be semantically nondecomposable. In brief, the dual idiom representation model indicates that semantic decomposability is related to the
link of constituents to metaphoric referents, and the IDH further indicates that degree of semantic decomposability is positively related to finding metaphoric referents.

Fourth, native speakers have more idiom entries than L2 learners as a result of extensive language exposure. Furthermore, Abel (2003) found that native speakers tended to judge idioms to be nondecomposable, while L2 learners tended to judge idioms to be decomposable. In the same study, L2 learners who self-reported being exposed to English every day had a tendency to judge idioms to be nondecomposable. In other words, the perception of semantic decomposability of an idiom is a function of the amount of language exposure. This suggestion corresponds to the integrated model (Wray & Perkins, 2000) in that, as language exposure increases, L1 language learners tend to view a frequently occurring word string as a formulaic expression. As long as the word string is viewed as a fixed, formulaic expression, it is conceived of as one big word rather than a string of individual words. Therefore, the fixed, formulaic expression, or the big word, tends to be conceived of as nondecomposable.

Conceptions of semantic decomposability reflect how idioms are stored in the mental lexicon. Several studies have investigated how NSs perceive the semantic decomposability of idioms (Abel, 2003; Gibbs & Nayak, 1989; Titone & Connine, 1994). Gibbs and Nayak (1989) precategorized idioms as normally decomposable, abnormally decomposable, and nondecomposable. Normally decomposable means that the constituents of the idioms are directly related to the figurative meanings, such as *pop the question*. *Pop* is related to “suddenly asking,” and *question* refers to a “marriage proposal.” Abnormally decomposable means that the constituents have a close relationship with the figurative meanings, but in a metaphorical way, such as *spill the beans* (“reveal a secret”). *Spill* is closely related to “reveal,” but *beans* is related to “secret” in a less direct, metaphorical way. Nondecomposable means that
the constituents have little relationship with the figurative meanings, such as *chew the fat* (“to talk aimlessly”). The participants were 24 English-speaking undergraduate students in the United States, and they were asked to express whether they agreed with the precategorization. A 75% agreement threshold was applied for data analysis, which means an idiom was regarded as one of the idiom categories at least 75% of the time. The results showed that the mean proportion of agreement was 86% for normally decomposable idioms, 79% for abnormally decomposable idioms, and 88% for nondecomposable idioms.

One limitation of Gibbs and Nayak’s (1989) study is that the idioms were precategorized by the researchers, instead of being categorized by the participants, which may have affected the results. Therefore, Titone and Connine (1994) carried out a similar study but allowed participants to freely categorize the idioms into one of the three categories. Fifty-six English-speaking undergraduate students in the United States were asked to freely categorize the semantic decomposability of the idioms taken from the studies done by Gibbs and Nayak (1989) and Gibbs, Nayak, and Cutting (1989). A 75% agreement threshold was applied, resulting in the categorization of 35% of the idioms as nondecomposable and 15% as decomposable. The researchers found that the participants’ categorization of abnormally decomposable idioms was unreliable; for this reason, I do not investigate this type of idiom in the present study. Because of the unreliability of categorizing abnormally decomposable idioms, Abel (2003) carried out a similar study with NNS participants in which he excluded the category of abnormally decomposable idioms. Abel asked 30 non–English speaking participants to categorize the 130 idioms taken from Titone and Connine’s (1994) study into semantically decomposable and semantically nondecomposable idioms. A 75% agreement threshold was applied, based on which
17.7% of the idioms were nondecomposable and 29.2% were decomposable. In general, NSs tend to perceive idioms as semantically nondecomposable (Titone & Connine, 1994), while NNSs tend to perceive idioms as semantically decomposable (Abel, 2003).

These three studies used a 75% agreement threshold for participants’ categorization. Titone and Connine (1994) investigated what percentage could be used as the agreement criterion. The researchers ran a Z approximation of the binominal distribution for the decomposability/nondecomposability sorting task, and found that 73% agreement in one category was significant at \( p < .01 \) in a two-tailed test. The null probability equals .5. This means the null hypothesis is that the probability of 73% of participants sorting an idiom into a category is .5, which implies the sorting is by chance. However, Titone and Connine found that the probability that 73% of the participants sorted an idiom into one category was less than .01. If this extreme probability occurs given the null hypothesis is true, the null hypothesis should be rejected. Moreover, when an idiom was rated as decomposable or nondecomposable at least 75% of the time, it means the rating result was at least significant at \( p < .01 \) in a two-tailed test.

In summary, the dual idiom representation model explains how idioms are stored and retrieved. Familiarity, semantic decomposability, and conceptual metaphor are the three factors playing an important role in idiom comprehension. Familiarity affects whether an idiom has an idiom entry or consists of constituent entries. Semantic decomposability affects how easily one can link the constituent meanings to the metaphoric referents. Both NSs and NNSs are more familiar with semantically decomposable idioms than semantically nondecomposable idioms. Furthermore, NNSs may make more use of conceptual metaphors to figure out the figurative
meanings of idioms than NSs, because conceptual metaphor is especially helpful for idiom comprehension when people do not have idiom entries.

Comparison of the dual idiom representation model with the three previous models

The dual idiom representation model highlights the importance of familiarity, semantic decomposability, and language exposure in L2 idiom processing. These three factors are not considered in the previous three models, namely the idiom list hypothesis, lexical representation hypothesis, and configuration model. Familiarity and semantic decomposability affect whether idioms are stored and retrieved as idiom entries, constituent entries, or both in the mental lexicon. When idioms are represented as idiom entries (e.g., familiar idioms) in the mental lexicon, they are lexicalized and processed as a whole. In contrast, when idioms are represented as constituent entries without any idiom entries (e.g., unfamiliar idioms) in the mental lexicon, they have not been lexicalized yet. According to Wray and Perkins (2000), one of the advantages of processing lexicalized phrases is that it can be done with increased speed, which allows people more time to be engaged in more complex cognitive activities (e.g., producing complex sentences). Therefore, idioms that are stored as idiom entries should be processed faster than idioms that are stored as constituent entries solely.

Furthermore, the length of language exposure affects the degree of familiarity and the perception of semantic decomposability, which in turn affects the amount of idiom entries and constituent entries in the mental lexicon. NSs are exposed to the target language longer than NNSs—classroom-taught foreign language learners in particular—so they are more familiar with idioms than NNSs, and have more idiom entries than NNSs do. Therefore, the fourth assumption of the dual idiom
representation model can be used to explain the findings from many studies that NSs process idioms faster than NNSs (Conklin & Schmitt, 2008; Siyanova et al., 2011; Underwood, Schmitt, & Galpin, 2004). Again, the idiom list hypothesis, lexical representation hypothesis, and configuration model cannot address the effects of length of language exposure, because the three models do not take the NNS mental lexicon into consideration.

In addition, in Siyanova et al.’s (2011) study, for example, NSs were found to process figurative meanings as fast as literal meanings, and NNSs were found to process figurative meanings more slowly than literal meanings. The NS finding was not consistent with the idiom list hypothesis, lexical representation hypothesis, or configuration model because the idiom list hypothesis and configuration model both argue that literal meanings should be activated faster than figurative meanings and the lexical representation hypothesis argues that figurative meanings should be activated faster than literal meanings. The dual idiom representation model, however, can account for Siyanova et al.’s results: The NSs were familiar with the idioms, and the familiar idioms have both idiom entries and constituent entries in the mental lexicon, so that when the NSs encountered familiar idioms, they were able to retrieve the figurative meanings from the idiom entries as fast as they retrieved the literal meanings from the constituent entries. In this sense, the dual idiom representation model can account for NS idiom processing better than the three traditional models.

The finding that the NNSs processed figurative meanings more slowly than literal meanings is not consistent with the lexical representation hypothesis either. However, viewed in terms of the dual idiom representation model, the NNSs were unfamiliar with the idioms, so they had more constituent entries than idiom entries in their mental lexicon, which means that they had to find the literal meanings of the
constituents of the idioms first, and make a connection between the literal meanings and their metaphoric referents in order to find the figurative meanings of the idioms. Therefore, the NNSs took longer to activate the figurative meanings of the idioms than the literal meanings of the idioms.

2.6 The bilingual lexical representation of idioms

The bilingual mental lexicon is composed of the words of the L1 and the L2. This section begins by discussing three models of the bilingual mental lexicon, namely the revised hierarchical model (RHM; Kroll & Stewart, 1994), the distributed feature model (DFM; De Groot, 1992, 1993), and the modified hierarchical model (MHM; Pavlenko, 2009). The three models explain how bilinguals deal with L1 and L2 nonformulaic words in terms of forms and lexical concepts. The purpose of the present discussion is to identify an appropriate model to explain how idioms are represented in the bilingual mind.

The revised hierarchical model claims that an L1 word and its L2 counterpart share the same concepts. As shown in Figure 1 below, the L1 word and the concepts have a strong conceptual link (i.e., the solid line in the figure), but the L2 lexical counterpart has a weak conceptual link or even no conceptual link with the concepts (i.e., the dotted line in the figure). The meaning of the L2 word can be activated by its lexical link to the L1 counterpart first, and then connected to the relevant concepts through the strong L1 conceptual link. Despite the strong L2–L1 lexical link, the lexical link of the L1 word to its L2 counterpart is weak. According to Chen and Leung (1989) and De Groot (2002), in translation tasks, participants took longer to translate from L1 to L2 than from L2 to L1. In addition, in picture naming tasks, which require the ability to link words and concepts, the participants took more time to name a picture in L2 than in L1. This is because the links between concepts and L2
words are weak or even nonexistent; thus, naming in L2 takes longer in comparison with naming in L1, which has a strong link between concepts and L1 words.

Figure 1. Revised hierarchical model (adopted from Pavlenko, 2009)

The revised hierarchical model takes into account the development of a person’s language proficiency, so that a weak link can become stronger as one’s language proficiency improves. However, the model offers no explanation about how NNSs deal with partially shared concepts between L1 and L2 words. It is very likely that an L1 word and its L2 equivalent do not refer to exactly the same concepts. Therefore, De Groot (1992, 1993) proposed the distributed feature model to address crosslinguistic differences. As shown in Figure 2, *idée* is the French equivalent of *idea*. The four white rectangles represent the concepts unique to *idée* and *idea*, respectively. The two black rectangles represent the concepts shared by the two words.
The more features shared by L1 and L2 words, the less time one needs for translation. However, the distributed feature model does not explain how learners’ access to the partially shared concepts of L2 words might change as their L2 language proficiency develops. Moreover, the model, which takes a feature-based approach, does not reflect how learners access the core features of a word and the peripheral features of that word. For instance, when a learner processes “bank” in “she went to a bank to deposit her money,” her/his processing time should be faster than when processing “bank” in “she went to the bank in order to ride in a boat.” This is because “bank” in the first sentence is used in its core meaning, while in the second sentence it is used in a peripheral meaning. In addressing the limitations of the revised hierarchical model and distributed feature model, Pavlenko (2009) put forward the modified hierarchical model. The modified hierarchical model attempts to maintain the strengths of the two previous models, namely the development of strong links between L1 and L2 at the lexical and conceptual levels, and the possibility of partially shared concepts between L1 and L2 words. A difference between the modified hierarchical model and the dual idiom representation model is that the modified hierarchical model addresses partially shared concepts between L1 and L2 words, while the dual idiom representation model does not.
The modified hierarchical model (Pavlenko, 2009), which is visualized in Figure 3, claims that the bilingual mental lexicon is composed of one shared conceptual component (i.e., ○3, ○4, and ○5) and two separate lexical components (i.e., ○1 and ○2).

![Modified hierarchical model](image)

*Figure 3. Modified hierarchical model (adopted from Pavlenko, 2009)*

In Figure 3, the solid line means the connection is strong, and the dotted line means that the connection is weak. The lexical level stores L1 and L2 word forms, which is different from the notion of the lexical level in the dual idiom representation model. In the dual idiom representation model, the lexical level stores the form and the meaning of words. Despite the difference between the two models, the present study borrowed the dual idiom representation model’s idea of differentiating the concepts of form and meaning, and incorporated this differentiation into the modified hierarchical model. Moreover, the dual idiom representation model only proposes a general conceptual level, but does not identify the differentiation of L1 and L2 concepts. Based on previous studies (Boers, 2000b; Pavlenko, 2008, 2009), we know
that L2 learners’ concepts are influenced by their L1 concepts, which leads to L1-specific categories, L2-specific categories, and shared categories at the conceptual level. The shared categories contain equivalent concepts or partially equivalent concepts. The other two categories contain language-specific concepts. The differentiation in the conceptual component may lead to comprehension difficulty. For example, Pavlenko (2008) found that in a film recall task, Russian-speaking English learners showed negative transfer, lexical borrowing, and even avoidance in their L2 narration when the L1–L2 words were either partially equivalent or nonequivalent. Pavlenko (2008, 2009) argued that when dealing with L2 words that have partially equivalent L1 conceptual counterparts, L2 learners need to restructure the concepts. When dealing with L2 words that have conceptually nonequivalent L1 counterparts, L2 learners need to develop new L2 conceptual categories. However, it is relatively simple to deal with conceptually equivalent words. The L1 words have a positive transfer to their L2 counterparts, so L2 learners just have to link the L2 words and the existing L1 concepts. The more effort L2 learners have to put into finding the relevant L1 concepts for the L2 words, the more difficulty they have in comprehending and producing the L2 words.

In the present study, it is assumed that when L1 and L2 share more similarity in metaphor, L2 learners should have less difficulty processing idioms because of the positive influence from the L1. On the contrary, when L1 and L2 share less similarity in metaphor, L2 learners might experience more difficulty processing idioms because of the negative influence from the L1. Because L1–L2 metaphoric equivalence is crucial to L2 idiom processing, the present study takes into account the differentiation of L1 and L2 conceptual metaphors to investigate idiom processing, which has not been done with the dual idiom representation model. The purpose of adding this
element to the dual idiom representation model model is to make it more useful for investigating how L2 learners process idioms.

2.7 Conceptual metaphors

According to the dual idiom representation model, conceptual metaphor is an important resource for people to infer the figurative meaning of an idiom. It exists at the conceptual level of the mental lexicon, in addition to world knowledge.

Gibbs et al. (1990) found that conceptual metaphors were important to idiom comprehension. They asked participants to describe idioms in terms of causation, intentionality, manner, consequences, negative consequences, and reversibility, and found that their imagery for idioms (e.g., “spill the beans”) was more consistent than for literal paraphrases of the idioms (e.g., “reveal a secret”) and literal controls (e.g., “spill the peas”). The consistency of imagery helped the researchers identify the conceptual metaphors of idioms. Unique and consistent conceptual metaphors contribute to the conventionalization of figurative meanings, which supports the cognitive semantics view that the meanings of idioms are not arbitrary. According to this view (Kövecses & Szabó, 1996), the mechanism of idiom processing is more than an arbitrary link between the linguistic features of an idiom (e.g., linguistic forms, syntactic properties, and meanings of the linguistic forms) and the figurative meaning of the idiom. Instead, the comprehension of idioms involves conceptual domains and cognitive mechanisms, as illustrated in Figure 4.
According to this understanding of idiom comprehension, as Figure 4 shows, a processor goes through two stages between seeing the linguistic form and disambiguating the meaning of the idiom; one stage is related to conceptual domains and the other to cognitive mechanisms. The conceptual domains of figurative language include the source domain and the target domain. The source domain refers to physiological effects of abstract emotions and feelings. The target domain refers to the abstract emotions and feelings. Take “anger is fire,” for example; the source domain is “fire” and the target domain is “anger.” In other words, people use the physiological effect of “fire” to describe the abstract emotion of “anger.” According to Gibbs and Nayak (1989), the more semantically decomposable an idiom is, the more easily one can find the source domain and the target domain of the idiom. The

\[ \text{Linguistic forms and their meanings} \]

\[ \text{Conceptual domains (i.e., source domain and target domain)} \]

\[ \text{Cognitive mechanisms (i.e., metaphors and world)} \]

\[ \text{Figurative meaning of an idiom} \]

Figure 4. The cognitive semantics view of idiom

According to this understanding of idiom comprehension, as Figure 4 shows, a processor goes through two stages between seeing the linguistic form and disambiguating the meaning of the idiom; one stage is related to conceptual domains and the other to cognitive mechanisms. The conceptual domains of figurative language include the source domain and the target domain. The source domain refers to physiological effects of abstract emotions and feelings. The target domain refers to the abstract emotions and feelings. Take “anger is fire,” for example; the source domain is “fire” and the target domain is “anger.” In other words, people use the physiological effect of “fire” to describe the abstract emotion of “anger.” According to Gibbs and Nayak (1989), the more semantically decomposable an idiom is, the more easily one can find the source domain and the target domain of the idiom. The
second stage, of cognitive mechanisms, refers to metaphors and world knowledge. In the case of bilinguals, cognitive mechanisms can be influenced by both L1 and L2 because, according to the dual idiom representation model, metaphors and world knowledge are stored at the conceptual level in the mental lexicon (Abel, 2003), and according to the modified hierarchical model, the concepts in the bilingual’s mental lexicon are affected by degrees of equivalence between L1 and L2 metaphors (Pavlenko, 2009).

As described above, Gibbs et al. (1990) used an off-line method, asking their participants to describe their imagery of idioms and their literal paraphrases. Gibbs et al. (1997) adopted an on-line method using priming to examine whether NSs activated conceptual metaphors when processing idioms. They found that the participants indeed activated conceptual metaphors when processing idioms, but not when processing literal paraphrases of the idioms. The difference between using idioms and novel expressions is that although people may sometimes create novel expressions by using metaphors, the mappings between the source domain and the target domain are not conventionalized. Understanding such novel expressions may require specific contexts or the speakers’ (or writers’) further explanations (Gibbs, 1996). On the contrary, understanding conventionalized expressions, such as idioms, usually does not require specific contexts or explanations. This is because the meanings of the conventionalized expressions have already been lexicalized in the mental lexicon (Swinney & Cutler, 1979). In summary, both the off-line and on-line studies show that conceptual metaphors are unique to idioms and affect idiom comprehension.

L1 influence on L2 occurs not only in comprehending lexical items, but also in comprehending conceptual metaphors. Boers (2000b) asked 118 French-speaking English learners to categorize the metaphors of the figurative expressions in a text,
and found that some learners categorized the metaphors incorrectly due to L1 interference. For example, when the participants were asked what metaphor was used for the figurative expression “a flourishing company,” they tended to categorize it as a “health” metaphor because they were influenced by an L1 French conventionalized figurative expression *une santé florissante*, meaning “a flourishing health,” even though in English this phrase is related to the metaphorical theme of “plants and gardening.” By extension, crosslinguistic differences in the conceptual component play a role in L2 learners’ idiom processing. As Pavlenko (2008, 2009) maintained, the degree of conceptual equivalence between L1 words and L2 words is related to the effort one has to expend to understand the L2 words. Following Pavlenko, the present study assumes that the more similar the L1 and L2 metaphors are, the less effort the L2 learners will have to expend to understand the L2 idioms. Therefore, the participants’ processing time is expected to be faster for the idioms that have high metaphoric equivalence between L1 and L2 than for those that have low metaphoric equivalence between L1 and L2.

Few studies have examined the influence of L1 metaphoric concepts on L2 idiom processing. Some studies have discussed the matter but focused only on the similarity between L1 and L2 forms. For instance, Irujo (1986) found that Spanish-speaking English learners could identify and produce more identical idioms (i.e., the forms and meanings are identical in Spanish and English) than similar idioms (i.e., the forms are slightly different, but the meanings are identical), followed by different idioms (i.e., the forms are completely different, but the meanings are identical). In addition, the learners showed greater interference from their L1 when they produced similar idioms as opposed to different idioms. When the L1–L2 forms were identical, there was positive transfer from L1 to L2. When the L1–L2 forms were similar, learners tended
to overgeneralize their L1 knowledge of idioms to the L2 idioms without paying attention to the slight differences. However, when the L1–L2 forms were different, the learners had no way to apply their L1 knowledge of idioms to the L2 idioms.

According to the modified hierarchical model, each word has a form representation and a corresponding conceptual representation. On this view, Irujo’s (1986) results can be explained not only by the degree of form equivalence, but also by the degree of conceptual equivalence, which is what Abel (2003) called conceptual metaphor. If we conceive an idiom as one big word, the bilingual mental lexicon of idioms is composed of a shared conceptual component and two separate lexical components. The shared conceptual component can be further divided into three subcomponents, namely L1-specific metaphor categories, L2-specific metaphor categories, and shared metaphor categories. The two separate lexical components store L1 idioms and L2 idioms, respectively. In Irujo’s study, it is possible that the strong positive L1 transfer of identical idioms was due to the L2 idioms and their L1 equivalents sharing identical conceptual metaphors, while the interference effects on both similar and different idioms were caused by partial equivalence or non-equivalence in conceptual metaphors.

2.7.1 Crosslinguistic influence on conceptual metaphors

Kövecses (2003), who investigated crosscultural metaphors, held that the use of metaphors is not merely a matter of cognition, but also a matter of language and culture. When two different languages express the same figurative meaning (i.e., target domain), the literal meanings in the two languages (i.e., source domain) can be either the same or different. Accordingly, the conceptual metaphors, which link figurative and literal meanings, can be either the same or different across languages. When the literal meanings are the same in two languages, the conceptual metaphors
used in the two languages are the same, too. However, when the literal meanings are different in the two languages, the conceptual metaphors used in the two languages are also different.

Some degree of similarity in conceptual metaphors, literal meanings, and figurative meanings exists in many languages. For instance, when describing the concept of “pressure,” English, Chinese, Japanese, and Hungarian all can represent it with the imagery of “hot fluid in a container” (Lantolf & Thorne, 2006). Yu (1995) compared figurative expressions of anger and happiness in Chinese and English, and found that the two languages generally use the same physiological effects (i.e., the source domain, in the cognitive semantics view) to express anger and happiness (i.e., the target domain, in the cognitive semantics view). For instance, English and Chinese both use “fire” to express anger, and use “up,” “light,” and “container” to express happiness. However, in some cases, the two languages have slight differences in choosing source domains to represent the same target domain.

English and Chinese both have the conceptual metaphor of “anger is fire.” In English, we have, “He was breathing fire,” and in Chinese we have, “Don’t set me on fire” (別讓我發火). The source domain (i.e., fire) and the target domain (i.e., anger) are the same in English and Chinese. In addition to “fire,” English uses “fluid” to express anger, while Chinese uses “gas.” For example, in English, we have, “I have reached the boiling point,” and in Chinese, we have, “He’s got big gas in his spleen” (他脾氣很大). In other words, while the target domain is anger in both Chinese and English, the source domains differ.

In other cases, the two languages draw on the same source domain to represent different target domains. For instance, “up” and “light (in weight)” can be used to
express happiness. When the conceptual metaphor is “happiness is up,” in English, we have, “They were in high spirits,” and “She felt light-hearted” and in Chinese, we have, “He is very high-spirited” (他很高興), and “She felt light and loose in her heart” (她心裡覺得輕鬆多了). The source domains (i.e., up/light) and the target domain (i.e., happiness) are the same in English and Chinese. But in addition, Chinese also uses “up/light” to express over-complacency, which has a negative meaning; for example, “He is light and floating” (他很輕浮). (It should be noted that although these examples use two different Chinese words for up/light, 高 ‘up, high’ and 輕 ‘up, light [in weight]’, the meanings overlap to a certain extent.) However, the metaphor of “over-complacency is up/light” is not a conventionalized use in English. In other words, although the source domain is the same (i.e., up and light), the target domains differ.

The examples above show that the use of the target domain and the source domain can be different to a certain extent across languages, which affects the use of conceptual metaphor by bilinguals. To be specific, the degree of metaphoric equivalence is caused by the degree of similarity in the languages’ use of the target domain and the source domain. In the current study, for the Chinese-speaking participants, the target domain and the source domain of the English idioms may not be exactly the same as they are for the equivalent idioms in Chinese. The more different the target domain and the source domain are in the two languages, the less metaphoric equivalence the idiom has.

2.7.2 The modified hierarchical model and crosslinguistic conceptual metaphors

The reason that the present study chooses the modified hierarchical model to illustrate idiom representation in the mental lexicon is because the differentiation of
L1 and L2 concepts is supported by studies of crosslinguistic conceptual metaphor. Lantolf and Thorne (2006) and Yu (1995) demonstrated that many conceptual metaphors are universal, such as “anger is fire” and “happiness is up,” and that people frequently use their common daily experiences, or world knowledge, to express similar emotions and thoughts. However, the target domains or the source domains can vary across languages. These variations can account for the differentiation between L1 and L2 metaphoric concepts. The variations support the modified hierarchical model’s understanding of the bilingual mental lexicon; that is, that it contains L1-specific categories, L2-specific categories, and shared categories at the conceptual level. An illustration of a Chinese–English bilingual conceptual representation of “up” and “light,” according to the modified hierarchical model, is presented in Figure 5.

![Diagram](image)

**Figure 5.** Crosslinguistic difference in metaphoric referents
Figure 5 shows the relationship between the lexical level and the conceptual level described by the modified hierarchical model (Pavlenko, 2009). The solid line means that the connection is strong, and the dotted line means that the connection is weak. In this example, the source domain is “up” in both Chinese (i.e., ①) and English (i.e., ②), and the shared target domain is “happiness” (i.e., ④). However, in Chinese, “up” can also refer to the concept of “over-complacency” (i.e., ③), which is L1-specific to Chinese-speaking English learners. Although “over-complacency” is an L1-specific concept in this case, there is still a connection between the L1-specific concept and the L2 word. This is because although “over-complacency is up and light” is not a conventionalized use in English, it could be a novel metaphor for English-speaking language users. The novel use has a weak connection between the concept and the L2 word. In this case, there is no L2-specific concept of “up” and “light” in English for the learners (i.e., ⑤).

2.7.3 Crosslinguistic conceptual metaphors and L2 idiom comprehension

Conceptual metaphor is crucial to L2 learners when they comprehend idioms, and the studies mentioned in the last section indicate that languages have an influence on the speakers’ use of conceptual metaphors. Crosslinguistic factors affect L2 learners’ interpretations of conceptual metaphors. Littlemore (2003) asked 18 Bangladeshi students who studied in an English university to interpret the metaphors used by their lecturers and complete a cultural values questionnaire. The researcher found crosslinguistic differences in interpretations of metaphors. The L2 learners’ interpretations of the lecturers’ meanings diverged from those of the lecturers, and both interpretations were found to be related to their cultural values according to the questionnaire. A possible explanation is that L1 schemata (i.e., the representations of the world in our minds) are entrenched in L2 learners’ minds, so that L2 learners tend
to rely on their L1 concepts to understand L2 metaphoric expressions. To truly understand L2 metaphoric expressions, they have to make additional effort if the L2 metaphors have no L1 equivalents (Cook, 1994).

Furthermore, crosslinguistic differences can cause comprehension difficulty for L2 learners. For example, Littlemore et al. (2011) asked 20 L2 learners who studied in an English university to identify words or multiwords that they found difficult to understand in lectures, and then examined the L2 learners’ interpretations of the metaphors used by their lecturers. The researchers found that nearly 40% of the expressions identified as difficult to understand were metaphoric expressions, even if they were composed of familiar words. This finding implies that L2 learners cannot just rely on the literal meanings of constituent entries (i.e., familiar words) in the hopes of figuring out metaphoric meanings. Having the appropriate conceptual metaphors is important in mapping the constituents onto the corresponding metaphoric referents.

2.8 Chapter summary

This chapter reviews four idiom processing models, namely the idiom list hypothesis, lexical representation hypothesis, configuration model, and dual idiom representation model. The dual idiom representation model highlights the importance of familiarity, semantic decomposability, and conceptual metaphor in idiom processing. The dual idiom representation model argues that idioms can be stored as idiom entries and/or constituent entries in the mental lexicon. The idiom entries and constituent entries are stored at the lexical level. If a person has only constituent entries for an idiom, s/he may rely on metaphoric referents at the conceptual level to find the figurative meaning. For NNSs, access to conceptual metaphors is affected by L1–L2 metaphoric equivalence. The present study uses the modified hierarchical
model, a model of the bilingual mental lexicon, to illustrate the relationship between the lexical level and the conceptual level, as well as L1–L2 metaphoric equivalence at the conceptual level in bilinguals’ mental lexicon. According to the modified hierarchical model, the link between an L1 word and its lexical concept is stronger than the link between an L2 word and its lexical concept. This explains why L2 learners prefer resorting to the L1 to understand the meanings of L2 words, especially when the links between L2 words and lexical concepts are weak.

This chapter also reviews three models of contextual influence on idiom processing, namely the direct access view, the reordered access model, and the graded salience hypothesis (GSH). The graded salience hypothesis highlights the importance of familiarity and context. When an idiom is familiar, the salient meaning is its figurative meaning. In contrast, when an idiom is unfamiliar, the salient meaning is its literal meaning. According to the graded salience hypothesis, the salient meaning will be activated first, regardless of contextual compatibility. The influence of context will become stronger as more contextual information is obtained.

In addition, this chapter reviews research that addresses the influence of language exposure on idiom processing, showing that NSs and NNSs process idioms differently. The importance of language exposure is also highlighted by the dual idiom representation model. However, because few studies have investigated the influence of familiarity and semantic decomposability on reaction time in priming and neutral contexts across language groups, it is worth conducting research that compares NS and NNS processing.

The final section of this chapter presents the current study’s research questions along with the predictions for each question, which are based on the reviewed literature.
2.9 The study: NS and NNS idiom processing

Before I discuss the research questions and predictions, I will briefly introduce the main task that participants completed in the current study, which is a sentence-by-sentence self-paced reading task. More details on the experiment’s design and procedure will be given in Chapter 3: Methods. This brief introduction is intended to provide readers with the basic knowledge needed to understand the research questions and predictions.

2.9.1 Sentence-by-sentence self-paced reading

The present study adopted a sentence-by-sentence self-paced reading (SPR) task to collect the data for its analysis. The data are the NS and NNS participants’ reaction times (RTs) as they processed sentences that contained idioms. The length of RTs is an indicator of participants’ idiom processing. An increase in RTs implies comprehension difficulty; a decrease in RTs implies comprehension ease (Katz & Ferreti, 2001).

Schmitt and Underwood (2004) used word-by-word self-paced reading to compare L1 speakers’ and L2 learners’ processing time of the terminal word of a formulaic expression and of the same terminal word in a nonformulaic expression. An example of a formulaic expression in their study is put all your eggs in one basket; the corresponding nonformulaic expression is drop my flower basket. Schmitt and Underwood found that both L1 speakers and L2 learners processed the terminal words in the formulaic sequences and nonformulaic sequences equally fast. This finding is counter to the argument that the terminal word in the formulaic sequence should be processed faster because the predictability of idioms facilitates processing. A presumption behind the focus on the terminal word is that it was expected to follow the point of idiom uniqueness, so its processing should be facilitated by the
participants’ recognition of the formulaic sequence. However, identification of the point of idiom uniqueness is hard to operationalize: The point varies from one idiom to another, and between L1 speakers and L2 learners. Schmitt and Underwood did not test the point of idiom uniqueness in each idiom or for either of the two language groups. Therefore, we do not know whether the lack of difference in the processing times was related to the point of idiom uniqueness.

In the present study, the sentence-by-sentence self-paced reading task focuses on participants’ RTs for whole idiom-containing sentences, rather than focusing on the processing of the terminal word (Schmitt & Underwood, 2004) or the RT difference between individual words prior to and after the point of idiom uniqueness (Siyanova et al., 2011). In other words, this study assumes that sentence-by-sentence self-paced reading is a better measurement than one that requires identifying the point of idiom uniqueness first.

2.9.2 Research questions and predictions

The purpose of the present study is to investigate how NSs and NNSs comprehend the figurative meanings of idioms. Therefore, the two types of context, priming and neutral, used in this study had to provide information for the participants to interpret the idioms figuratively. In both types of context, the figurative meaning was the only contextually appropriate meaning. The priming context was unambiguous in that participants were given contextual clues biased toward the figurative meaning before encountering the idiom. Therefore, the priming context should facilitate participants’ activation of the figurative meaning. The neutral context was ambiguous in that participants were given no contextual clues before encountering the idiom. Therefore, in the neutral context, although figurative meanings and literal meanings could both be activated at the beginning of processing
the word string, the literal meaning should be suppressed and the figurative meaning ultimately accepted. For this reason, processing idioms in the neutral context was expected to take longer than processing idioms in the priming context. An RT difference between the priming context and the neutral context (RT\text{prime} - RT\text{neu}) of less than zero signifies facilitation of processing, while a difference of zero or greater than zero indicates no facilitation of processing.

The following section presents my five research questions and predictions for the study’s outcomes.

RQ1: Is there a main effect for context type on idiom processing?

Prediction 1: There is a main effect for context type.

According to the graded salience hypothesis, both the contextually appropriate (figurative) and inappropriate (literal) meanings are activated at the beginning of sentence processing. As time goes by, the contextually appropriate meaning is maintained, while the contextually inappropriate meaning is abandoned. The priming contexts in this study can provide extra information to the participants, while the neutral contexts cannot. The extra information is predicted to help the participants to select the contextually appropriate, figurative meanings as soon as possible. However, in the neutral contexts, the participants are not able to select the contextually appropriate meanings until they collect sufficient contextual clues while reading toward the end of the sentences. These different kinds of processing should cause the participants to take more time in the neutral contexts than in the priming contexts.

RQ2: Is there an interaction effect between language group and context type?

Prediction 2: There are two possibilities. One is that no interaction effect is found between language group and context type. The other is that an interaction effect is found.
We know that idioms should be processed faster in the priming contexts than in the neutral contexts for both language groups. Therefore, the RTs in the two context types across the two language groups are not expected to pattern like the hypothetical data shown in Figure 6, which shows the Chinese group processing idioms more slowly in the priming context than the neutral context.

![Figure 6. Hypothetical graph of an interaction effect between language group and context type](image)

According to Siyanova et al. (2011), the NNSs in their study were not as good as the NSs in making use of context to understand idioms, while Conklin and Schmitt (2008) argued that the NNSs in their study were as good as NSs in making use of context. If Siyanova et al. (2011) were right, and NSs are better at using context, then the RTs in the two context types across the two language groups should pattern like those in the graph of hypothetical data in Figure 7. The graph shows an RT difference between the priming context and the neutral context that is larger for the English group than for the Chinese group. In this case, an interaction effect is found.
However, if Conklin and Schmitt (2008) were right, the RTs in the two context types across the two language groups should pattern like those shown in the hypothetical graph in Figure 8. The slope representing the NS data parallels the slope representing the NNS data, implying that the NNSs are as good as the NSs in making use of context to understand idioms. In this case, no interaction effect is found.
RQ3: Do native speakers of Chinese and native speakers of English differ in their conception of English idioms as semantically decomposable or semantically nondecomposable?

Prediction 3: Native speakers of Chinese tend to perceive English idioms as semantically decomposable rather than semantically nondecomposable, unlike native speakers of English.

In this study, I used the participants’ rating of the semantic decomposability of idioms to categorize the idioms as either decomposable or nondecomposable. According to previous studies (Abel, 2003; Gibbs & Nayak, 1989; Titone & Connine, 1994), NSs have a tendency to consider idioms as semantically nondecomposable, while NNSs have a tendency to see idioms as semantically decomposable.

RQ4: Which factors significantly influence the English idiom processing of native speakers of English and native speakers of Chinese: familiarity, semantic decomposability, and metaphoric equivalence?
Prediction 4: Familiarity can explain NS idiom processing. However, semantic decomposability and/or metaphoric equivalence should play an important role in NNS idiom processing.

According to the dual idiom representation model, familiarity is crucial to idiom comprehension because if people are familiar with an idiom, the salient meaning is the figurative meaning; however, if they are unfamiliar with an idiom, the salient meaning will be the literal meaning. Native speakers of English should be familiar with English idioms, which distinguishes them from nonnative speakers, because the salient meanings of idioms for the native speakers are the figurative meanings; therefore, they should be able to process idioms faster than non-native speakers. In addition, because NSs are familiar with the idioms, they do not have to rely on semantic decomposability to figure out the figurative meanings of idioms, according to the dual idiom representation model. Therefore, semantic decomposability might not be a significant predictor in NS idiom processing.

Native speakers of Chinese will be less familiar with the English idioms; therefore, when they rate their familiarity with the idioms in this study, it is very likely that they will rate most idioms as unfamiliar. Familiarity does not play as important a role for NNSs as it does for NSs. Therefore, familiarity is not a significant predictor for NNSs’ idiom processing behavior. Because they are unfamiliar with the idioms, the native speakers of Chinese will have to rely on the semantic decomposability of the idioms in order to link the literal meanings of the idiom constituents to their metaphoric referents, and then figure out the figurative meanings by conceptual mapping. Therefore, it is likely that both semantic decomposability and metaphoric equivalence between L1 and L2 will be significant predictors in NNS idiom processing, or metaphoric equivalence between L1 and L2 will play a more
important role than semantic decomposability in idiom processing because of a strong L1 influence (Irujo, 1986; Pavlenko, 2009).

RQ5: Is there any interaction effect among familiarity, semantic decomposability, and metaphoric equivalence in the English idiom processing of native speakers of English and native speakers of Chinese?

Prediction 5: There is an interaction effect between familiarity and semantic decomposability in the English idiom processing for native speakers of English. Moreover, there might be an interaction effect between semantic decomposability and metaphoric equivalence for native speakers of Chinese.

For NSs, because familiarity plays an important role in idiom processing while semantic decomposability does not, the interaction between the two factors could be crucial to their idiom processing. According to the dual idiom representation model, a high-familiarity, decomposable idiom can develop an idiom entry more easily than a low-familiarity, decomposable idiom. The existence of an idiom entry can facilitate the retrieval of the figurative meaning of an idiom. In other words, the more familiar and decomposable an idiom is, the less processing time it needs. The interaction between familiarity and semantic decomposability could yield two possibilities in the NS participants’ reaction times in priming and neutral contexts (RT_{prime} - RT_{neu}). The first possibility is that both RT_{prime} and RT_{neu} decrease, but RT_{prime} decreases more dramatically than RT_{neu} as degree of familiarity and semantic decomposability both increase. As a result, the RT difference between the two context types would increase. The other possibility is that both RT_{prime} and RT_{neu} decrease, but RT_{neu} decreases more dramatically than RT_{prime} as degree of familiarity and semantic decomposability both increase. As a result, the RT difference between the two context types would decrease.
For NNSs, as semantic decomposability increases, an idiom becomes more semantically transparent, which should help idiom processing. In addition, according to the modified hierarchical model, a strong L1 and L2 link can facilitate learners’ ability to link L2 words to L1 words, and then to link the L1 words to the corresponding concepts with little difficulty. Therefore, as semantic decomposability and metaphoric equivalence increase, the processing time for idioms decreases accordingly. The interaction between semantic decomposability and metaphoric equivalence could yield two possibilities in the NNS participants’ reaction times in priming and neutral contexts. The first possibility is that both \( RT_{\text{prime}} \) and \( RT_{\text{neu}} \) decrease, but \( RT_{\text{prime}} \) decreases more dramatically than \( RT_{\text{neu}} \) as degree of semantic decomposability and metaphoric equivalence both increase. As a result, the RT difference between the two context types would increase. The other possibility is that both \( RT_{\text{prime}} \) and \( RT_{\text{neu}} \) decrease, but \( RT_{\text{neu}} \) decreases more dramatically than \( RT_{\text{prime}} \) as degree of familiarity and semantic decomposability both increase. As a result, the RT difference between the two context types decreases.
CHAPTER 3

METHODS

The study aims to investigate whether the three factors of familiarity, semantic decomposability, and metaphoric equivalence affect L1 and L2 language users’ comprehension of idioms, and whether there are any interaction effects among these factors. In addition, the study investigates how idioms are represented in L1 and L2 language users’ mental lexicons in terms of semantic decomposability. Furthermore, the study investigates whether familiarity and context significantly affect L1 and L2 language users’ idiom comprehension.

The participants are native speakers of English and native speakers of Chinese. In the study’s main experiment, the participants read passages containing idioms in English. As they read, their reaction times were recorded. In addition, the English native speaker participants completed three surveys, and the Chinese native speaker participants completed four surveys. All of the participants completed a language background survey, an idiom familiarity survey, and a semantic decomposability survey. The L1 Chinese participants also completed a Chinese–English metaphoric equivalence survey. All participants also took a language proficiency test. The remainder of this chapter will describe the research instruments, the research participants, the development of the instruments, the method of idiom contextualization, the experimental procedures, and the data collection.
3.1 Research instruments

The experiment used a self-paced reading (SPR) task, which is a method first introduced by Just, Carpenter, and Woolley (1982). In an SPR, readers encounter text on a computer screen; they control the length of time that they see a given section of text by pressing a button to change the section that appears on the screen. The computer records the readers’ reaction times (RTs), which is the time between button presses. The reaction times are measured and analyzed to draw inferences about the cognitive processing of language. Longer RTs imply more processing difficulty, while shorter RTs imply less processing difficulty (Jegerski, 2014). The instruments used in idiom processing studies mainly include SPR (e.g., Conklin & Schmitt, 2008; Matlock & Heredia, 2002; Schmitt & Underwood, 2004) and eye tracking (e.g., Cieślicka & Heredia, 2013; Siyanova et al., 2011; Underwood et al., 2004). Witzel, Witzel, and Forster (2012) compared how well these two methods predict RT patterns for ambiguous sentences, such as sentences containing relative clause attachments or adverb attachments and coordination sentences. The results indicated that both SPR and eye tracking could reveal the predicted patterns of RT differences for relative clause attachment and adverb attachment sentences. In addition, both methods could accurately predict the RT patterns at the critical region for adverb attachment sentences. Furthermore, a cost-free SPR package is available through Linger. Therefore, SPR was utilized in the present study to investigate participants’ idiom processing.

For this study’s main experiment, participants were seated in front of a computer screen, on which they read passages of several sentences. They were required to press a computer key to control their exposure duration to each sentence while they were reading. One of the sentences in each passage was the critical sentence because it
contained the target idiom. The study assumed that the latencies of pressing the button would be affected by the properties of the idioms being read (e.g., familiarity, semantic decomposability, and metaphoric equivalence), because these properties should affect the time readers need to integrate an idiom’s meaning into the context in which the idiom occurs.

The task was administered on a Windows 8 ASUS laptop with Linger installed. Linger is a Windows-based program for SPR. The program randomizes the passages, so each participant read the priming contexts, neutral contexts, and fillers in different orders. In addition, the participants took the three or four surveys and the language proficiency test through SurveyGizmo, which is an online survey platform. The results were saved in Excel files.

There are different types of SPR, depending on whether the previous text is still displayed on the screen when the subsequent text is shown, and whether the text appears in the center of the screen or from left to right on the screen. The present study used a moving-window SPR task, in which only one sentence was visible at a time. The first screen the participants saw displayed a series of dashes corresponding to each character in each word of the entire passage. When the participants pressed the spacebar on the keyboard, the first sentence replaced the corresponding dashes. When the participants pressed the spacebar the second time, the first sentence was replaced by dashes, and the second sentence appeared. In other words, each button press caused the current screen content to disappear and new content to appear. The reason to use a sentence-by-sentence SPR instead of a word-by-word SPR, is that, as described in Chapter 2, each idiom might have a different point of idiom uniqueness, and the points of idiom uniqueness might differ for NSs and NNSs, as well. In other words, it is difficult to choose a certain point in an idiom in order to compare the RTs
before and after that point. Figure 9 below illustrates how the participants performed the SPR on the computer.
(First, the participants saw dashes on the screen.)

**John found out his wife had been dating another man.**

(Second, the participants saw the first sentence on the screen.)

(Third, the participants saw the second sentence on the screen. The first sentence had disappeared.)

**She tried to cover her tracks.**

(Fourth, the participants saw the third sentence, or the critical sentence. The first and the second sentence had disappeared.)

**John decided to divorce her.**

(Fifth, the participants saw the fourth sentence. The previous three sentences had disappeared.)

**Did John's wife try to hide her love affair?**

"F" for yes. "J" for no.

(Last, all sentences had disappeared. The participants saw a comprehension question and instructions for answering it.)

*Figure 9.* Screenshots from the self-paced reading task
Participants read 48 passages, half with idioms (i.e., critical items), and half without idioms (i.e., fillers). After the participants read each passage, they had to answer a comprehension question. If the answer was YES, they were to press J, and if the answer was NO, they were to press F.

After the SPR, the participants completed the surveys. All participants took the language background survey and the language proficiency test. All participants also took a survey that measured how familiar they were with the idioms and a survey that measured the extent to which they conceived the idioms to be semantically decomposable. Only the Chinese native speakers took the fourth survey, which measured the extent to which they conceived each of the English idioms to be metaphorically equivalent to a given Chinese idiom. The development of the rating surveys is described in section 3.3.2, and the surveys are provided in Appendix C.

In the analysis, the participants’ reaction times in the SPR were the dependent variable, and the participants’ ratings of familiarity, semantic decomposability, and metaphoric equivalence from the surveys were the independent variables. The significance level was set at .05.

3.2 Participants

3.2.1 Establishing sample sizes

A power analysis was conducted to determine the appropriate sample size by using the program G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). According to Faul et al. (2007), Cohen defined small, medium, and large effect sizes in an ANOVA as .10, .25, and .40, respectively. In my review of previous idiom comprehension research, I found none that reported effect sizes. A general understanding is to choose a medium effect size so that any significant difference can be detected even if it is not large. In the present study, the effect size was set at .25, α was set at .05, power was set at .75. The statistical test of “repeated-measures ANOVA, between factors” was chosen. Number of groups was set at 2, and number of measurements was set at 4 (2 levels of language group X 2 levels of context type in RQ1). G*Power does not have an option specifically for mixed-effect modeling, so the present study used RQ1 as an example to estimate the sample size. RQ1 can be analyzed either by repeated-measures ANOVAs or mixed-effect modeling, and G*Power has an option to estimate sample sizes for repeated-measures ANOVAs. Therefore, the present study chose
repeated-measures ANOVAs when estimating the sample size. The suggested sample size from the G*Power analysis was 72. The present study has a sample of 76.

3.2.2 Characteristics of the participants

Thirty-six native speakers of Chinese and 40 native speakers of English took part in the study. All the participants were either undergraduate or graduate students at the University of Hawai‘i at Mānoa (UHM). The participants were recruited through either the Linguistics Beyond the Classroom (LBC) program or an advertisement distributed through flyers and email. Participants received modest compensation for their time. Students who were recruited through the LBC program were also given extra course credits based on the LBC policy.

The native-Chinese-speaking participants were born in either Mainland China or Taiwan. Of them, 26 were female and 10 were male. Their average age was 28 years old. All of them indicated that Mandarin Chinese was their dominant language, which they had learned before age 6, and that their education had been in Mandarin Chinese. On average, they had studied Mandarin Chinese as a school subject for 13 years, and English as a foreign language in school for 11 years. As for their education level, two people indicated that their highest degree achieved was senior high school, 13 had BAs, 17 had MAs, and four had PhDs.

Among the native-English-speaking participants, one was born in the Philippines, one in India, and the others in the United States. Twenty-five were female and 15 were male. Their average age was 25.3 years old. All of them indicated that English was their first language, which they had learned before age 6. They also received education in English in school. On average, they had studied English as a subject for 11 years. As for foreign languages studied, four had studied French, two German, one Hindi, seven Japanese, two Korean, two Chinese (one for three years, and one for one year), and 10 Spanish. As for education level, 15 people indicated that their highest degree achieved was senior high school, 18 had BAs, six had MAs, and one had a PhD.

3.2.3 Language proficiency test

All the participants in both language groups took an English language proficiency test, which was a modified C-test (Raatz & Klein-Braley, 2002) (Appendix B), after they finished the SPR task in Session 1. The test consists of a passage with 20 blanks. The participants had to type in one correct word in each blank in order to get one point. If they did not fill in a blank, misspelled a word, or chose an incorrect tense, they were given zero points for that
blank. The total possible score was 20 points. The test was intended to provide information
on the participants’ general language proficiency (Raatz & Klein-Braley, 2002). The purpose
of a C-test is to measure test takers’ abilities to restore missing text by making use of the
general redundancy of the language as a whole. To restore missing text, test takers have to
rely on lexical, morphological, and syntactic knowledge on the sentence level, and
knowledge of cohesion and rhetorical organization on the text level.
3.3 Materials
3.3.1 Process of idiom selection

Idioms were selected from the 320 idioms used by Abel (2003). I eliminated 170 idioms
based on word frequency and currency. The word frequency of each idiom was examined
using the Compleat Lexical Tutor (http://www.lexutor.ca/vp/). Only idioms within the first
2000 high frequency words were retained. The purpose of this process was to ascertain that
the idioms used in this study did not contain unfamiliar words that might affect participants’
processing time. The currency of each idiom was established through use of two online
dictionaries, Cambridge Dictionaries Online (http://dictionary.cambridge.org/us/) and the
Free Dictionary (http://idioms.thefreedictionary.com). Idioms that did not occur in either of
the dictionaries were excluded, as were idioms that either of the dictionaries specified as
Australian or British. The remaining 150 idioms were randomly split into two lists for rating
purposes.

To determine the selected idioms’ degree of familiarity, semantic decomposability, and
metaphoric equivalence between L1 and L2, a preliminary online survey was conducted with
60 native speakers of Chinese (35 female; 25 male), who did not take part in the main study.
They were either undergraduate or graduate students in Taiwan. They spoke Mandarin
Chinese as their first language. Their average age was 22.6, and they had studied Chinese for
12.5 years on average and English for 10 years on average. They also took the language
proficiency test, with an average result of 13 out of 20 items correct.

This preliminary rating study involved two phases. The purpose of splitting the rating
study into two steps was to increase the response rate by shortening the time needed to
complete each phase. On average, the first phase took 40 minutes and the second phase took
15 minutes. Section 3.3.2 provides details of the rating study and its rubric. In the first phase,
each of the 60 participants was randomly assigned to one of the two idiom lists. They rated
familiarity and semantic decomposability of each idiom on a 5-point Likert scale. If they
gave an idiom 4 or 5 points, they had to write the definition of the idiom. The mean scores for the two characteristics were calculated for each idiom; if one of the mean scores fell between 2.5 and 3.5, the idiom was excluded. The reason to adopt a range instead of a cut-off point (e.g., 3) was to ascertain that the idioms selected had a clear inclination in familiarity and semantic decomposability. For instance, if an idiom has an average familiarity score of 3.1, it remains unclear if the idiom is significantly more familiar to most of the participants, compared to an idiom having an average familiarity score of 2.9. However, it is very likely that an idiom having an average familiarity score of 3.6 is significantly more familiar to most of the participants, compared to an idiom having an average familiarity score of 2.4. After the first phase, 54 idioms remained. In the second phase, which took place one week after the first phase, a metaphoric equivalence survey was administered. The same 60 participants were asked to rate the 54 idioms in terms of metaphoric equivalence between English and Chinese. As with the analysis of the first phase, a mean score was calculated for each idiom, and the idioms with mean scores between 2.5 and 3.5 were excluded. At this point, 24 idioms remained, and these idioms were used for the main study (Appendix A). In addition, the familiarity rating and the semantic decomposability rating together were used to examine the validity and reliability of the surveys that the participants in the main study completed.

3.3.1.1 Validity of idiom ratings

A validity check was carried out to examine whether the surveys measured what they were designed to measure. In other words, we want to know whether the familiarity rating survey measured the concept of familiarity, and the semantic decomposability rating survey measured the concept of semantic decomposability. Factor analysis was used to examine validity. It was expected that all the idiom ratings in the familiarity rating survey would belong to one factor (e.g., Factor 1), and all the idiom ratings in the semantic decomposability rating survey would belong to another factor (e.g., Factor 2). If one idiom rating in the semantic decomposability rating survey was found in the other factor, it would imply that the participants did not truly rate the idiom based on the concept of semantic decomposability. Therefore, the rating results for that idiom would be problematic.

The results of a factor analysis of a survey can show how many constructs exist in the survey. Ideally, each survey measures only a single construct. Principal component analysis was the extraction method used, and varimax was the rotation method. Principal component analysis is preferred if the goal is to explore patterns in the data, and varimax is preferred.
over other rotation methods in an exploratory analysis (Kim & Mueller, 1978). Before factor analysis, the distribution of familiarity and semantic decomposability was examined in the NS and the NNS groups’ data. The purpose was to investigate whether the two language groups had different distributions. If so, two separate factor analyses, instead of one factor analysis, are required. Four distribution charts are shown in Appendix E. The charts suggest that the NS group and the NNS group had different distributions of familiarity and semantic decomposability ratings. Therefore, two separate factor analyses should be carried out. However, in the present study the NS group had 40 participants, and the NNS group had 36 participants. Carrying out a factor analysis with such small sample sizes could run into a risk of failure of the solution to converge (Tabachnick & Fidell, 2013). Therefore, the present study adopted a single factor analysis by combining the NS and the NNS data. Further analysis is needed with increased sample sizes for which two separate factor analyses could be run.

Before factor analysis was conducted, the assumptions of normality and linearity were checked. Skewness and kurtosis were used to check normality. According to Tabachnick and Fidell (2013), an acceptable range of skewness and kurtosis is ±1.5. If variables (i.e., items) do not fall into the acceptable range, data transformation is required. Appendix F displays the skewness and kurtosis of raw data and transformed data. According to Tabachnick and Fidell (2007) and Howell (2007), if the distributions are moderately negatively skewed, a square-root transformation method is used (i.e., new X = sqrt (K-X); K = a constant which makes the smallest score 1). If the distributions are substantially negatively skewed, a logarithmic transformation is used (i.e., new X = LG10 (K-X)). As shown in the normality check table, when the raw data were examined, 33 skewness or kurtosis values were out of the normal range. When the square-root method was used, 49 skewness or kurtosis values were out of the normal range. When log10 transformation was used, 56 skewness or kurtosis values were out of the normal range. Given that more values were out of the normal range after data transformation, the present study kept both the familiarity and semantic decomposability ratings data untransformed.

According to Tabachnick and Fidell (2013), when there are too many items to run a linearity check, a practical method is to select several items with the worst skewness and kurtosis values and examine their linearity by scatterplots. If no curvilinearity is found, the linearity assumption is valid. In the present study, the three items with the worst skewness
and kurtosis values were used to check linearity (i.e., Idiom 9f, Idiom 9d, and Idiom 17f). As shown in Appendix F, the scatterplots showed that no curvilinearity was found, so the linearity assumption was valid.

Table 1 shows the results of the factor analysis: Except for Idiom 3, all the idioms in the familiarity rating survey had a higher loading on Factor 1, and all of the idioms in the semantic decomposability rating survey had a higher loading on Factor 2. Idiom 3 should have had a higher loading on Factor 2, but in fact had a higher loading on Factor 1. In other words, the familiarity rating survey measured the construct of familiarity, and the semantic decomposability rating survey measured the construct of semantic decomposability, but Idiom 3 did not measure the construct of semantic decomposability well. Idiom 3 was therefore removed from the materials.

Table 1. Factor Analysis

<table>
<thead>
<tr>
<th>Idioms</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiom8f</td>
<td>.832</td>
<td>.241</td>
</tr>
<tr>
<td>Idiom16f</td>
<td>.794</td>
<td>.090</td>
</tr>
<tr>
<td>Idiom5f</td>
<td>.791</td>
<td>.231</td>
</tr>
<tr>
<td>Idiom12f</td>
<td>.780</td>
<td>.022</td>
</tr>
<tr>
<td>Idiom2f</td>
<td>.779</td>
<td>-.011</td>
</tr>
<tr>
<td>Idiom23f</td>
<td>.757</td>
<td>.094</td>
</tr>
<tr>
<td>Idiom13f</td>
<td>.745</td>
<td>.103</td>
</tr>
<tr>
<td>Idiom24f</td>
<td>.738</td>
<td>.051</td>
</tr>
<tr>
<td>Idiom11f</td>
<td>.737</td>
<td>-.009</td>
</tr>
<tr>
<td>Idiom1f</td>
<td>.682</td>
<td>.065</td>
</tr>
<tr>
<td>Idiom10f</td>
<td>.642</td>
<td>.065</td>
</tr>
<tr>
<td>Idiom7f</td>
<td>.636</td>
<td>.146</td>
</tr>
<tr>
<td>Idiom4f</td>
<td>.574</td>
<td>.049</td>
</tr>
<tr>
<td>Idiom19f</td>
<td>.566</td>
<td>-.014</td>
</tr>
<tr>
<td>Idiom6f</td>
<td>.551</td>
<td>.146</td>
</tr>
<tr>
<td>Idiom9f</td>
<td>.524</td>
<td>-.011</td>
</tr>
<tr>
<td>Idiom</td>
<td>Score 1</td>
<td>Score 2</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Idiom20f</td>
<td>.491</td>
<td>.059</td>
</tr>
<tr>
<td>Idiom15f</td>
<td>.484</td>
<td>-.097</td>
</tr>
<tr>
<td>Idiom14f</td>
<td>.470</td>
<td>.049</td>
</tr>
<tr>
<td>Idiom22f</td>
<td>.449</td>
<td>-.098</td>
</tr>
<tr>
<td>Idiom21f</td>
<td>.429</td>
<td>.053</td>
</tr>
<tr>
<td>Idiom3d</td>
<td>.378</td>
<td>.297</td>
</tr>
<tr>
<td>Idiom3f</td>
<td>.377</td>
<td>.182</td>
</tr>
<tr>
<td>Idiom18f</td>
<td>.356</td>
<td>.026</td>
</tr>
<tr>
<td>Idiom17f</td>
<td>.274</td>
<td>-.048</td>
</tr>
<tr>
<td>Idiom19d</td>
<td>-.042</td>
<td>.707</td>
</tr>
<tr>
<td>Idiom22d</td>
<td>-.419</td>
<td>.672</td>
</tr>
<tr>
<td>Idiom13d</td>
<td>.392</td>
<td>.652</td>
</tr>
<tr>
<td>Idiom8d</td>
<td>.362</td>
<td>.637</td>
</tr>
<tr>
<td>Idiom5d</td>
<td>.320</td>
<td>.631</td>
</tr>
<tr>
<td>Idiom23d</td>
<td>.416</td>
<td>.609</td>
</tr>
<tr>
<td>Idiom11d</td>
<td>.243</td>
<td>.609</td>
</tr>
<tr>
<td>Idiom17d</td>
<td>-.340</td>
<td>.603</td>
</tr>
<tr>
<td>Idiom21d</td>
<td>-.071</td>
<td>.594</td>
</tr>
<tr>
<td>Idiom10d</td>
<td>-.005</td>
<td>.573</td>
</tr>
<tr>
<td>Idiom24d</td>
<td>.338</td>
<td>.557</td>
</tr>
<tr>
<td>Idiom18d</td>
<td>-.182</td>
<td>.543</td>
</tr>
<tr>
<td>Idiom15d</td>
<td>-.316</td>
<td>.540</td>
</tr>
<tr>
<td>Idiom7d</td>
<td>.149</td>
<td>.536</td>
</tr>
<tr>
<td>Idiom6d</td>
<td>.159</td>
<td>.518</td>
</tr>
<tr>
<td>Idiom4d</td>
<td>-.373</td>
<td>.516</td>
</tr>
<tr>
<td>Idiom12d</td>
<td>.107</td>
<td>.513</td>
</tr>
<tr>
<td>Idiom9d</td>
<td>-.260</td>
<td>.500</td>
</tr>
<tr>
<td>Idiom1d</td>
<td>.105</td>
<td>.497</td>
</tr>
<tr>
<td>Idiom20d</td>
<td>-.070</td>
<td>.461</td>
</tr>
<tr>
<td>Idiom16d</td>
<td>.273</td>
<td>.370</td>
</tr>
<tr>
<td>Idiom2d</td>
<td>.157</td>
<td>.370</td>
</tr>
</tbody>
</table>
3.3.1.2 Reliability of the idiom rating surveys

The reliability of the familiarity and semantic decomposability surveys was checked using Cronbach’s alpha. The results of Cronbach’s alpha indicate the degree to which items consistently measure the same construct (i.e., in this case, familiarity and semantic decomposability). In the present study, the reliability of the two surveys was 0.92. This indicates that the rating results were highly reliable.

3.3.2 Familiarity, semantic decomposability, and metaphoric equivalence surveys

In the familiarity rating survey, the participants rated each idiom based on five options:

1. *I haven’t seen this phrase before, and I don’t know what it means.*
2. *I have seen this phrase before, but I don’t know what it means.*
3. *I haven’t seen this phrase before, but I can roughly guess what it means.*
4. *I have seen this phrase before, and I think I know what it means.*
5. *I know this phrase.*

If the participants selected the third, fourth, or fifth option, they were required to provide the meaning of the idiom to verify that they were truly familiar with it.

Figure 10 illustrates what the participants saw on the computer screen when rating familiarity, and how the participants were asked to write the definition of the idiom if they indicated that they could guess the meaning or knew the meaning.
This study’s familiarity rating survey was based on the Vocabulary Knowledge Scale (VKS) developed by Wesche and Paribakht (1996) and was modified based on Joe (1995). There are two distinct features of the familiarity check used in the present study. First, the participants were required to write down the figurative meanings of the idioms they claimed to know to some extent. Second, the VKS includes only four options; the current study added the third option, *I haven’t seen this phrase before, and I think I know what it means*, based on Joe’s (1995) argument that learners are able to guess word meanings by using their vocabulary knowledge, such as prefixes and suffixes. According to the dual idiom representation model (Abel, 2003), L2 learners are able to semantically decompose idioms in order to “guess” the meaning by linking constituent meanings to conceptual metaphors, so the inclusion of this description is compatible with the theoretical framework that the present study assumes.

In the VKS, the researchers gave more points as the participants indicated higher degrees of vocabulary knowledge. The rating scales and principles for scoring used in the present study were adapted with slight modifications from those used in the VKS. In the present study, the participants received more points when they indicated a higher degree of
familiarity (I haven’t seen this phrase before, and I don’t know what it means, 1 point; I have seen this phrase before, but I don’t know what it means, 2 points; I haven’t seen this phrase before, but I can roughly guess what it means, 3 points; I have seen this phrase before, and I think I know what it means, 4 points; and I know this phrase, 5 points). If the participants selected the third, fourth, or fifth option, but did not give an appropriate definition, points were deducted (i.e., without an appropriate definition, I haven’t seen this phrase before, but I can roughly guess what it means received 1 point; I have seen this phrase before, and I think I know what it means received 2 points; and I know this phrase also received 2 points).

In the semantic decomposability survey, each idiom appeared with its definition. Definitions were extracted from Cambridge Dictionaries Online (http://dictionary.cambridge.org/us/) and the Free Dictionary (http://idioms.thefreedictionary.com). The instructions, which were based on those used by Gibbs and Nayak (1989), asked the participants to evaluate how well they were able to understand the idiom meaning (i.e., relate it to the provided definition) by considering the individual words. The participants chose from five options: not at all, a little bit, somewhat, a lot, and very much. Figure 11 illustrates what the participants saw on the computer screen when rating semantic decomposability.
Likewise, in the metaphoric equivalence survey, the participants rated each idiom by choosing from five options: not at all, a little bit, somewhat, a lot, and very much. No previous survey on metaphoric equivalence was available to my knowledge, so I constructed a survey and tested it in a pilot study. For the details of how the equivalent Chinese idioms and metaphors were selected, please refer to section 3.5.3: Pilot study three. In the instructions, I gave an example to explain how to take the survey. Figure 12 illustrates what the participants saw on the computer screen when rating metaphoric equivalence. The English translation below each screenshot was not shown in the main experiment.
Example: *read between the lines* means to try to understand what is meant by something that is not written explicitly or openly. This English idiom uses the concept of “read / between the lines” to express its figurative meaning. To what extent do you think the concept of “music outside the strings” in Chinese is similar to the concept of “read between the lines” in English?

How to answer: In English, we use the concept of “to read between the lines of a text” (read / between the lines) to express something that is not written explicitly or openly. In the example above, the Chinese use the concept of “to play music which is outside the strings” to express something that is not written explicitly or openly. Therefore, please answer the question based on the degree of similarity between the English concept and the Chinese concept.

1. *praise something to the skies* means to give something much praise. This English idiom uses the concept of praise / to the skies to express the figurative meaning of the idiom. To what extent do you think the concept of “lift the praise up to the sky” in Chinese is similar to the concept of “praise to the skies” in English?

<table>
<thead>
<tr>
<th>「捧上天」和「贊美到天邊」這兩個概念的相關程度</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
</table>

The degree of similarity between the concept of “lift the praise up to the sky” and the concept of “praise to the skies”

**Figure 12.** A screenshot from the metaphoric equivalence rating survey

### 3.3.3 Idiom contexts

Two types of contexts were created, namely priming contexts and neutral contexts. Each idiom occurred in both types of contexts, so that I could investigate how much time the participants spent on reading the idioms in the two different contexts. The priming contexts have extra contextual clues right before the critical sentence, while the neutral contexts do not. The structure of the two contexts is illustrated in (3.1) and (3.2). A full list of idiom contexts can be found in Appendix D.
(3.1) Priming context
Structure:
opening/elaboration/priming the figurative meaning/idiom (critical sentence)/closing
Example:
*Alice baked a cake./Her boyfriend was eager to taste it./He kept telling her the cake was delicious./He praised the cake to the skies./Alice was very touched.*

(3.2) Neutral context
Structure:
opening/elaboration/idiom (critical sentence)/closing
Example:
*Alice baked a cake./Her boyfriend was eager to taste it./He praised the cake to the skies./Alice was very touched.*

In both the priming and the neutral contexts, the figurative meaning is the contextually appropriate meaning, while the literal meaning is the contextually inappropriate meaning. In the priming contexts, the extra contextual clues give the figurative meaning of the idiom. Thus, the participants should more easily comprehend the idioms in the priming contexts than in the neutral contexts. In the neutral context, which is an ambiguous context, both literal and figurative meanings may be available to the readers initially. The literal meaning is contextually inappropriate, while the figurative meaning is contextually appropriate. Therefore, participants have to suppress the literal meaning in order to activate the figurative meaning. Consequently, it should take more time to process idioms in the neutral contexts than the priming contexts. According to Jegerski (2014), a longer RT implies processing difficulty, while a shorter RT implies ease of processing. The participants were expected to find the critical sentences easier to process in the priming contexts, leading to shorter RTs, and more difficult in the neutral contexts, leading to longer RTs.

When the participants took part in the SPR in Session 1, they were randomly assigned to either List 1 or List 2. In List 1, the priming contexts occurred in odd passages, and the neutral contexts occurred in even passages. In List 2, the neutral contexts occurred in odd passages, and the priming contexts occurred in even passages. However, the computer program presented the passages to the participants in a random order, so no pattern could lead
the participants to anticipate the type of context they would encounter. If the participants were assigned to List 1 in Session 1, they were assigned to List 2 in Session 2. If they were assigned to List 2 in Session 1, they were assigned to List 1 in Session 2. There were 12 priming contexts, 12 neutral contexts, and 24 fillers in each list. The fillers were passages that did not contain idioms. The ratio of fillers to idioms was 1 to 1.

3.4 Procedures

In Session 1, the researcher began by explaining how to use the keyboard to carry out the SPR task, and the participants received four practice trials before the main experiment. If they had any questions during the practice trials, they could ask the researcher for clarification. However, they were not allowed to ask questions during the main experiment. This was to avoid any interruption during the experiment.

The participants then read the 48 passages (24 with idioms, 24 fillers) and answered the comprehension question that followed each passage. The comprehension questions tested whether the participants had paid attention to the passages while reading. After the participants finished the SPR, they were given the language background survey and the language proficiency test. Session 1 took 20 minutes.

In Session 2 (one week later), the participants again carried out the SPR task. The procedure was the same as it was for Session 1. After they finished the SPR, the native-English-speaking participants were given the familiarity and semantic decomposability surveys. The native-Chinese-speaking participants were given the same two surveys and the metaphoric equivalence survey. Session 2 took 30 minutes.

At the end of Session 2, the participants who were recruited through the LBC program were given a feedback form, a questionnaire, and compensation. The participants who were recruited through flyers and email were given compensation.

3.5 Pilot studies

Four pilot studies were conducted prior to the main study. The first three pilot studies were done to support the development of the experimental materials, as described above and in the following subsections. The fourth pilot study was conducted to test the procedures for the main experiment.

3.5.1 Pilot study one: Selection of idioms

The process of selecting the idioms was described in section 3.3.1.
3.5.2 Pilot study two: Naturalness of contexts

After the idioms were selected, the researcher created the priming contexts and the neutral contexts for the idioms, working with a proofreader to make the contexts as natural as possible. The proofreader was a native speaker of English with an MA in linguistics from UHM. After the contexts were revised, the researcher asked two native speakers of English to read the materials and give feedback on the naturalness of the contexts. Both of these consultants agreed that the contexts were sufficiently natural.

3.5.3 Pilot study three: Development of the metaphoric equivalence survey

There was no existing metaphoric equivalence survey available when the study was carried out. The researcher first utilized online resources (e.g., Dictionary of Chinese Idioms: http://dict.idioms.moe.edu.tw/cydic/index.htm ) to find equivalent Chinese idioms and the metaphors used by the idioms, and then asked two native speakers of Chinese to take the metaphoric equivalence survey and give feedback on the idioms and metaphors used in Chinese. Through discussion, these two consultants and the researcher reached agreement on the meaning of the Chinese idioms and metaphors to be used in the main experiment. In addition, the researcher asked two native speakers of English to read the English metaphors and the contexts and give feedback. Through discussion, these two consultants and the researcher reached agreement on the meaning of the English metaphors and the naturalness of the contexts to be used in the main experiment.

3.5.4 Pilot study four: Procedures of the experiment

After all the materials were created, eight native-Chinese-speaking participants and eight native-English-speaking participants took part in a pilot study to test the procedures and the materials of the experiment. The English participants indicated that they were familiar with most of the idioms. Even though they were unfamiliar with a very few idioms, they were still able to understand the meanings through the contexts. The Chinese participants indicated that they were familiar with some idioms, and when they did not know the idioms, they used contextual clues to help them figure out the figurative meanings of the idioms. In addition, all participants were able to answer at least 80% of the comprehension questions correctly in the two SPR sessions. Furthermore, all the participants indicated that the instructions for the surveys were clear to them. As for the language proficiency test, the English participants indicated that it was easy for them to produce the answers to fill in the blanks. By contrast, the Chinese participants indicated that it was a little challenging for them to produce the
answers to fill in the blanks. On average, the English participants got 16 out of 20 items correct, and the Chinese participants got 12 out of 20 items correct.

3.6 Data collection

Participants’ reaction times and responses to comprehension questions were automatically recorded by Linger. Their responses to the four surveys and the language proficiency test were saved in SurveyGizmo. The researcher saved individuals’ data and responses to the online surveys in Excel files.
CHAPTER 4
RESULTS AND DISCUSSION

The present study aims to investigate how context types affect idiom processing (RQ1 and RQ2), how English idioms are represented in English-speaking participants’ and Chinese-speaking participants’ mental lexicon (RQ3), and how participants’ conception of idioms affects their idiom processing (RQ4 and RQ5). Table 2 summarizes the methods employed to analyze the different types of data collected in the various parts of this study: participants’ answers to comprehension questions, scores on the language proficiency test, RTs from the self-paced reading task (RQ1 and RQ2), and ratings on the surveys on the familiarity, semantic decomposability, and crosslinguistic metaphoric equivalence of the idioms.
Table 2. Summary of Data Analysis

<table>
<thead>
<tr>
<th>Question</th>
<th>Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension questions</td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>Language proficiency test</td>
<td>Independent samples ( t )-test</td>
</tr>
<tr>
<td>Self-paced reading task (RQ1 and RQ2)</td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>Semantic decomposability survey (RQ3)</td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>Self-paced reading task (RQ4 and RQ5)</td>
<td>Mixed-effects modeling</td>
</tr>
</tbody>
</table>

4.1 Comprehension questions and general language proficiency test

4.1.1 Data analysis

An initial analysis was carried out to demonstrate that the English-speaking participants and the Chinese-speaking participants represented the two populations of interest: native speakers of English and English learners. In order to ensure that the participants were engaged in the SPR task, they answered a comprehension question after they read an idiom context. Their accuracy rates on the comprehension questions were examined prior to the analyses that were conducted to answer the research questions. A very low accuracy rate would suggest that the participant was not paying attention to the task; therefore, their performance would not be of interest (Jegerski, 2014). The cut-off was 80% accuracy, based on the results of the fourth pilot study (see Section 3.5.4 in Chapter 3), in which the participants answered at least 80% of the questions correctly. In the current study, all of the participants’ average accuracy rates on the comprehension questions were above 80%, so no participant’s data were excluded from the analysis. (The Chinese-speaking participants’
accuracy rate was 92% on average, and the English-speaking participants’ accuracy rate was slightly higher, at 93.5% on average.)

In addition, Cronbach’s alpha was used to examine the reliability of the general language proficiency test. An independent samples t-test was conducted to compare average accuracy rates between the two language groups in order to investigate whether general language proficiency was significantly different between the English-speaking participants and the Chinese-speaking participants.  

4.1.2 Results

For the general language proficiency test, Cronbach’s alpha was .77, which indicates that the test was moderately reliable. An independent samples t-test examined whether there was a significant difference in the accuracy rates of the general language proficiency test between the two language groups. Prior to the t-test, the assumptions of no outliers, homogeneity, and normality were tested. As shown in Figure 13, there was one outlier in the English group and one outlier in the Chinese group. According to Tabachnick and Fidell (2013), if outliers are from the intended population, they may be retained but the values should be converted in order to normalize the distribution. The two outliers in the present study were from the intended populations based on their demographic information; therefore, the outliers were replaced by the lower bounds of the whiskers of the boxplots to reduce their impact on the subsequent analyses.
After replacing the outliers with the minimum, the English-speaking participants’ average accuracy rate on the general language proficiency test was 85% with a standard deviation of 7%. A Shapiro-Wilk test showed that the English-speaking participants’ general language proficiency was not normally distributed ($W = .92, p = .01, \alpha = .05$). However, the independent samples $t$-test is robust to non-normal distributions if the sample size is large enough (e.g., $N > 30$); therefore, the English-speaking participants’ data were not transformed to a normal distribution. For the Chinese-speaking participants, the average accuracy rate on the general language proficiency test was 65% with a standard deviation of 15%. A Shapiro-Wilk test showed that the Chinese-speaking participants’ general language proficiency was normally distributed ($W = .97, p = .4, \alpha = .05$). Figure 14 below shows that the distribution of NS general language proficiency was negatively skewed, while that of NNS general language proficiency was normally distributed.
Figure 14. Distribution of accuracy rates on the language proficiency test ($N_{\text{English}} = 39$; $N_{\text{Chinese}} = 35$)

Levene’s test showed that the error variances were not equivalent in the two language groups ($F = 22.08, p < .001, \alpha = .05$). Therefore, the degrees of freedom were adjusted from 72 to 47. The independent samples $t$-test showed that the participants’ scores were significantly different between the two language groups ($t(47) = 6.93, p < .001, \alpha = .05, 95\% \text{ CI} [13.61\%, 24.76\%]$). The results of the $t$-test indicate that the English-speaking participants’ general language proficiency was significantly higher than the Chinese-speaking participants’.

Moreover, the effect size using Cohen’s $d$ is 1.71, which means the mean difference between the two language groups is 1.71 times the standard deviation. According to Cohen (1988), a $d$ of .2 indicates a small effect size; a $d$ of .5, a medium effect size, and a $d$ of .8, a large effect size. Therefore, the mean difference in general language proficiency between the English and the Chinese groups is large. The results of the independent $t$-test suggest that the Chinese-speaking participants and the English-speaking participants were not from the same population in terms of English language proficiency. The Chinese-speaking participants cannot be considered advanced language users because their general language proficiency was not native-like; hence, they were suitable participants for the study, whose intended
target population was classroom-taught foreign language learners. Table 3 summarizes the results of the general language proficiency test.

Table 3. Summary of General Language Proficiency Test Results

<table>
<thead>
<tr>
<th>Language</th>
<th>Average accuracy rate</th>
<th>SD</th>
<th>Reliability</th>
<th>Levene’s test</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>85%</td>
<td>7%</td>
<td>0.77</td>
<td>$F = 22.08$</td>
<td>6.93</td>
</tr>
<tr>
<td>Chinese</td>
<td>65%</td>
<td>15%</td>
<td>(p &lt; .001)</td>
<td>(p &lt; .001)</td>
<td></td>
</tr>
</tbody>
</table>

4.2 Influence of context and language on NS and NNS idiom processing (RQ1 & RQ2)

**RQ1**: *Is there a main effect for context type on idiom processing?*

**RQ2**: *Is there an interaction effect between language group and context type?*

4.2.1 Data analysis

All the NS and NNS RTs in the priming and the neutral contexts were used for log transformation. After the RTs were transformed, the RTs that fell outside of 3SDs were replaced by the value of 3SDs. Nine data points out of 3496 were identified as outliers and then replaced with the value of 3SDs, accounting for 0.3% of the data.

A mixed-effect modeling was applied to examine whether context and the interaction between language group and context were significant predictors. The dependent variable was the transformed RT. The independent variables were language group, context type, and the interaction between language group and context type. Language group was not an interest of the present study, but its interaction with context was. To investigate the interaction effect, language group was added to the model. Language group, context type, and the interaction term were the fixed effects. Subjects and items were the random effects.

4.2.2 Results

Figures 15 and 16 below show the distribution of RTs in each context type. In the priming contexts, the English-speaking participants’ RTs ($M = 1392.02$ ms; $SD = 772.21$ ms) were shorter than the Chinese-speaking participants’ ($M = 2829.53$ ms; $SD = 2051.68$ ms). In the neutral contexts, the English-speaking participants’ RTs ($M = 1548.31$ ms; $SD = 889.41$ ms) were shorter than the Chinese-speaking participants’ ($M = 3038.01$ ms; $SD = 1976.69$ ms). These results are reasonable because native speakers tend to process sentences faster.
than non-native speakers. In addition, the figures show that the RTs are more normally distributed after data transformation than before data transformation (Appendix G).

Figure 15. Distribution of RTs in priming contexts ($N_{\text{English}} = 40; N_{\text{Chinese}} = 36$)
None of the previous studies employing reaction time data has reported the effect size of any effect found. Therefore, no effect size information can be used to index the magnitude of effects found in this study; nevertheless, I provide the effect size as a reference for other researchers when they use the same predictors to examine the effects on reaction time in their own studies. The effect size in mixed-effects modeling can be calculated in the same way as the standardized coefficient in a multiple regression analysis (Seongah. Im, personal communication, January 13, 2014). In a multiple regression analysis, standardization of coefficients is used to measure the effects of predictors on the dependent variable. Moreover, standardized coefficients are created by standardizing all the variables in the multiple regression analysis. The formula is as follows:

\[
\frac{\text{(Unstandardized coefficient of the predictor} \times \text{its standard deviation})}{\text{Standard deviation of the dependent variable}}
\]
Standardized coefficients are interpreted as the change in the dependent variable measured in standard deviations.

This study first tried a two-predictor model, including language group and context as the fixed effects and subject and item as random intercepts and random slopes (the slope of context was allowed to vary for each subject and item). However, the results showed that the slope of context did not vary for each subject (variance = 0). Therefore, in a following model, the context as a random slope by subject was excluded. In the following model (three-predictor model), language group and context remained the fixed effects, and the interaction between language group and context was added as a new fixed effect. The random effects included random intercepts by subject and item, and the random slope of context by item.

Table 4 shows the results of a three-predictor model. There were main effects for language group ($t = -9.85; p < .001; \text{effect size} = .5$) and context ($t = -3.52; p < .001; \text{effect size} = .08$), but no interaction effect between language group and context ($t = -.37; p = .71; \text{effect size} = .01$). Although language group was a significant predictor, it was not the interest of the present study because NSs tended to process English sentences faster than NNSs regardless of whether the sentences contained idioms. A further analysis examined NSs’ and NNSs’ processing of fillers (sentences with no idioms). Language group was the fixed effect, and subjects and items were random intercepts. The results show that language group was a significant predictor ($t = -10.98; p < .001; \text{effect size} = .007$). The mean RT of the NSs was 1524.08 ms, and that of the NNSs was 2487.98 ms. The NSs processed sentences with literal interpretations faster than the NNSs did.

Figure 17 shows the relationship between language group and context. The solid line represents Chinese-speaking participants’ RTs in the priming and neutral contexts. The dotted line represents English-speaking participants’ RTs in the priming and neutral contexts. The two lines are parallel, which suggests that language group did not interact with context. The graph supports my findings reported above.
Figure 17. Language group and context
Table 4. Results of the Three-Predictor Model

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>b</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effect (γ0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>49.76</td>
<td>90.40</td>
<td>&lt;0.001</td>
<td>3.31</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>Language Group</td>
<td>79.86</td>
<td>-9.85</td>
<td>&lt;0.001</td>
<td>0.50</td>
<td>-0.36</td>
<td>-0.24</td>
</tr>
<tr>
<td>Context</td>
<td>33.72</td>
<td>-3.52</td>
<td>&lt;0.001</td>
<td>0.08</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
<tr>
<td>Interaction</td>
<td>3374.00</td>
<td>-0.37</td>
<td>0.71</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>0.02</td>
<td>5.85</td>
<td>&lt;0.001</td>
<td>0.01</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>0.02</td>
<td>3.19</td>
<td>&lt;0.001</td>
<td>0.01</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>0.001</td>
<td>2.44</td>
<td>0.01</td>
<td>0.0005</td>
<td>0.0025</td>
<td></td>
</tr>
<tr>
<td>Residual (eSi)</td>
<td>0.03</td>
<td>41.07</td>
<td>&lt;0.001</td>
<td>0.029</td>
<td>0.032</td>
<td></td>
</tr>
</tbody>
</table>

Session was another factor affecting participants’ RTs. It is very likely that participants processed sentences faster in session 2 than session 1. However, the present study counterbalanced context type across sessions, so session should not be a confounding factor affecting the interpretation of results. This study first tried a two-predictor model, including context and session as the fixed effects and subject and item as random intercepts and random slopes. However, the results showed that the slope of context did not vary for each subject (variance = 0). Therefore, in a following model, the context as a random slope by subject was excluded. In the following model, context and session remained the fixed effects, and the interaction between context and session was added as a new fixed effect. The random effects included random intercepts by subject and item, the random slope of context by item, and the random slope of session by subject and item.
Table 5 shows that context ($t = -4.05; p < .001; \text{effect size} = .09$) and session ($t = 9.09; p < .001; \text{effect size} = -.2$) were significant predictors of processing idioms, but no interaction effect was found ($t = .78; p = .44; \text{effect size} = -.01$). The result of no interaction supports my claim that session did not interact with context to affect processing time.

Table 5. Results of the Three-Predictor Model

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>$t$</th>
<th>$p$</th>
<th>$b$</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Fixed effect ($\gamma_0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>55.19</td>
<td>81.91</td>
<td>&lt;0.001</td>
<td>3.08</td>
<td>3.24</td>
</tr>
<tr>
<td>Context</td>
<td>32.30</td>
<td>-4.05</td>
<td>&lt;0.001</td>
<td>.09</td>
<td>-0.08</td>
</tr>
<tr>
<td>Session</td>
<td>73.19</td>
<td>9.09</td>
<td>&lt;0.001</td>
<td>-.2</td>
<td>0.09</td>
</tr>
<tr>
<td>Interaction</td>
<td>3276.98</td>
<td>0.78</td>
<td>0.44</td>
<td>-.01</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Variance</th>
<th>Wald Z</th>
<th>$p$</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.04</td>
<td>5.86</td>
<td>&lt;0.001</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Item</td>
<td>0.02</td>
<td>3.15</td>
<td>0.002</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item (context)</td>
<td>0.001</td>
<td>2.58</td>
<td>0.01</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>Subject (session)</td>
<td>0.002</td>
<td>4.15</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>Item (session)</td>
<td>0.001</td>
<td>2.10</td>
<td>0.04</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Residual ($\varepsilon_i$)</td>
<td>0.03</td>
<td>40.48</td>
<td>&lt;0.001</td>
<td>0.02</td>
<td>0.03</td>
</tr>
</tbody>
</table>

To sum up, there was a main effect for context type, but no interaction effect between language group and context type; the result of context as a main effect supports the study’s prediction. Moreover, the result of no interaction effect supports Conklin and Schmitt’s (2008) claims, in that the NNSs in the current study were as good as the NSs in making use of contexts to understand idioms.
4.2.3 Discussion

The results show that the participants spent significantly less reading time in the priming contexts than in the neutral contexts. The priming contexts provided more contextual clues than the neutral contexts, which may have primed the participants before they encountered idioms during reading. The implication is that the extra contextual clues helped the participants to select contextually appropriate meanings (i.e., figurative meanings) and abandon contextually inappropriate meanings (i.e., literal meanings) as soon as possible although the two meanings were accessible in the initial processing. In this sense, the processing in the priming contexts looked similar to the direct access view and the reordered access model in that the contextually inappropriate meanings were abandoned early. On the other hand, the neutral contexts did not have the extra contextual clues to help participants select contextually appropriate meanings (i.e., figurative meanings) as soon as possible, so both contextually appropriate meanings and contextually inappropriate meanings (i.e., literal meanings) were activated in the initial processing, and as the participants collected more information toward the end of their reading of a sentence, as opposed to the initial processing, they abandoned the contextually inappropriate meanings and maintained the contextually appropriate meanings. The results of the neutral contexts could not be explained by the direct access view and the reordered access model because, according to the two models, the figurative meaning was the dominant meaning of an idiom and also the contextually appropriate meaning in the neutral context; therefore, the processing in the neutral context should be as fast as that in the priming context. However, the results did not support the two models.

According to the graded salience hypothesis, when saliency of meaning is taken into account, the most salient meaning is activated in the initial processing regardless of whether it is contextually appropriate or contextually inappropriate. Context comes into play as language users collect more contextual clues, which help them eventually select the contextually appropriate meaning. The English-speaking participants, as advanced English language users, were expected to be familiar with English idioms. For them, then, the salient meanings of the idioms would be the figurative meanings. The Chinese-speaking participants’ significantly longer reading times imply that the salient meanings of the idioms for them were not the figurative meanings but the literal meanings.
When the Chinese-speaking participants processed idioms in the priming contexts, the literal meanings of the idioms were activated in the initial processing; however, as mentioned earlier, the priming contexts helped the participants to abandon contextually inappropriate meanings as soon as possible, so the Chinese-speaking participants abandoned the literal meanings as soon as they were primed by the extra contextual clues. In contrast, the English-speaking participants’ salient meanings were figurative meanings, so when they read idioms in the priming contexts, their salient meanings were contextually appropriate. Unlike the Chinese-speaking participants, the English-speaking participants did not have to abandon contextually inappropriate meanings first; instead, they were able to activate the contextually appropriate meanings immediately.

The processing of the sentences was expected to be harder in the neutral contexts than in the priming contexts. When the Chinese-speaking participants read idioms in the neutral contexts, the salient, literal meanings were activated in the initial processing. As they read more sentences in the neutral contexts, the non-salient figurative meanings were also activated. The participants had difficulty selecting between figurative and literal meanings because they were familiar with the literal meanings of the idioms, but these were contextually inappropriate. On the other hand, the emerging figurative meanings were contextually appropriate, but less familiar (i.e., non-salient). (See Table 9 and Table 12 in Section 4.3.3 for the two language groups’ familiarity ratings of the idioms.) According to the graded salience hypothesis context comes into play to help readers select the appropriate meanings eventually. The Chinese-speaking participants selected the figurative meanings and abandoned the literal meanings only after encountering enough context to support the decision, and this process cost time. In contrast, for the English-speaking participants, the salient meanings of the idioms were the figurative meanings, so even though the literal meanings were available while they were reading in the neutral contexts, they could easily rely on context to select the figurative meanings and abandon the literal meanings. This is because their salient meanings were contextually appropriate, which means that even if they might have been confused by literal meanings during reading, it did not take them much time to decide to abandon these meanings as contextually inappropriate.

In brief, the present study found that the participants spent more time reading in neutral contexts than in priming contexts. The graded salience hypothesis can explain the results that show a context effect.
4.3 Idiom representation in the NS and NNS mental lexicon (RQ3)

**RQ3:** Do native speakers of Chinese and native speakers of English differ in their conception of English idioms as semantically decomposable or semantically nondecomposable?

4.3.1 Data analysis of previous studies

According to Abel (2003), the NS conception of semantic decomposability differs from the NNS conception. NSs tend to conceive of idioms as nondecomposable, while NNSs tend to conceive of idioms as decomposable. To understand how the participants in the present study conceived of the idioms used in the study, I compared the percentages of English-speaking participants and Chinese-speaking participants who judged the idioms to be semantically decomposable and semantically nondecomposable.

Previous studies have used two methods to investigate how NSs and NNSs conceive of idioms in the mental lexicon. Both methods use ratings of idiom decomposability by participants. First, Gibbs and Nayak (1989) showed NS participants idioms that the researchers had designated as semantically decomposable or semantically nondecomposable. The participants decided whether they agreed or disagreed with each idiom’s categorization. The criterion was 75% agreement. For example, if 80% of the participants agreed that an idiom was semantically decomposable and 20% disagreed, the idiom was categorized as semantically decomposable. However, if 65% of the participants agreed that the idiom was semantically decomposable and 35% disagreed, the idiom would not be categorized as either decomposable or nondecomposable. Then, including only the data from the idioms that at least 75% of the participants had agreed were either decomposable or nondecomposable, Gibbs and Nayak calculated the mean percentages of participants who agreed with either categorization. For the decomposable category, the mean percentage was 86%. For the nondecomposable category, the mean percentage was 88%.

Titone and Connine (1994) and Abel (2003) also employed 75% agreement as the criterion to categorize idioms. In addition, these researchers calculated the mean percentages of idioms that were categorized as semantically decomposable or semantically nondecomposable without the criterion. Making the calculation without such a criterion can give us a general picture of how idioms are conceived of in the NSs’ and NNSs’ mental lexicons, while using the criterion can help us understand the tendencies of NSs and NNSs. In Titone and Connine’s study, the mean percentage of the idioms categorized as semantically decomposable by the NS participants was 41.9%, while that of the idioms categorized as
semantically nondecomposable was 58.1%. When the 75% agreement rate was employed, 15% of the idioms (i.e., 26 out of 171) were judged as decomposable, while 35% (i.e., 61 out of 171) were judged as nondecomposable.

The studies conducted by Gibbs and Nayak (1989) and Titone and Connine (1994) both focused on native speakers. In order to compare NS and NNS conceptions of idioms, Abel (2003) asked NNSs to rate the same idioms that Titone and Connine used. Abel found that 55.2% of the NNS participants categorized the idioms as semantically decomposable, while 44.8% of the participants categorized the idioms as semantically nondecomposable. Employing the 75% agreement criterion, Abel found that 29.2% of the idioms (i.e., 38 out of 130) were judged as semantically decomposable, and 17.7% (i.e., 23 out of 130) as semantically nondecomposable.

4.3.2 Data analysis of the present study

First, a regression analysis was employed to examine the relationship between language group and semantic decomposability. The reason to examine this relationship is to support the later analysis and discussion of how NSs and NNSs conceive of the semantic decomposability of idioms. In other words, the later analysis and discussion will be meaningful only if there is a relationship between language group and semantic decomposability. Language group is the independent variable, and the 5-point ratings of semantic decomposability are the dependent variable.

The present study used the same methods as Titone and Connine (1994) and Abel (2003). The mean percentages for semantically decomposable and semantically nondecomposable idioms were calculated for the NS and NNS groups. The purpose was to investigate how NSs and NNSs conceived of idioms in general. In addition, the 75% agreement criterion was used in order to better understand NSs’ and NNSs’ tendencies in how they conceive of idioms in the mental lexicon.

In the present study, the participants rated the idioms’ semantic decomposability on a 5-point Likert scale. If 75% or more of the participants rated an idiom as 4 or 5, the idiom was considered to be semantically decomposable. If 75% or more of the participants rated an idiom as 1 or 2, the idiom was considered to be semantically nondecomposable. Ratings of 3 were excluded because they did not show a tendency toward either category. This method differs from that used in the previous studies (Abel, 2003; Gibbs & Nayak, 1989; Titone & Connine, 1994), in which the participants sorted idioms into different categories on their own.
In the present study, I sorted the idioms into the categories of semantically decomposable and semantically nondecomposable based on the participants’ 5-point Likert scale ratings. It should be noted that because 3-point ratings were excluded from the data analysis, the number of participants whose ratings were included in the calculations is not the same as the total number of participants. As a result, the percentages of participants in the two categories do not add up to 100%.

The present study asked participants to rate semantic decomposability on a continuum because semantic decomposability is a matter of degree. The rating method used in this study is thus able to reflect the nature of semantic decomposability better than the dichotomous method used in the previous studies. One limitation of this study, however, is that it loses information by excluding the 3-point data. A further analysis was conducted to examine the percentages of the excluded 3-point data. For NSs, 115 out of 960 data points were excluded (12%). For NNSs, 134 out of 864 data points were excluded (16%).

4.3.3 Results

The descriptive statistics of semantic decomposability ratings by language group is presented in Table 8. Before answering RQ3, a regression analysis was carried out to examine the relationship between language group and semantic decomposability to ascertain that the two factors were related, and that it was worth analyzing the different language groups’ perception of semantic decomposability. Table 6 below shows the results of the regression analysis. The relationship between language group and semantic decomposability is -.54, which means that when language group changes from NS to NNS, the mean semantic decomposability rating decreases by .54 units. In addition, the $R^2$ squared is .22, which means language group can explain 22% of the participants’ ratings of semantic decomposability. The regression model is significant when language group is added as the predictor ($F = 20.75, p < .001, t = -4.56, p < .001, \alpha = .05$).

Table 6. Results of Regression Analysis for Effects of Language Group on Semantic Decomposability Ratings

<table>
<thead>
<tr>
<th>B</th>
<th>R Squared</th>
<th>F-value</th>
<th>F Change</th>
<th>p-value</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.54</td>
<td>.22</td>
<td>20.75</td>
<td>20.75</td>
<td>&lt;0.001</td>
<td>-4.56</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
In Table 7 (and in Figures 21 and 22 below), the order of idioms is based on the Chinese-speaking participants’ decomposability ratings from the lowest to the highest. Idiom 3 was excluded (see Chapter 3). As the table shows, for most of the idioms, the percentages of participants who conceived of the idioms as semantically decomposable are higher than the percentages of participants who conceived of them as semantically nondecomposable, regardless of language group. The mean percentage of English-speaking participants who conceived of the idioms as semantically decomposable is 80%, and the mean percentage of English-speaking participants who conceived of the idioms as semantically nondecomposable is 9%. In addition, the mean percentage of Chinese-speaking participants who conceived of the idioms as semantically decomposable is 64%, and the mean percentage of Chinese-speaking participants who conceived of the idioms as semantically nondecomposable is 21%. In Table 7, the percentages exceeding the 75% agreement criterion are in boldface. Many percentages in the semantic decomposability category exceed 75%, but none in the semantically nondecomposable category does. More details of the descriptive statistics are shown in Table 8.
Table 7. NS and NNS Ratings of Idioms as Semantically Decomposable and Semantically Nondecomposable

<table>
<thead>
<tr>
<th>Idioms</th>
<th>NS (N = 40)</th>
<th>NS (N = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>10 make no bones about something</td>
<td>14</td>
<td>35%</td>
</tr>
<tr>
<td>5 run something into the ground</td>
<td>28</td>
<td>70%</td>
</tr>
<tr>
<td>23 pay through the nose</td>
<td>27</td>
<td>68%</td>
</tr>
<tr>
<td>13 kick up one's heels</td>
<td>25</td>
<td>63%</td>
</tr>
<tr>
<td>8 hit below the belt</td>
<td>31</td>
<td>78%</td>
</tr>
<tr>
<td>21 cool one's heels</td>
<td>15</td>
<td>38%</td>
</tr>
<tr>
<td>6 force one's hand</td>
<td>28</td>
<td>70%</td>
</tr>
<tr>
<td>19 climb the walls</td>
<td>24</td>
<td>60%</td>
</tr>
<tr>
<td>11 seize the day</td>
<td>34</td>
<td>85%</td>
</tr>
<tr>
<td>24 strike a bargain</td>
<td>38</td>
<td>95%</td>
</tr>
<tr>
<td>16 by word of mouth</td>
<td>36</td>
<td>90%</td>
</tr>
<tr>
<td>7 swallow one's pride</td>
<td>34</td>
<td>85%</td>
</tr>
<tr>
<td>12 fan the flames</td>
<td>33</td>
<td>83%</td>
</tr>
<tr>
<td>2 cover one's tracks</td>
<td>39</td>
<td>98%</td>
</tr>
<tr>
<td>15 burn the midnight oil</td>
<td>34</td>
<td>85%</td>
</tr>
<tr>
<td>22 break the ice</td>
<td>32</td>
<td>80%</td>
</tr>
<tr>
<td>14 call a meeting</td>
<td>39</td>
<td>98%</td>
</tr>
<tr>
<td>1 praise something to the skies</td>
<td>38</td>
<td>95%</td>
</tr>
<tr>
<td>20 follow the crowd</td>
<td>40</td>
<td>100%</td>
</tr>
<tr>
<td>18 break the bank</td>
<td>33</td>
<td>83%</td>
</tr>
<tr>
<td>4 kill two birds with one stone</td>
<td>36</td>
<td>90%</td>
</tr>
<tr>
<td>17 play with fire</td>
<td>36</td>
<td>90%</td>
</tr>
<tr>
<td>9 lend someone a hand</td>
<td>39</td>
<td>98%</td>
</tr>
</tbody>
</table>

*Note.* The numbers represent the numbers of participants. The percentages in the decomposable columns indicate the proportions of participants who rated the idioms at 4–5; the percentages in the nondecomposable columns indicate the proportions of participants who rated the idioms at 1–2.

The percentage of participants who rated idioms as semantically decomposable ranged from 17% to 100% in the Chinese-speaking group, and from 35% to 98% in the English-speaking group. The percentage of participants who rated idioms as semantically nondecomposable ranged from 0% to 64% in the Chinese-speaking group, and from 0% to 40% in the English-speaking group. In other words, the Chinese-speaking participants’
The conception of semantic decomposability had more variation than the English-speaking participants’, regardless of type of semantic decomposability. A similar trend can be found by examining the standard deviations. For semantically decomposable idioms, the mean standard deviation in the NNS data was 30%, and in the NS data was 18%. For semantically nondecomposable idioms, the mean standard deviation in the NNS data was 24%, and in the NS data was 11%. In other words, the English-speaking participants’ conception was more consistent than the Chinese-speaking participants’, regardless of type of semantic decomposability.

### Table 8. Summary of Descriptive Statistics of Percentages by Semantic Decomposability and Language Group

<table>
<thead>
<tr>
<th></th>
<th>Chinese (N = 36)</th>
<th>English (N = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semantically Decomposable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>64%</td>
<td>80%</td>
</tr>
<tr>
<td>SD</td>
<td>30%</td>
<td>18%</td>
</tr>
<tr>
<td>Range</td>
<td>83%</td>
<td>65%</td>
</tr>
<tr>
<td>Min</td>
<td>17%</td>
<td>35%</td>
</tr>
<tr>
<td>Max</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Semantically Nondecomposable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>21%</td>
<td>9%</td>
</tr>
<tr>
<td>SD</td>
<td>24%</td>
<td>11%</td>
</tr>
<tr>
<td>Range</td>
<td>64%</td>
<td>40%</td>
</tr>
<tr>
<td>Min</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Max</td>
<td>64%</td>
<td>40%</td>
</tr>
</tbody>
</table>

The mean percentages in each category for each language group suggest that both Chinese-speaking and English-speaking participants tended to conceive of idioms as semantically decomposable rather than semantically nondecomposable. Figure 18 below presents the ratio of the two types of semantic decomposability in the two language groups. For the Chinese-speaking participants, the ratio of semantically decomposable to semantically nondecomposable idioms was 3:1 (64%/21%). For English-speaking
participants, the ratio was 8.9:1 (80%/9%). In other words, the English-speaking participants tended to conceive of idioms as semantically decomposable much more than the Chinese-speaking participants did.

Figure 18. Ratio of two types of semantic decomposability

Figures 19 and 20 below visualize the information in Table 7 in bar charts; like Figure 18, they depict the NSs’ stronger tendency to conceive of idioms as semantically decomposable. For most of the idioms, higher percentages of NSs than NNSs judged them to be semantically decomposable. However, for a few of the idioms, higher percentages of NNSs than NSs considered them to be semantically decomposable: Idioms 15, 22, 18, 4, 17, and 9. These specific idioms will be discussed in detail with the presentation of Table 9.
Figure 19. Percentages in the semantically decomposable category across languages
Figure 20. Percentages in the semantically nondecomposable category across languages
Table 9 below presents the mean scores on familiarity, semantic decomposability, and metaphoric equivalence for the six idioms (Idioms 15, 22, 18, 4, 17, and 9) that higher percentages of NNSs than NSs considered semantically decomposable. In the table, the grand mean is the mean score of all the idioms across all the participants. For these six specific idioms, the mean scores of semantic decomposability from the NNS group are all higher than those from the NS group. All the mean scores from the NNS group are above 4. In the semantic decomposability survey, participants rated how well they thought they could understand an idiom from considering the individual words, and a 4-point rating meant “a lot.” In addition, although the NNSs’ mean scores of familiarity for these six idioms are lower than the NSs’, all the NNSs’ mean scores for these idioms are above the overall NNS mean score of familiarity (i.e., 3.1). In the NNS group, the mean scores of familiarity for most idioms are above 4, except for Idiom 15. In the familiarity survey, the description of 4 points means the NNSs thought they had seen the phrases before, and they thought they knew the meanings. As for metaphoric equivalence between L1 and L2 for the six specific idioms, the NNSs’ mean scores for all except one (Idiom 22) are above 4. In the metaphoric equivalence survey, the participants rated the degree to which they thought the L1 metaphor was similar to the L2 metaphor; a 4-point rating meant “a lot.” All the mean scores of metaphoric equivalence for these six idioms are above the overall mean score of metaphoric equivalence (i.e., 3.49).
Before making a conclusion regarding these six idioms, I examined the relationship between semantic decomposability, familiarity, and metaphoric equivalence by language group. In general, NSs tended to conceive of idioms as decomposable more than NNSs did (see Table 8). However, for these six idioms, the percentage of native Chinese speakers who perceived them to be decomposable was higher than that of native English speakers. Therefore, the present study investigated the characteristics of the six idioms and compared them with the other 17 idioms (i.e., the idioms for which the percentage of native Chinese speakers who perceived them to be decomposable was lower than that of native English speakers). The examination of the correlations among these ratings will facilitate a later discussion of the characteristics of the idioms that were conceived of as semantically decomposable and semantically nondecomposable by most of the NNSs. In addition to calculating the two language groups’ mean scores on these six idioms, I conducted correlation analysis of the ratings of all 23 idioms for each language group separately. Table 10 shows that the NNSs’ ratings of familiarity, semantic decomposability, and metaphoric equivalence were significantly correlated, and Table 11 shows that the NSs’ ratings of familiarity and semantic decomposability were significantly correlated.
Table 10. NNS: Correlations between Familiarity, Semantic Decomposability, and Metaphoric Equivalence

<table>
<thead>
<tr>
<th></th>
<th>Familiarity</th>
<th>Semantic Decomposability</th>
<th>Metaphoric Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarity</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Decomposability</td>
<td>.92**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Metaphoric Equivalence</td>
<td>.85**</td>
<td>.94**</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. ** significant at the .01 level

Table 11. NS: Correlations between Familiarity and Semantic Decomposability

<table>
<thead>
<tr>
<th></th>
<th>Familiarity</th>
<th>Semantic Decomposability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarity</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Semantic Decomposability</td>
<td>.85**</td>
<td>1</td>
</tr>
</tbody>
</table>

** significant at the .01 level

As discussed above, for the six idioms that the NNSs rated semantically decomposable at higher rates than the NSs, the mean semantic decomposability ratings, the mean familiarity ratings, and the mean metaphoric equivalence ratings are all above the grand means. The correlation analysis shows a strong tendency toward positive correlations between the NNSs’ ratings on the three surveys. In other words, when most of the Chinese-speaking participants conceived of an idiom as highly semantically decomposable, they were also very familiar with the idiom and conceived of it as highly metaphorically equivalent between L1 and L2.

Based on these results, a follow-up question is to investigate whether these correlations only apply to these six specific idioms. If similar results are obtained for the 17 idioms that the NNSs rated as semantically decomposable at lower rates than the NSs, then the correlations are not unique to the idioms that the NNSs rated as semantically decomposable at higher rates than the NSs. Table 12 below shows the mean scores from the three surveys for
the other 17 idioms. In this table, mean scores lower than the grand mean are in boldface. Only four out of the 17 idioms (24%) have no mean scores below the grand mean (Idioms 7, 14, 16, and 20). For the remaining 13 idioms (76%), at least one mean score is below the grand mean, and for nine (69%), the mean scores are all below the grand mean. In other words, when lower percentages of NNSs than NSs conceived of an idiom as semantically decomposable, the NNSs were more likely to consider the idiom to be less familiar, less semantically decomposable, and less metaphorically equivalent to an L1 idiom than the idioms that higher percentages of NNSs than NSs conceived of as semantically decomposable. Therefore, these results support that the characteristics shared by six specific idioms among the 23 idioms in the study were unique to these six idioms.
Table 12. NS and NNS: Mean Scores for the Other 17 Idioms

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Familiarity</th>
<th>SD</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>praise something to the skies (Idiom 1)</td>
<td>3.03</td>
<td>4.4</td>
<td>4.56</td>
</tr>
<tr>
<td>cover one’s tracks (Idiom 2)</td>
<td>2.94</td>
<td>4.68</td>
<td>4.22</td>
</tr>
<tr>
<td>run something into the ground (Idiom 5)</td>
<td>1.83</td>
<td>4.48</td>
<td>2.42</td>
</tr>
<tr>
<td>force one’s hand (Idiom 6)</td>
<td>3.03</td>
<td>4.33</td>
<td>3</td>
</tr>
<tr>
<td>swallow one’s pride (Idiom 7)</td>
<td>3.22</td>
<td>4.73</td>
<td>4.11</td>
</tr>
<tr>
<td>hit below the belt (Idiom 8)</td>
<td>1.97</td>
<td>4.58</td>
<td>2.72</td>
</tr>
<tr>
<td>make no bones about something (Idiom 10)</td>
<td>1.44</td>
<td>3</td>
<td>2.19</td>
</tr>
<tr>
<td>seize the day (Idiom 11)</td>
<td>2.86</td>
<td>4.48</td>
<td>3.25</td>
</tr>
<tr>
<td>fan the flames (Idiom 12)</td>
<td>3.03</td>
<td>4.65</td>
<td>4</td>
</tr>
<tr>
<td>kick up one’s heels (Idiom 13)</td>
<td>1.89</td>
<td>3.95</td>
<td>2.28</td>
</tr>
<tr>
<td>call a meeting (Idiom 14)</td>
<td>3.86</td>
<td>4.7</td>
<td>4.64</td>
</tr>
<tr>
<td>by word of mouth (Idiom 16)</td>
<td>3.28</td>
<td>4.88</td>
<td>3.86</td>
</tr>
<tr>
<td>climb the walls (Idiom 19)</td>
<td>2.28</td>
<td>3.43</td>
<td>3.25</td>
</tr>
<tr>
<td>follow the crowd (Idiom 20)</td>
<td>4.19</td>
<td>4.7</td>
<td>4.78</td>
</tr>
<tr>
<td>cool one’s heels (Idiom 21)</td>
<td>1.92</td>
<td>2.43</td>
<td>2.75</td>
</tr>
<tr>
<td>pay through the nose (Idiom 23)</td>
<td>1.75</td>
<td>3.93</td>
<td>2.33</td>
</tr>
<tr>
<td>strike a bargain (Idiom 24)</td>
<td>2.33</td>
<td>4.28</td>
<td>3.64</td>
</tr>
<tr>
<td>Grand Mean</td>
<td>3.1</td>
<td>4.36</td>
<td>3.75</td>
</tr>
</tbody>
</table>

Figure 20 above presents a graphic comparison of the relative percentages of NNSs and NSs conceiving of idioms as semantically nondecomposable. For 13 idioms (i.e., Idioms 10, 5, 23, 13, 8, 21, 11, 24, 16, 7, 12, 2, and 14), higher percentages of NNSs than NSs considered them to be nondecomposable. Taking the results displayed in Figure 20 and Table 12 together, the mean scores of the NNSs’ survey ratings of seven out of the 13 idioms (54%; Idioms 10, 5, 23, 13, 8, 21, and 24) were below the grand means in the NNS group on all three surveys. In other words, when more NNSs than NSs conceived of an idiom as semantically nondecomposable (Figure 20), the NNSs also considered the idiom to be less familiar, less semantically decomposable, and less metaphorically equivalent than the idioms that more NNSs than NSs conceived of as semantically decomposable (Figure 19).
In general, more NSs than NNSs conceived of the idioms as semantically decomposable. However, when more NNSs than NSs considered an idiom as semantically decomposable, it meant that the NNSs also tended to think that the figurative meaning of the idiom could be inferred from its constituents at higher rates than the NSs (this is how participants were asked to rate idioms in the semantic decomposability survey). The highly semantically decomposable idioms for the NNSs (Figure 19) have two common features: These idioms were very familiar to the NNSs, and they had close metaphorical equivalents in the L1 and L2. By contrast, when more NNSs than NSs conceived of an idiom as semantically nondecomposable, the NNSs also tended to find the idiom less familiar, less semantically decomposable, and less metaphorically equivalent in L1 and L2 than the other idioms.

As Table 7 above shows, based on the criterion of 75% agreement, the English-speaking participants considered 16 out of 23 idioms (69.6%) to be semantically decomposable, and the Chinese-speaking participants considered 12 out of 23 idioms (52.2%) to be semantically decomposable. By this criterion, neither group considered any of the idioms to be semantically nondecomposable (Table 7). The two language groups’ conception of 12 idioms (52%; Idioms 1, 2, 4, 7, 9, 12, 14, 15, 17, 18, 20, and 22) overlapped; that is, these 12 can be placed in the semantically decomposable category according to the 75% agreement criterion for both groups. The mean scores of semantic decomposability of these overlapping idioms were all above the grand means for both language groups. For the NSs, the mean familiarity ratings of all of the overlapping idioms except for Idiom 15 (4.33) are all above the grand mean (4.36). For the NNSs, the mean familiarity ratings of the overlapping idioms do not show a clear pattern; some of the mean scores are above the grand mean, while others are below it. However, for the NNSs, the means of the metaphoric equivalence ratings of all 12 overlapping idioms are above the grand mean.

To sum up, using the 75% agreement criterion, 12 idioms were conceived of as semantically decomposable by both the NSs and the NNSs. For the NSs, the 12 overlapping idioms were highly semantically decomposable and familiar. For the NNSs, the 12 overlapping idioms were highly semantically decomposable and metaphorically equivalent between L1 and L2.

The use of the 75% agreement criterion shows that the English-speaking participants tended to conceive of more idioms as semantically decomposable than the Chinese-speaking participants did. This result strengthens the finding on the ratios of the two types of semantic
decomposability rating, presented in Figure 18. However, for the Chinese-speaking participants, when the 75% agreement criterion was applied, the numbers of semantically decomposable and semantically nondecomposable idioms were almost the same (12 and 11 respectively), which implies that the Chinese-speaking participants did not have a tendency to conceive of the idioms as all belonging to either one of the categories. This result differs from the results presented in Table 8, which shows that the mean percentages were higher for the semantically decomposable idioms than the semantically nondecomposable idioms (64% vs. 21%). Furthermore, the use of the 75% agreement criterion provides a clear picture of the tendencies of NS and NNS perceptions.

In brief, Figure 18 shows that both native speakers of Chinese and native speakers of English tended to conceive of the English idioms used in the study as semantically decomposable as opposed to semantically nondecomposable, with the native speakers of English showing this tendency more strongly. These results differ from the study’s prediction. However, as Figure 19 shows, there are some idioms that more native speakers of Chinese than native speakers of English conceived of as semantically decomposable. The characteristics of these idioms were high familiarity, high semantic decomposability, and high metaphoric equivalence between L1 and L2, according to the NNSs. In addition, as Figure 20 shows, the NNSs conceived of more idioms as semantically nondecomposable (N = 15) than did the NSs (N = 4). The idioms that the NNSs conceived of as being semantically nondecomposable were less familiar, less semantically decomposable, and less metaphorically equivalent to the NNSs, compared with the other idioms that they considered to be semantically nondecomposable.

4.3.4 Discussion

The idioms used in this study were extracted from those used by Titone and Connine (1994) and Abel (2003). The current study shows that both NSs and NNSs tended to conceive of the idioms as semantically decomposable rather than nondecomposable. The NNSs’ results are similar to those of the previous studies, but the NSs’ results are different from those of the previous studies, which reported that NSs tended to conceive of idioms as semantically nondecomposable. Table 13 below is a comparison of the findings of the previous studies and the current study. In the table, “mean percentage” indicates the average percentage of participants who judged idioms as belonging to one of the two semantic decomposability categories. “Mean percentage (75% agreement)” indicates only the idioms for which the
percentages exceeded 75%. “N of idioms (75% agreement)” and “% of idioms (75% agreement)” indicate the number and percentage of idioms for which the 75% agreement criterion was met.
Table 13. Comparison of the Current Study and the Previous Studies

<table>
<thead>
<tr>
<th>Studies</th>
<th>Semantically decomposable</th>
<th>Semantically nondecomposable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gibbs &amp; Nayak (1989) (NS, 36 idioms)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean percentage (75% agreement)</td>
<td>86%</td>
<td>88%</td>
</tr>
<tr>
<td><strong>My Study (NS, 23 idioms)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean percentage (75% agreement)</td>
<td>90%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Titone &amp; Connine (1994) (NS, 171 idioms)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean percentage</td>
<td>41.9%</td>
<td>58.1%</td>
</tr>
<tr>
<td>N of idioms (75% agreement)</td>
<td>26</td>
<td>61</td>
</tr>
<tr>
<td>% of idioms (75% agreement)</td>
<td>15.2%</td>
<td>35.7%</td>
</tr>
<tr>
<td><strong>My Study (NS, 23 idioms)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean percentage</td>
<td>80%</td>
<td>9%</td>
</tr>
<tr>
<td>N of idioms (75% agreement)</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>% of idioms (75% agreement)</td>
<td>69.6%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Abel (2003) (NNS) Study 2 (130 idioms)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean percentage</td>
<td>55.2%</td>
<td>44.8%</td>
</tr>
<tr>
<td>N of idioms (75% agreement)</td>
<td>38</td>
<td>23</td>
</tr>
<tr>
<td>% of idioms (75% agreement)</td>
<td>29.2%</td>
<td>17.7%</td>
</tr>
<tr>
<td><strong>My Study (NNS, 23 idioms)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean percentage</td>
<td>64%</td>
<td>21%</td>
</tr>
<tr>
<td>N of idioms (75% agreement)</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>% of idioms (75% agreement)</td>
<td>52.2%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Note. Abel’s (2003) Study 1 is not relevant to the current study; therefore, only Study 2’s results appear in the table.
In the current study, when the 75% agreement criterion was applied, the mean percentage of semantically decomposable idioms for the NS group was 90%, which is close to Gibbs and Nayak’s (1989) finding of 86%. However, Gibbs and Nayak also found a mean percentage of semantically nondecomposable idioms of 88%, while the current study found that no idiom was seen as semantically nondecomposable by this criterion. In addition, the current study’s participants judged the idioms to be semantically decomposable at higher rates (NS: mean percentage, 80%; NNS: mean percentage, 64%) than did the participants in Titone and Connine’s (NS: 41.9%) and Abel’s (NNS: 55.2%) studies.

When the 75% agreement criterion is applied, the difference in percentages between the current study and the previous studies is larger. In the current study, 69.6% of the idioms were judged as semantically decomposable by the NSs, but in Titone and Connine’s (1994) study, only 15.2% of the idioms were judged as semantically decomposable by the NS participants. Moreover, in the current study, 52.2% of the idioms were judged as semantically decomposable by the NNSs, but in Abel’s (2003) study, only 29.2% of the idioms were judged as semantically decomposable by the NNS participants. In the current study, when the 75% agreement criterion was applied, no idiom was judged as semantically nondecomposable. However, in Titone and Connine’s study, 35.7% of the idioms were judged as semantically nondecomposable, and in Abel’s, 17.7% of the idioms were judged as semantically nondecomposable.

The NNSs’ results in the current study are consistent with those in the previous studies when the mean percentages and the percentages of idioms under the 75% agreement criterion are taken into consideration. The NNSs tended to perceive idioms as semantically decomposable rather than semantically nondecomposable. However, the NS results in the current study are inconsistent with those in the previous studies. In the previous studies, when the mean percentages and percentages of idioms under the 75% agreement criterion were taken into consideration, the NSs were found to conceive of idioms as nondecomposable, while in the current study, the NSs were found to conceive of idioms as decomposable. The inconsistency may result from the tasks given to the participants. As already discussed, the previous studies asked the participants to consciously sort the idioms into either the semantically decomposable category or the semantically nondecomposable category. However, the current study asked the participants to rate the semantic decomposability of the idioms on a 5-point scale.
One problem of the sorting method used in the previous studies is that the participants were instructed to judge idioms as semantically decomposable if the individual components could contribute to the overall figurative meanings, and as semantically nondecomposable if not. However, it is very likely that some constituents could contribute to the overall figurative meanings, while others could not. The instructions did not tell the participants how to determine to what degree the components must contribute to the overall figurative meaning for the idiom to be judged as semantically decomposable (or nondecomposable). Moreover, the previous studies did not report reliability. Therefore, it is unknown how reliable their sorting method is.

In the present study, the participants were allowed to judge semantic decomposability on a continuum. This means that the participants could judge the idioms based on the degree to which individual components contributed to the overall figurative meanings. Moreover, the reliability of the semantic decomposability survey was 0.92, which means the results are highly reliable.

The present results for the NSs’ and NNSs’ judgments of semantic decomposability can be explained by degree of familiarity. The correlation between these two factors is .92 for the NNSs (Table 10), and .85 for the NSs (Table 11). In other words, the more semantically decomposable an idiom is to a language user, the more familiar it is to the language user, and vice versa. Abel (2003) also found that the degree of semantic decomposability is related to the degree of familiarity. In Abel’s study, for the NSs, the familiarity mean score of semantically decomposable idioms was 5.92 on a scale of 7, and the familiarity mean score of semantically nondecomposable idioms was 5.76. For the NNSs, the familiarity mean score of semantically decomposable idioms was 4.9 on a scale of 7, and the familiarity mean score of semantically nondecomposable idioms was 2.99. Moreover, these results show that the NSs’ familiarity with semantically decomposable idioms was higher than the NNSs’, which is consistent with the correlation results found in the present study.

In the current study, the NSs’ mean familiarity rating score is 4.31 on a scale of 5 points (SD = .42), and that of the NNSs is 3.08 (SD = .66). An independent samples t-test shows that the English-speaking participants’ familiarity was significantly higher than the Chinese-speaking participants’ (t-value = 9.70; df = 74; p < .001; α = .05). The effect size of the mean difference is 2.22 standard deviations. According to Cohen’s d index, this effect size is large. In other words, the NSs were much more familiar with the idioms in the study than the NNSs.
This result is consistent with Abel’s findings. Table 14 gives a summary of the results of the independent samples $t$-test.

Table 14. Summary of Independent Samples $t$-Test of Mean Familiarity Rating Scores

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>$t$-test</th>
<th>$p$-value</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>4.31</td>
<td>0.42</td>
<td>9.7</td>
<td>$p &lt; .001$</td>
<td>2.22</td>
</tr>
<tr>
<td>Chinese</td>
<td>3.08</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because familiarity is positively related to semantic decomposability, and the Chinese-speaking participants tended to conceive of idioms as semantically decomposable, it is very likely that the English-speaking participants, whose familiarity with the idioms was significantly higher, also conceived of the idioms as semantically decomposable. In Abel’s dual idiom representation model, semantic decomposability is related to how idioms are stored and retrieved in the mental lexicon. A decomposable idiom has both an idiom entry and constituent entries in the mental lexicon, while a nondecomposable idiom has an idiom entry only. In the present study, although both NSs and NNSs conceived of the idioms as semantically decomposable more than semantically nondecomposable, this finding does not necessarily mean that the participants in both language groups stored and retrieved the idioms as both constituent entries and idiom entries in their mental lexicons. This is because the English-speaking participants were very familiar with the idioms, which would have facilitated their ability to establish idiom entries; according to the dual idiom representation model, language users can develop an idiom entry for a high-familiarity, decomposable idiom more easily than for a low-familiarity, decomposable idiom. In contrast, the Chinese-speaking participants were less familiar with the idioms, so it is very likely that they did not have idiom entries for most of the idioms; as a result, according to the dual idiom representation model, they would have tended to store the idioms as constituent entries in their mental lexicons. When they encountered the unfamiliar idioms, the Chinese-speaking participants would have had to rely on the literal meanings of the constituents and link the literal meanings to their metaphoric referents in order to try to understand the figurative meanings of the idioms. In other words, the results for semantic decomposability and familiarity show that the native speakers of English tended to have idiom entries, while the native speakers of Chinese tended to have constituent entries, which is consistent with the
dual idiom representation model’s assumption that, as a result of extensive language exposure (i.e., high familiarity with idioms) native speakers have more idiom entries than L2 learners have.

In summary, this study’s results differ from those of previous studies, probably for two main reasons: First, this study’s sorting method differs from the previous studies’ method, and second, the NS participants in this study were very familiar with most of the idioms.

4.4 Factors influencing NS and NNS idiom processing (RQ4 & RQ5)

**RQ4:** Which factors significantly influence the English idiom processing of native speakers of English and native speakers of Chinese: familiarity, semantic decomposability, and metaphoric equivalence?

**RQ5:** Is there any interaction effect among familiarity, semantic decomposability, and metaphoric equivalence in the English idiom processing of native speakers of English and native speakers of Chinese?

4.4.1 Data analysis

In the present study, the 5-point ratings from the familiarity, semantic decomposability, and metaphoric equivalence surveys were treated as continuous variables. Rhemtulla, Brosseau-Liard, and Savalei (2012) compared the categorical least squares (cat-LS) method and the maximum likelihood (ML) method when estimating confirmatory factor analysis (CFA) models with ordinal variables (i.e., two to seven categories). The cat-LS is a common method of parameter estimation for categorical data, as the ML is for continuous nonnormal data. The researchers found that once the number of categories reached five, the ML method and the cat-LS method performed equally well. In other words, the cat-LS method is good at parameter estimation for ordinal data, but when the number of categories reaches five, the ML method, which is used for continuous data, is acceptable, too. Therefore, in the present study, the rating data, which was ordinal in nature, was treated as continuous data because the number of categories reached five, and the ML method was employed.

As reported in Section 3.3.1.1, factor analysis showed that all the familiarity ratings loaded on Factor 1 and most of the semantic decomposability ratings loaded on Factor 2 except for Idiom 3, for which the semantic decomposability ratings loaded on Factor 1. Therefore, all the data points related to Idiom 3 were excluded from further analysis. The remaining data for each language group were examined for outliers, which were replaced by the maximum/minimum of the whiskers of the boxplots.
Mixed-effects modeling was applied to investigate what factors significantly affected the participants’ idiom processing. The dependent variable was the reaction time difference between priming contexts and neutral contexts. The independent variables for the English-speaking group were familiarity, semantic decomposability, and the interaction between familiarity and semantic decomposability. The independent variables for the Chinese-speaking group were familiarity, semantic decomposability, metaphoric equivalence, and the interactions among familiarity, semantic decomposability, and metaphoric equivalence. In the mixed-effects modeling, these independent variables were fixed effects, and subjects and items were random effects. The reason to include subjects and items as random effects is to account for the variance associated with the random effects in the error term in order to obtain a more conservative and generalizable result, compared with the result obtained from a fixed effects model. Therefore, the conservative result is more generalizable to the population and to all similarly selected idiom items.

4.4.1.1 Model construction in a comparable previous study

The present study adopted the procedures described by Locker, Hoffman, and Bovaird (2007) when carrying out mixed-effects modeling. In Locker et al.’s study, 38 undergraduate students read either words or non-words on a computer screen, and had to decide whether what they read was a word or non-word as quickly as possible. The 39 words in the test were categorized by the frequency of orthographic neighbors (high vs. low) and semantic neighborhood (large vs. small). Before the researchers examined whether subjects and items should be considered as random effects, a baseline (empty) model with no random effects (i.e., only one error term) was examined for future comparison. The equation is shown below:

\[ Y_{si} = \gamma_0 + \varepsilon_{si} \]  

(1)

Where \( Y_{si} \) = the observed RT for subject \( s \) and item \( i \);

\( \gamma_0 \) = the intercept, or expected mean RT for the overall sample;

\( \varepsilon_{si} \) = the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

Random effects for subjects and items were added next. The equation is shown below:

\[ Y_{si} = \gamma_0 + U_{0s} + V_{0i} + \varepsilon_{si} \]  

(2)

Where \( Y_{si} \) = the observed RT for subject \( s \) and item \( i \);
\( \gamma_0 = \) the intercept, or expected mean RT for the overall sample;  
\( U_{0s} = \) the random effect of subject \( s \);  
\( V_{0i} = \) the random effect of item \( i \);  
\( \epsilon_{si} = \) the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

To estimate whether adding random effects would improve the model fit, the value of -2 restricted log likelihood (LL) in the subject-and-item as random effects model was subtracted from that in the baseline model. The difference of the model deviances is distributed as a chi-square. The degrees of freedom are equal to the difference in the number of parameters estimated within each model, and alpha is set at .05. If the difference in the -2 restricted LL is greater than the critical value in the chi-square distribution, the model is improved after adding the random effects. If the difference in the -2 restricted LL is less than the critical value in the chi-square distribution, the model is not improved, which means adding the random effects is unnecessary.

In Locker et al.'s (2007) study, adding the two random effects improved the model fit, so the random effects were kept and the two fixed effects (i.e., neighborhood frequency and semantic neighborhood) were added. The equation is as below:

\[
Y_{si} = \gamma_0 + \gamma_1(\text{neighborhood frequency}) + \gamma_2(\text{semantic neighborhood}) + \gamma_3(\text{neighborhood frequency})(\text{semantic neighborhood}) + U_{0s} + V_{0i} + \epsilon_{si} \tag{3}
\]

Where \( Y_{si} = \) the observed RT for subject \( s \) and item \( i \);  
\( \gamma_0 = \) the intercept, or expected mean RT for the overall sample;  
\( \gamma_1 = \) the main effect of neighborhood frequency;  
\( \gamma_2 = \) the main effect of semantic neighborhood;  
\( \gamma_3 = \) the interaction effect of neighborhood frequency and semantic neighborhood;  
\( U_{0s} = \) the random effect of subject \( s \);  
\( V_{0i} = \) the random effect of item \( i \);  
\( \epsilon_{si} = \) the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

The present study used the same procedure to estimate whether adding the random effects, the fixed effects, and the interaction effect would improve the model fit.

4.4.1.2 Model construction in the present study
In the present study, a baseline model and a subject-and-item as random effects model were constructed first to examine whether random effects should be considered, and then language group as a fixed effect was added. The reason to examine whether language group is a significant factor affecting RT differences is because, if it is, then it should be included in the models along with the other identified factors, namely familiarity and semantic decomposability. However, in addition to familiarity and semantic decomposability, the Chinese-speaking participants’ idiom processing was affected by metaphoric equivalence. Therefore, it is necessary to examine the influence of this factor in a separate model. And if language group is not a significant factor, it should not be kept in the model. In other words, if language group is not a significant factor, the NS data and NNS data can be analyzed separately. In this case, the NS models should include familiarity and semantic decomposability, and the NNS models should include familiarity, semantic decomposability, and metaphoric equivalence.

The following section presents the formula of the model when language group is a fixed effect and when it is not. Given the discussion above, the equation for language group as a fixed effect is as follows:

$$Y_{si} = \gamma_0 + \gamma_1(\text{language group}) + U_0s + V_0i + \epsilon_{si} \quad (4)$$

Where $Y_{si}$ = the observed RT for subject $s$ and item $i$;
$\gamma_0$ = the intercept, or expected mean RT for the overall sample;
$\gamma_1$ = the main effect of language group;
$U_0s$ = the random effect of subject $s$;
$V_0i$ = the random effect of item $i$;
$\epsilon_{si}$ = the residual deviation from the sample mean RT for subject $s$ and item $i$.

If language group is a significant factor, it should be included in the data analysis. In this case, the equation is as follows:

$$Y_{si} = \gamma_0 + \gamma_1(\text{language group}) + \gamma_2(\text{familiarity}) + \gamma_3(\text{semantic decomposability}) + U_0s + V_0i + \epsilon_{si} \quad (5)$$

Where $Y_{si}$ = the observed RT for subject $s$ and item $i$;
$\gamma_0$ = the intercept, or expected mean RT for the overall sample;
\( \gamma_1 \) = the main effect of language group;
\( \gamma_2 \) = the main effect of familiarity;
\( \gamma_3 \) = the main effect of semantic decomposability;
\( U_{0s} \) = the random effect of subject \( s \);
\( V_{0i} \) = the random effect of item \( i \);
\( \varepsilon_{si} \) = the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

Because metaphoric equivalence data is only available for the Chinese-speaking participants, a separate analysis was needed to investigate the influence of metaphoric equivalence on NNS idiom processing. The equation is as follows:

\[
Y_{si} = \gamma_0 + \gamma_1(\text{metaphoric equivalence}) + U_{0s} + V_{0i} + \varepsilon_{si}
\]  
(6)

Where \( Y_{si} \) = the observed RT for subject \( s \) and item \( i \);
\( \gamma_0 \) = the intercept, or expected mean RT for the overall sample;
\( \gamma_1 \) = the main effect of metaphoric equivalence;
\( U_{0s} \) = the random effect of subject \( s \);
\( V_{0i} \) = the random effect of item \( i \);
\( \varepsilon_{si} \) = the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

However, if language group is not a significant predictor, the equations are as follows:

**English-speaking participants**

\[
Y_{si} = \gamma_0 + \gamma_1(\text{familiarity}) + \gamma_2(\text{semantic decomposability}) + U_{0s} + V_{0i} + \varepsilon_{si}
\]  
(7)

Where \( Y_{si} \) = the observed RT for subject \( s \) and item \( i \);
\( \gamma_0 \) = the intercept, or expected mean RT for the overall sample;
\( \gamma_1 \) = the main effect of familiarity;
\( \gamma_2 \) = the main effect of semantic decomposability;
\( U_{0s} \) = the random effect of subject \( s \);
\( V_{0i} \) = the random effect of item \( i \);
\( \varepsilon_{si} \) = the residual deviation from the sample mean RT for subject \( s \) and item \( i \).
An interaction term between familiarity and semantic decomposability is added to examine whether there is an interaction effect. The equation is shown below:

\[ Y_{si} = \gamma_0 + \gamma_1(familiarity) + \gamma_2(semantic \ decomposability) + \gamma_3(familiarity)(semantic \ decomposability) + U_0s + V_0i + \epsilon_{si} \]  

(8)

Where \( Y_{si} \) = the observed RT for subject \( s \) and item \( i \);
\( \gamma_0 \) = the intercept, or expected mean RT for the overall sample;
\( \gamma_1 \) = the main effect of familiarity;
\( \gamma_2 \) = the main effect of semantic decomposability;
\( \gamma_3 \) = the interaction effect between familiarity and semantic decomposability;
\( U_0s \) = the random effect of subject \( s \);
\( V_0i \) = the random effect of item \( i \);
\( \epsilon_{si} \) = the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

**Chinese-speaking participants**

\[ Y_{si} = \gamma_0 + \gamma_1(familiarity) + \gamma_2(semantic \ decomposability) + \gamma_3(meta\phoric \ equivalence) + U_0s + V_0i + \epsilon_{si} \]  

(9)

Where \( Y_{si} \) = the observed RT for subject \( s \) and item \( i \);
\( \gamma_0 \) = the intercept, or expected mean RT for the overall sample;
\( \gamma_1 \) = the main effect of familiarity;
\( \gamma_2 \) = the main effect of semantic decomposability;
\( \gamma_3 \) = the main effect of metaphoric equivalence;
\( U_0s \) = the random effect of subject \( s \);
\( V_0i \) = the random effect of item \( i \);
\( \epsilon_{si} \) = the residual deviation from the sample mean RT for subject \( s \) and item \( i \).

In addition to testing the main effects of familiarity, semantic decomposability, and metaphoric equivalence, three interaction terms are added to examine whether there is any interaction effect among the three factors.

\[ Y_{si} = \gamma_0 + \gamma_1(familiarity) + \gamma_2(semantic \ decomposability) + \gamma_3(meta\phoric \ equivalence) \]
Where $Y_{si} = \text{the observed RT for subject } s \text{ and item } i$;

$\gamma_0 = \text{the intercept, or expected mean RT for the overall sample;}$

$\gamma_1 = \text{the main effect of familiarity;}$

$\gamma_2 = \text{the main effect of semantic decomposability;}$

$\gamma_3 = \text{the main effect of metaphoric equivalence;}$

$\gamma_4 = \text{the interaction effect between familiarity and semantic decomposability;}$

$\gamma_5 = \text{the interaction effect between familiarity and metaphoric equivalence;}$

$\gamma_6 = \text{the interaction effect between semantic decomposability and metaphoric equivalence;}$

$U_{0s} = \text{the random effect of subject } s;$

$V_{0i} = \text{the random effect of item } i;$

$\epsilon_{si} = \text{the residual deviation from the sample mean RT for subject } s \text{ and item } i.$

4.4.2 Results

As mentioned earlier, the present study adopted Locker et al.’s (2007) procedures to conduct mixed-effects modeling. SPSS 19 (George & Mallery, 2003) was used for modeling purposes. In Locker et al.’s study, the authors only examined random intercepts, but in the present study, I also examined random slopes. Furthermore, Locker et al.’s study constructed its models from the simple to the complex, but this study followed the common approach to mixed-effect models within the sentence processing community by constructing models from the complex to the simple.

To address RQ4 and RQ5, the dependent variable was the reaction time difference between the priming and the neutral contexts. After the RTs had been log transformed, and the outliers were replaced by 3SDs, the RTs in the priming contexts were subtracted from the RTs in the neutral contexts.

4.4.2.1 The language group model

In this model, language group was included as the fixed effect, and subjects and items were included as the random intercepts. Because language group was a between-subject factor, it was not included as the random slope in the model. Table 15 shows that language group was not a good predictor ($t = .36; p = .72; \text{effect size} = -.01$).
Table 15. Language Group Predictor Model

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>b</th>
<th>B</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Fixed effect (γ0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>38.03</td>
<td>-3.75</td>
<td>&lt;0.001</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.02</td>
</tr>
<tr>
<td>Language Group</td>
<td>1724.00</td>
<td>0.36</td>
<td>0.72</td>
<td>0.004</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Variance</th>
<th>Wald Z</th>
<th>p</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Random effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.00</td>
<td>2.37</td>
<td>0.02</td>
<td>0.001</td>
</tr>
<tr>
<td>Item</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual (εsi)</td>
<td>0.06</td>
<td>29.36</td>
<td>&lt;0.001</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Given this finding, the models for the English-speaking group and the Chinese-speaking group will be discussed separately. For the English-speaking group, two models were examined. One is a two-predictor model (i.e., familiarity and semantic decomposability). The other is a three-predictor model (i.e., familiarity, semantic decomposability, and familiarity x semantic decomposability). For the Chinese-speaking group, as well, two models were examined. One is a three-predictor model (i.e., familiarity, semantic decomposability, and metaphoric equivalence). The other is a six-predictor model (i.e., familiarity, semantic decomposability, metaphoric equivalence, familiarity x semantic decomposability, semantic decomposability x metaphoric equivalence, and familiarity x metaphoric equivalence).

4.4.2.4 English-speaking group: Two-predictor model

Familiarity and semantic decomposability were the two predictors added to the model. Familiarity and semantic decomposability were the fixed effects, subjects and items were the random intercepts, and familiarity and semantic decomposability were the random slopes by subject and item. The results show that neither familiarity nor semantic decomposability were
significant predictors (familiarity: $t = .47; p = .64$; effect size: .03; semantic decomposability: $t = -.97; p = .33$; effect size: -.03). In addition, none of the random intercepts and random slopes were significant. In a following model, an interaction term between familiarity and semantic decomposability was added as a new fixed effect to test whether it was a significant factor accounting for the RT difference.

4.4.2.5 English-speaking group: Three-predictor model

An interaction term for familiarity and semantic decomposability was added to this model. Familiarity, semantic decomposability, and the interaction between familiarity and semantic decomposability were added as the fixed effects. Table 16 shows that after the interaction term was added, familiarity, semantic decomposability, and the interaction term were significant predictors (familiarity: $t = -2.08; p = .04$; effect size = -.19; semantic decomposability: $t = -2.52; p = .01$; effect size = -.24; interaction term: $t = 2.33; p = .02$; effect size = .34). As the degrees of familiarity and semantic decomposability increased by one unit, the RT difference between the two context types ($RT_{\text{prime}}-RT_{\text{neu}}$) decreased by .05 log ms and .06 log ms, respectively (log transformation). Because the RTs in the priming contexts were shorter than the RTs in the neutral contexts (RQ1), the RTs in the neutral contexts should decrease greatly, while the RTs in the priming contexts should not, so that the RT difference would grow smaller as the degrees of familiarity and decomposability increased. On the contrary, as the degrees of the interaction between familiarity and semantic decomposability increased by one unit, the RT difference between the two context types decreased by .01 log ms (log transformation). More details on the interpretation of the interaction effect are discussed in section 4.4.3.
Table 16. English-speaking Group: Three-predictor Model

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>b</th>
<th>95% Confidence Interval Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect (γ0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>741.30</td>
<td>1.83</td>
<td>0.07</td>
<td>0.19</td>
<td>-0.01</td>
<td>0.39</td>
</tr>
<tr>
<td>Familiarity</td>
<td>885.98</td>
<td>-2.08</td>
<td>0.04</td>
<td>-0.05</td>
<td>-0.10</td>
<td>-0.003</td>
</tr>
<tr>
<td>Semantic Decomposability</td>
<td>902.87</td>
<td>-2.52</td>
<td>0.01</td>
<td>-0.06</td>
<td>-0.24</td>
<td>-0.01</td>
</tr>
<tr>
<td>Interaction</td>
<td>915.23</td>
<td>2.33</td>
<td>0.02</td>
<td>0.01</td>
<td>0.34</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Variance</th>
<th>Wald Z</th>
<th>p</th>
<th>95% Confidence Interval Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>0.002</td>
<td>1.90</td>
<td>0.06</td>
<td>0.001</td>
<td>0.01</td>
</tr>
<tr>
<td>Item</td>
<td>0.06</td>
<td>21.14</td>
<td>&lt; 0.001</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Residual (εsi)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following discussion examines the two models used for the Chinese-speaking group’s data.

4.4.2.6 Chinese-speaking group: Three-predictor model

Familiarity, semantic decomposability, and metaphoric equivalence were the three predictors added to the model. In addition, subjects and items were random intercepts. The slopes of familiarity, semantic decomposability, and metaphoric equivalence were random slopes by subject and item. Table 17 shows that semantic decomposability and metaphoric equivalence were two significant predictors accounting for the NNS’s RT difference in the priming and neutral contexts (semantic decomposability: $t = 2.13; p = .03$; effect size = .07; metaphoric equivalence: $t = -2.29; p = .02$; effect size = -.14). Item as random intercept was significant ($p = .01$).
Table 17. Chinese-speaking Group: Three-predictor Model

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>b</th>
<th>B</th>
<th>95% Confidence Interval</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Fixed effect ($\gamma_0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>101.26</td>
<td>-1.26</td>
<td>0.21</td>
<td>-0.06</td>
<td>-0.15</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>71.26</td>
<td>0.51</td>
<td>0.61</td>
<td>0.004</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Semantic Decomposability</td>
<td>801.97</td>
<td>2.13</td>
<td>0.03</td>
<td>0.02</td>
<td>0.07</td>
<td>0.001</td>
<td>0.04</td>
</tr>
<tr>
<td>Metaphoric Equivalence</td>
<td>647.95</td>
<td>-2.29</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.14</td>
<td>-0.04</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Variance</th>
<th>Wald Z</th>
<th>p</th>
<th>95% Confidence Interval</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Random effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>0.01</td>
<td>2.58</td>
<td>0.01</td>
<td>0.004</td>
<td>0.02</td>
</tr>
<tr>
<td>Item</td>
<td>0.07</td>
<td>19.78</td>
<td>&lt;0.001</td>
<td>0.06</td>
<td>0.07</td>
</tr>
</tbody>
</table>

As the degree of semantic decomposability increased by one unit, the RT difference increased by .02 log ms. Because the RTs in the priming contexts were shorter than those in the neutral contexts in general (RQ1), the RTs in the priming contexts should decrease more than the RTs in the neutral contexts, so that the RT difference between the two context types increased as the degree of semantic decomposability increased. In addition, as the degree of metaphoric equivalence increased by one unit, the RT difference decreased by .02 log ms. Because the RTs in the priming contexts were shorter than those in the neutral contexts in general (RQ1), the RTs in the neutral contexts should decrease more than the RTs in the priming contexts, so that the RT difference between the two context types decreased as the degree of metaphoric equivalence increased.
4.4.2.7 Chinese-speaking group: Six-predictor model

In addition to the three predictors mentioned in the three-predictor model (familiarity, semantic decomposability, and metaphoric equivalence), three interaction terms were added to the model. The first interaction term was the interaction between familiarity and semantic decomposability. The second interaction term was the interaction between familiarity and metaphoric equivalence. The third interaction term was the interaction between semantic decomposability and metaphoric equivalence. Item as random intercept was included in the model, too. The results show that none of the predictors of the fixed effects were significant (familiarity: $t = -0.71; p = 0.48$; effect size = -0.09; semantic decomposability: $t = 0.40; p = 0.69$; effect size = 0.04; metaphor: $t = -1.93; p = 0.054$; effect size = -0.25; interaction 1: $t = 0.99; p = 0.37$; effect size = 0.02; interaction 2: $t = 1.10; p = 0.27$; effect size = 0.19; interaction 3: $t = 0.17; p = 0.86$; effect size = 0.03).

To sum up, familiarity, semantic decomposability, and the interaction between the degrees of familiarity and semantic decomposability played an important role in idiom processing for the native speakers of English. The findings of a main effect for familiarity and an interaction effect between familiarity and semantic decomposability support my predictions. However, the present study also found a main effect for semantic decomposability, which did not support my prediction. Further, semantic decomposability and metaphoric equivalence played an important role in idiom processing for the native speakers of Chinese. This finding supports my prediction.

4.4.3 Discussion

The present study found that the NSs were highly familiar with idioms (mean =4.31 out of 5), which can explain why familiarity was found to be an important factor affecting NS idiom processing. Furthermore, the dual idiom representation model suggests that semantic decomposability can come into play when people are unfamiliar with idioms. However, in the present study, the NSs were found to be familiar with the idioms, so they should not have to rely on other resources, such as semantic decomposability, to facilitate their understanding of the idioms. However, the result shows that semantic decomposability was a good predictor. A possible explanation is that familiarity and semantic decomposability are positively correlated ($r = 0.85$). Therefore, when NSs perceived an idiom to be familiar, they also perceived the idioms to be semantically decomposable. Table 18 below summarizes the significant predictors accounting for RT differences in the two language groups.
Table 18. Summary of Significant Predictors in the Two Language Groups

<table>
<thead>
<tr>
<th>Significant Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>English-speaking</td>
</tr>
<tr>
<td>1. Familiarity</td>
</tr>
<tr>
<td>2. Semantic decomposability</td>
</tr>
<tr>
<td>3. Interaction between familiarity and semantic decomposability</td>
</tr>
<tr>
<td>Chinese-speaking</td>
</tr>
<tr>
<td>1. Semantic decomposability</td>
</tr>
<tr>
<td>2. Metaphoric equivalence</td>
</tr>
</tbody>
</table>

The results for the English-speaking participants indicate that the interaction between familiarity and semantic decomposability played an important role in the reaction time difference between priming and neutral contexts. This means that the effect of semantic decomposability on RT differences is different at different degrees of familiarity. As Table 16 above shows, when the interaction between familiarity and semantic decomposability increases by one unit, the NS RT difference increases by 0.01 log ms (log transformation). To understand how the increase of familiarity and semantic decomposability and the interaction between familiarity and semantic decomposability affect NS RT differences, we can analyze the effect of semantic decomposability on the reaction time difference when degree of familiarity varies. Equation (8) can be re-written as follows:

\[ Y_{si} = (0.19) + (-0.05)(\text{familiarity}) + (-0.06)(\text{semantic decomposability}) \\
+ (0.01)(\text{familiarity})(\text{semantic decomposability}) + 0.002 + 0.06 \]

The effect of semantic decomposability is represented by everything that is multiplied by semantic decomposability in the model: \((-0.06)(\text{semantic decomposability}) + (0.01)(\text{familiarity})(\text{semantic decomposability})\).

Table 19 below illustrates the effect of semantic decomposability on the reaction time difference. When familiarity = 1 (i.e., participants had not seen the idiom before), as semantic decomposability increases by one unit, the reaction time difference between higher semantic decomposability (e.g., semantic decomposability = 2) and lower semantic decomposability (e.g., semantic decomposability = 1) decreases by 0.05 log ms (i.e., \((-0.06) * 1 + 0.01 * 1 * 1 = -0.05\)). When familiarity = 2 (i.e., participants had seen the idiom before, but do not know
what it means), as semantic decomposability increases by one unit, the reaction time
difference between higher semantic decomposability (e.g., semantic decomposability = 2)
and lower semantic decomposability (e.g., semantic decomposability = 1) decreases by 0.04
log ms (i.e., (-0.06) * 1 + 0.01 * 2 * 1 = -0.04). As shown in the rightmost column of Table
19 (i.e., Effect), the reaction time difference between priming and neutral contexts decreases
as degree of familiarity and degree of semantic decomposability increase.

Table 19. Effect of Semantic Decomposability on the Reaction Time Difference
in the English-Speaking Group

<table>
<thead>
<tr>
<th>Familiarity</th>
<th>Semantic decomposability</th>
<th>Interaction</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.06</td>
<td>0.01</td>
<td>-0.05</td>
</tr>
<tr>
<td>2</td>
<td>-0.06</td>
<td>0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>3</td>
<td>-0.06</td>
<td>0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>4</td>
<td>-0.06</td>
<td>0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>5</td>
<td>-0.06</td>
<td>0.05</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

The decrease of the RT difference (i.e., $RT_{\text{prime}} - RT_{\text{neu}}$) is very likely due to the
possibility that the RTs in both context types decreased as the degrees of familiarity and
decomposability increased. According to Abel (2003), it is easier to form an idiom entry for a
high-familiarity, semantically decomposable idiom than for a low-familiarity, semantically
decomposable idiom. The formation of an idiom entry facilitates idiom processing.
Therefore, in the present study, the increase in the degree of familiarity and the degree of
semantic decomposability should have helped reduce the processing time. However, the RTs
in the neutral contexts decreased greatly in comparison with the RTs in the priming contexts;
as a result, the RT difference in the two context types decreased.

In answering RQ1, we learned that the participants processed idioms significantly
faster in the priming contexts than in the neutral contexts in general. However, it is possible
for participants to reduce their processing time of idioms in the neutral contexts. We have
further learned that the participants could process idioms significantly faster in the neutral
contexts when they were more familiar with the idioms and conceived the idioms to be more
semantically decomposable. However, being more familiar with the idioms and conceiving of
them as more semantically decomposable in the priming contexts did not change the participants’ RTs much.

The results for the Chinese-speaking participants indicate that semantic decomposability and metaphoric equivalence played an important role in the reaction time difference between priming and neutral contexts, while familiarity was not good predictors of the reaction time difference. According to the dual idiom representation model, when people are unfamiliar with an idiom, they attempt to semantically decompose it in order to find out the metaphoric referents. Moreover, based on the modified hierarchical model, the distance between L1 and L2 concepts will affect the reaction time difference. Similarly, the distance between L1 and L2 metaphors will affect the reaction time of idiom comprehension. In the current study, the NNSs’ average familiarity rating of the idioms was 3.08, meaning that they had not seen most of the idioms before, but they could roughly guess their meanings. Because the participants were unfamiliar with most of the idioms, it was more likely that semantic decomposability and metaphoric equivalence would play a role in their idiom processing. The study found that both semantic decomposability and metaphoric equivalence did play a significant role in the reaction time difference between the two context types. These results suggest that for NNSs, both the L1 influence and the linguistic feature of semantic decomposability are important in understanding L2 idioms.

Although semantic decomposability and metaphoric equivalence were good predictors of reaction time difference for the NNSs, metaphoric equivalence did not interact with semantic decomposability. My original prediction was that as the degrees of semantic decomposability and metaphoric equivalence increase, the reaction time difference will decrease. However, this was not the case in this study. This exploratory analysis has demonstrated that semantic decomposability and metaphoric equivalence between L1 and L2 accounted for most of the variance in the reaction time differences between the two context types.
CHAPTER 5
CONCLUSION

This dissertation has compared how classroom-taught English language learners and native speakers of English understand English idioms, and investigated how English idioms are represented in the mental lexicon of both groups. This chapter begins by summarizing the tasks that the participants completed and the findings from the analysis of the tasks’ results. The findings are then discussed in terms of idiom processing models and L2 idiom learning and teaching. The chapter concludes by explaining the study’s limitations and contributions, and suggesting directions for future research.

5.1 Summary of tasks and findings

This dissertation study consisted of two main parts: an experiment employing a self-paced reading task, and a set of surveys. The self-paced reading task was used to measure native English speakers’ and non-native English speakers’ response times when reading English passages containing idioms. The surveys included one in which the participants rated the familiarity of the idioms; another in which they rated the semantic decomposability of the idioms; and a third, which only the NNSs completed, in which they rated the metaphoric equivalence of the idioms used in this study and similar idioms in their first language, Chinese.

The study found that all of the participants processed the idioms significantly faster in priming contexts than in neutral contexts. The context effect on participants’ reaction times was equally strong for the English-speaking and the Chinese-speaking participants; that is, no interaction effect was found between context and language.

The study also found that both the English-speaking and the Chinese-speaking participants tended to conceive of the idioms in the study as semantically decomposable, which was different from previous findings. The previous studies have suggested that NSs tend to conceive of idioms as semantically nondecomposable, while NNSs tend to conceive of them as semantically decomposable. The difference could be due to the tasks that were used to sort idioms into different types of semantic decomposability. However, the English-
speaking participants in this study showed a stronger tendency to consider idioms as semantically decomposable in comparison with the Chinese-speaking participants.

According to the results of the analyses conducted to address RQ1, all of the participants processed idioms in the priming contexts significantly faster than those in the neutral contexts, and all of the participants conceived of the idioms as semantically decomposable. The results of the analyses that addressed RQ4 and RQ5 demonstrated that the English-speaking participants relied on the interaction of familiarity and semantic decomposability to understand the idioms. The more familiar and semantically decomposable an idiom was, the smaller the reaction time difference between the priming and neutral contexts for the English-speaking participants. However, for the Chinese-speaking participants, semantic decomposability and metaphoric equivalence between L1 and L2 played an important role in idiom comprehension. The more semantically decomposable and equivalent the metaphor was between L1 and L2, the less time the Chinese-speaking participants needed in both types of contexts in general. The difference between semantic decomposability and metaphoric equivalence is that the RT difference increased as the degree of semantic decomposability increased, while the RT difference decreased as the degree of metaphoric equivalence increased.

5.2 Implications of the findings

The findings in this study contribute to our understanding of the connections between context, familiarity, semantic decomposability, and metaphoric equivalence in idiom processing. Furthermore, the findings provide pedagogical implications on L2 idiom learning.

The findings for RQ1 and RQ2 show that degree of familiarity affects how easily people understand idioms in priming contexts and neutral contexts. According to the graded salience hypothesis, the salient meanings in the mental lexicon are always activated first. The salience of meanings is related to familiarity. The Chinese-speaking participants’ familiarity ratings of idioms were significantly lower than those of the English-speaking participants, and the Chinese-speaking participants’ RTs in both priming and neutral contexts were significantly higher than the RTs of the English-speaking participants. These results suggest that the salient meanings of the English idioms for the NNSs were the literal meanings, which is consistent with the literal salience hypothesis. However, to illustrate how native speakers of English and native speakers of Chinese understand idioms in both types of contexts, it is useful to apply the viewpoint of the graded salience hypothesis of language processing. The
present study’s findings suggest that different models of idiom processing might apply depending on how much context is available, which would explain the differences in processing speed in the priming and neutral contexts. Specifically, in the priming contexts, both the native speakers of English and the native speakers of Chinese may have activated the figurative meanings of the idioms as soon as possible. However, in the neutral contexts, both native speakers of English and native speakers of Chinese might have had access to both figurative meanings and literal meanings at the very beginning of processing, only abandoning the inappropriate meanings as time went by and more information became available; both situations correspond to the graded salience hypothesis (Giora, 1997).

Furthermore, the results imply that classroom-taught foreign language learners—in this study, native speakers of Chinese—are, like native speakers of English, able to utilize contextual clues to understand English idioms. In other words, L2 learners can utilize context to learn idioms, instead of learning by rote. Moreover, priming contexts can be more helpful in idiom learning than neutral contexts, especially for beginning learners, because the priming contexts provide richer contextual clues. A possible explanation, which comes from cognitive semantics, is that the rich contextual clues help L2 learners who are unfamiliar with L2 idioms to apply their world knowledge or their knowledge of metaphors to understand the L2 idioms (Kövecses & Szabó, 1996).

The findings for RQ3 show that the non-native speakers of English conceived of the English idioms as semantically decomposable. Because they were unfamiliar with most of the idioms, it is very likely that the idioms were stored and retrieved as constituent entries in their mental lexicon. By extension, it is also likely that they tended to make use of the literal meanings of the constituents of the idioms and link the literal meanings to the metaphoric referents in order to understand the idioms (RQ4). The degree of ease of linking constituents to their metaphoric referents in an L2 depends on how similar the metaphor is in L1 and L2. Boers (2000a) found that guiding L2 learners to pay attention to the literal meanings of figurative expressions instead of only teaching them the figurative meanings can enhance the learners’ metaphoric awareness and improve their learning of figurative expressions. According to the modified hierarchical model, the links between L2 words and L1-specific concepts are weak, but the links between L2 words and L1–L2 shared concepts are strong. In other words, the more similarity between the L1 and L2 metaphorical concepts, the less
difficulty L2 learners have when linking constituent meanings to metaphoric referents because the links between constituents and metaphorical concepts are strong.

In L2 idiom teaching, language teachers can make use of learners’ L1 schemata, and help them find similarities between L1 and L2 metaphorical concepts. According to Irujo (1986), L2 learners have no problem comprehending and producing L2 idioms when the L1 and L2 idioms are the same. This shows positive L1 transfer (Pavlenko, 2008, 2009). In other words, being aware that L1 metaphorical concepts can be transferred can facilitate classroom-taught language learners’ comprehension and production of L2 idioms. Moreover, when L2 idioms partially share metaphorical concepts with L1 idioms, L2 learners have to reconstruct their concepts to distinguish the subtle differences between the two languages (Pavlenko, 2008, 2009). According to Irujo (1986), when L1 idioms and L2 idioms are similar but not identical, L2 learners have difficulty comprehending the L2 idioms because they tend to overgeneralize their L1 knowledge. Therefore, when language teachers teach this kind of idiom, they can explicitly guide students to find the source domain (i.e., the physiological effects associated with abstract emotions and feelings) and the target domain (i.e., the abstract emotions and feelings) in L1 and L2. According to Yu (1995), there are patterns in English and Chinese when people link certain source domains to specific target domains. For example, English uses “fluid” for “anger,” while Chinese uses “gas” for “anger.” By being aware of the subtle differences in the source and target domains between L1 and L2, L2 learners can avoid overgeneralizing their L1 idiomatic knowledge to the understanding of L2 idioms. When the L2 idioms use L2-specific metaphors, the L2 learners have to develop new L2 metaphorical concepts, just like they do when they learn an L2 word that has an L2-specific concept (Pavlenko, 2008, 2009). The development of new L2 metaphorical concepts can be challenging because their L1 schemata are active in the L2 learners’ minds, and they have to make additional efforts to develop and then strengthen the connections between L2 idioms and L2-specific metaphorical concepts (Cook, 1994).

Boers (2000b) found that using cognitive effort to identify source domains and target domains and categorize them into different conceptual mapping patterns (e.g., anger as fire) can facilitate deep-level cognitive processing (Craik & Lockhart, 1972), which helps learners to remember the L2 idioms. After learners develop the connection between form and meaning in their minds, they also need to be repeatedly exposed to the L2 idioms in priming contexts. The repeated exposure can help strengthen the connection and build an idiom entry for each
new L2 idiom in the long run. Once the L2 learners have idiom entries for L2 idioms, they are able to retrieve the idioms as soon as possible. The more idiom entries L2 learners have, the more native-like their language proficiency is.

5.3 Limitations and future research questions

This study has small effect sizes for the significant predictors found in the analyses that addressed RQ4 and RQ5. The sample size of the present study is 76. It would be worth investigating whether the effect size would be larger if the sample size increased. In addition, session 2 of the experiment was held one week after session 1. It is possible that when participants completed their second session of the self-paced reading task, they still remembered what they had read in the first session, which might have affected their reaction times. Therefore, it would be worth investigating whether the analyses addressing RQ4 and RQ5 would obtain the same results if the interval between session 1 and session 2 were longer, such as one month instead of one week. Another limitation is that the numbers of factors entered into the mixed-effects modeling were not the same for the NS and NNS data because the English-speaking participants did not take the metaphoric equivalence survey. In future research, the native English speakers could be L2 Chinese learners who would also take the metaphoric equivalence survey, which would allow the use of the same factors for the mixed effects modeling for both language groups. It would be interesting to know how familiarity, semantic decomposability, and metaphoric equivalence might affect both English language learners and Chinese language learners, and how the three factors would interact in affecting RT differences between priming and neutral contexts. Moreover, the present study adapted the vocabulary knowledge scale by adding a category (3 points) to allow participants to guess the meaning of idioms. A further analysis is recommended to examine whether each of the five categories captures sufficient information by means of item response theory (IRT).

5.4 Significance and contribution of this study

None of the previous idiom processing models takes both L1 and L2 language users into account. The idiom list hypothesis, the lexical representation hypothesis, and the configuration model focus on L1 idiom comprehension. The literal salience hypothesis focuses on L2 idiom comprehension. However, the present study uses the dual idiom representation model as a framework to explain how native speakers of English and native speakers of Chinese comprehend English idioms. In the dual idiom representation model, familiarity and semantic decomposability are the two most important factors that are
commonly mentioned in other idiom-related studies (Cieślicka, 2006; Conklin & Schmitt, 2008; Siyanova et al., 2011; Titone & Connine, 1994; Underwood et al., 2004). However, not all studies that investigate both L1 and L2 idiom comprehension measure L1 and L2 language users’ idiom familiarity. As a result, these studies do not tell us whether the reaction times they measure are for figurative meanings or literal meanings, because the salient meanings of idioms are related to the degree of the idioms’ familiarity. To address this limitation, the present study measured the familiarity of the idioms to both L1 and L2 language users, and thus was able to investigate whether familiarity is a good predictor for NS and NNS idiom processing.

Moreover, the dual idiom representation model is not only useful for investigating idiom processing. By using the dual idiom representation model, the present study also connects with language exposure research. For example, Ellis et al. (2008) indicated that the long-term exposure that L1 language users have to formulaic expressions results in strong form-meaning-function links in their minds; these links are the important characteristic of what Abel (2003) calls “idiom entries” in the mental lexicon. The findings of the present study also suggest that both NSs and NNSs tend to conceive of idioms as semantically decomposable. Regarding NSs, these findings differ from those of previous research. According to the dual idiom representation model, native speakers are highly familiar with idioms and conceive of them as decomposable, which makes it easy for them to form idiom entries. In other words, conceiving of idioms as decomposable, rather than non-decomposable, does not necessarily slow down processing time. It is believed that when idioms are conceived of as non-decomposable, they are processed as a whole. In fact, as long as language users have “idiom entries” in their mental lexicon, they are able to process idioms as a whole and enjoy processing benefits.

To make the dual idiom representation model a better model of L2 idiom processing, the present study utilizes the modified hierarchical model, a bilingual mental lexicon model, to illustrate the relationships between L1 and L2 idioms and differentiated metaphoric concepts (i.e., shared concepts, L1-specific concepts, and L2-specific concepts). As predicted, the study found that metaphoric equivalence between L1 and L2 played an important role in L2 idiom processing. This result offers pedagogical implications on L2 idiom learning, as mentioned previously.
Furthermore, the present study used the vocabulary knowledge scale as a framework to measure participants’ familiarity with idioms, modifying it by adding one more scale option, which allows L2 learners to report that they guess the meanings of idioms. By including this option, the current study provides information about L2 idiom familiarity that previous studies did not. The modified scale has good validity and reliability, which means that future researchers can make use of it to measure familiarity of idioms as well. Last but not least, to my knowledge, none of the previous research has measured metaphoric equivalence between L1 and L2. While its validity and reliability need further research, the metaphoric equivalence survey created and used in this study may be the first of its kind.

APPENDIX A

Idioms

praise something to the skies
cover someone’s tracks
split the difference
kill two birds with one stone
run something into the ground
force one’s hand
swallow one’s pride
hit below the belt
lend someone a hand
make no bones about
seize the day
fan the flames
kick up one’s heels
call a meeting
burn the midnight oil
by word of mouth
play with fire
break the bank
climb the walls
follow the crowd
cool one’s heels
break the ice
pay through the nose
strike a bargain

APPENDIX B

Language background survey and modified C-test

Thank you for taking the time to complete this survey. This survey consists of two sections. Section 1 is about your language background, and Section 2 is about your language proficiency. This survey should only take about 15 minutes of your time. Your answers will be completely anonymous. Each question requires an answer in order to progress through the survey. If you have any questions about the survey, please tell the experimenter at any point during the process of completing the survey.

Section 1: Language Background. Please answer in English.

1. What is your language identity?
   - [ ] Chinese-speaking English Learners
   - [ ] Native speakers of English

2. What was(were) the first language(s) you learned before age 6?
3. How many years have you learned the first language(s) as a subject in school (by years)?
Example: Chinese 10 years.
If you have never learned the language as a subject in school, please write 0 year.

4. Have you ever learned a second language before age 15?
   ○ Yes
   ○ No

5. What was(were) the second language(s) you learned before age 15?

6. How many years have you learned the second language(s) as a subject in school (by years)? Example: Chinese 10 years.
   If you have never learned the language as a subject in school, please write 0 year.

7. What is your gender?
   ○ Female
   ○ Male

8. What is your age?

9. What is your highest education level (degree you have achieved)?*
   ○ Senior high
   ○ Undergraduate
   ○ MA
   ○ PhD

Please complete the rest of the word.
Example: I went to the zoo yes____.
Answer: I went to the zoo yesterday.
Please pay attention to tense and spelling.

The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction for helping children make better choices when it comes to what they eat and drink. Childhood obe(1) has bec(2) a ser(3) problem in th(4) country a(5) children cons(6) more sugar-based fo(7) and sp(8) less ti(9) getting the nece(10) exercise. Many par(11) have quest(12) schools’ deci(13) to al(14) vending machines which disp(15) candy and so(16) drinks. Many schools, tho(17), have co(18) to re(19) on the mo(20) these machines generate through agreements with the companies which make soft drinks and junk food.
Thank you for taking the time to complete this survey. The survey consists of two sections. Section 1 aims to find out how familiar you are with selected English phrases, and Section 2 aims to find out to what extent you can understand the overall meaning of the English phrases from the individual words in the phrases. This survey is estimated to take 15 to 20 minutes of your time. Your answers will be completely anonymous. Each question must be answered before you can move onto the next question. Please tell the experimenter at any point if you have a question during the process of completing the survey.

Section 1: Familiarity. For each English phrase, there are five descriptions of degree of familiarity. Please select the description that best fits how familiar you were with the English phrase BEFORE you took part in the experiment.

1. How much are you familiar with “praise something to the skies”?
   (1) I haven’t seen this phrase before, and I don’t know what it means.
   (2) I have seen this phrase before, but I don’t know what it means.
   (3) I haven’t seen this phrase before, but I can roughly guess what it means.
   (4) I have seen this phrase before, and I think I know what it means.
   (5) I know this phrase.

2. How much are you familiar with “cover someone’s tracks”?
   (1) I haven’t seen this phrase before, and I don’t know what it means.
   (2) I have seen this phrase before, but I don’t know what it means.
   (3) I haven’t seen this phrase before, but I can roughly guess what it means.
   (4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

3. How much are you familiar with “split the difference”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

4. How much are you familiar with “kill two birds with one stone”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

5. How much are you familiar with “run something into the ground”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

6. How much are you familiar with “force one’s hand”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

7. How much are you familiar with “swallow one’s pride”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

8. How much are you familiar with “hit below the belt”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

9. How much are you familiar with “lend someone a hand”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

10. How much are you familiar with “make no bones about”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

11. How much are you familiar with “seize the day”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.
12. How much are you familiar with “fan the flames”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

13. How much are you familiar with “kick up one’s heels”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

14. How much are you familiar with “call a meeting”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

15. How much are you familiar with “burn the midnight oil”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

16. How much are you familiar with “by word of mouth”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

17. How much are you familiar with “play with fire”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

18. How much are you familiar with “break the bank”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

19. How much are you familiar with “climb the walls”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

20. How much are you familiar with “follow the crowd”?
(1) I haven’t seen this phrase before, and I don’t know what it means.
(2) I have seen this phrase before, but I don’t know what it means.
(3) I haven’t seen this phrase before, but I can roughly guess what it means.
(4) I have seen this phrase before, and I think I know what it means.
(5) I know this phrase.

21. How much are you familiar with “cool one’s heels”?
Please provide a definition for each of the English phrases below. You may answer in either Mandarin Chinese or English.
25. praise something to the skies
26. cover someone’s tracks

27. split the difference

28. kill two birds with one stone

29. run something into the ground
30. force one’s hand

31. swallow one’s pride

32. hit below the belt

33. lend someone a hand

34. make no bones about
35. seize the day

36. fan the flames

37. kick up one’s heels

38. call a meeting
39. burn the midnight oil

40. by word of mouth

41. play with fire

42. break the bank

43. climb the walls
44. follow the crowd

45. cool one’s heels

46. break the ice

47. pay through the nose
Section 2: Understanding the overall meaning of a phrase from the individual words. This section aims to find out to what extent you are able to get the overall meaning of the English phrase from the meaning of the individual words in the phrase.

The overall meaning is provided in parentheses. Please read each phrase and its overall meaning before deciding the degree to which you are able to get the overall meaning of the phrase from the meaning of the individual words.

括号中的解释为惯用语的整体语意。请您先看完每个惯用语后，再读括号中的解释。

并在 5 个选项中勾选您认为惯用语的整体语意可以从构成这个惯用语的个别单字的字意推测出来的程度。

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>praise something to the skies (to give something much praise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cover someone’s tracks (to conceal one’s past activities)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>split the difference (to accept only part of what was originally wanted)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kill two birds with one stone (to solve two problems at one time with a single action)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phrase</td>
<td>address</td>
<td>state</td>
<td>city</td>
<td>zip</td>
<td>phone</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
<td>------</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>run something into the ground (to treat something very badly)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>force one’s hand (to make someone do something sooner than they want to)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>swallow one’s pride (to forget one’s pride and accept something humiliating)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hit below the belt (to be insulting and unfair)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lend someone a hand (to give someone some help)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>make no bones about (to say clearly what you think although you may embarrass or offend someone)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>seize the day (to take advantage of an opportunity when offered)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fan the flames (to make a situation worse)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kick up one’s heels (to do things that you enjoy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>call a meeting (to ask that people assemble for a meeting)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>burn the midnight oil (to stay up working, especially studying, late at night)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>by word of mouth (by speaking rather than writing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>play with fire (to do something dangerous or risky)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>break the bank (to use up all one’s money)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
climb the walls (to be extremely nervous or upset)  
follow the crowd (to do what everyone else is doing)  
cool one’s heels (to wait or to be kept waiting)  
break the ice (to initiate social interchanges and conversation)  
pay through the nose (to pay too much for something)  
strike a bargain (to reach an agreement on a price or negotiation for something)

Thank You!

Thank you for taking our survey. Your response is very important to us. Should you have any questions or feedback, please contact the experimenter.

Metaphoric equivalence survey

本问卷想要了解英文惯用语所使用的概念和中文惯用语所使用的概念其相关程度为何。每一道题皆有一段简短的说明，解释英文惯用语所使用的概念为何，请务必读完此一说明再作答。请注意本问卷不是想了解中英文翻译的适切程度，而是想了解中英文惯用语其背后所各自使用的概念的相关程度。作答时间约需 10 分钟。
例子：read between the lines 比喻 to try to understand what is meant by something that is not written explicitly or openly。这个英文惯用语使用 read / between the lines 的概念来表达此一比喻。您觉得中文的「弦外之音」所使用的概念和英文的 read between the lines 所使用的概念，相关程度为何？

如何作答：英文用「读在行与行之间」(read / between the lines) 的概念来表示某些没有挑明说的事情。而上述例子中的「弦外之音」使用「弹出既有的弦以外的音符」的概念来表示某些没有挑明说的事情。所以当您作答时，请就「读在行与行之间」和「弹出既有的弦以外的音符」的这两个概念的相近程度来勾选。

1. praise something to the skies 比喻 to give something much praise。这个英文惯用语使用 praise / to the skies 的概念来表达此一比喻。您觉得中文的「将某事物捧上天」所使用的「捧上天」的概念和英文的 praise to the skies 所使用的「赞美到天边」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「捧上天」和「赞美到天边」這兩個概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. cover someone's tracks 比喻 to conceal one's past activities。这个英文惯用语使用cover / one's tracks 的概念来表达此一比喻。您觉得中文的「掩盖行踪」所使用的「掩藏足迹」的概念和英文的 cover one's tracks 所使用的「遮盖足迹」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「掩藏足迹」和「遮盖足迹」這兩個概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. split the difference 比喻 to accept only part of what was originally wanted。这个英文惯用语使用split / difference 的概念来表达此一比喻。您觉得中文的「各退一步」所使用的「彼此往后退」的概念和英文的 split the difference 所使用的「将彼此间的差异对半分开」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「彼此往后退」和「将彼此间的差异对半分开」這兩個</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. **kill two birds with one stone** 比喻 to solve two problems at one time。这个英文惯用语使用 two birds / one stone 的概念来表达此一比喻。您觉得中文的「一石二鸟」所使用的「用一颗石头击中两只鸟」概念和英文的 kill two birds with one stone 「用一颗石头杀死两只鸟」所使用的概念，相关程度为何?

<table>
<thead>
<tr>
<th>观念的相关度</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「用一颗石头击中两只鸟」和「用一颗石头杀死两只鸟」</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. **run something into the ground** 比喻 to treat something very badly。这个英文惯用语使用 run / into the ground 的概念来表达此一比喻。您觉得中文的「过度使用」所使用的「使用太频繁」概念和英文的 run something into the ground 所使用的「将某物往地里开下去」的概念，相关程度为何？

<table>
<thead>
<tr>
<th>观念的相关度</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
</table>
6. force someone's hand 比喻 to make someone do something they do not want to do。这个英文惯用语使用 force / one's hand 的概念来表达此一比喻。您觉得中文的「迫使某人就范」所使用的「强迫某人照你的意思做」概念和英文的 force someone's hand 所使用的「限制某人双手行动的自由」概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「强迫某人照你的意思做」和「限制某人双手行动的自由」這兩個概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. swallow one's pride 比喻 to accept something although it is humiliating。这个英文惯用语使用 swallow / one's pride 的概念来表达此一比喻。您觉得中文的「忍气吞声」所使用的「忍住怒气和吞下抱怨」概念和英文的 swallow one's pride 所使用的「吞下某人的自尊」概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. hit below the belt 比喻 If a remark is below the belt, it is very insulting and unfair。这个英文惯用语使用 hit / below the belt 的概念来表达此一比喻。您觉得中文的「暗箭伤人」所使用的「在暗处向某人射箭」概念和英文的 hit below the belt 所使用的「将球击向棒球打者腰部以下」概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「在暗处向某人射箭」和「将球击向棒球打者腰部以下」這兩個概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. lend someone a hand 比喻 to give someone some help。这个英文惯用语使用 lend / hand 的概念来表达此一比喻。您觉得中文的「助一臂之力」所使用的「用一只手臂的力气来协助别人」概念和英文的 lend a hand 所使用的「借某人一只手」概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
10. make no bones about 比喻 to say clearly what you think or feel about something。这个英文惯用语使用 make / no bones 的概念来表达此一比喻。您觉得中文的「直言不讳」所使用的「毫不隐藏的说出口」概念和英文的 make no bones about 所使用「在汤中没有任何骨头,所以喝起来没有困难」的概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「毫不隐藏的说出口」和「在汤中没有任何骨头,所以喝起来没有困难」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11. seize the day 比喻 to take advantage of an opportunity when offered。这个英文惯用语使用 seize / the day 的概念来表达此一比喻。您觉得中文的「把握良机」所使用「抓紧机会」的概念和英文的 seize the day 所使用「抓紧时间」的概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「抓紧时间」和「抓紧机会」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
「抓紧机会」和「抓紧时间」这两个概念的相关程度

12. fan the flames 比喻 to make a situation worse。这个英文惯用语使用 fan / flames 的概念来表达此一比喻。您觉得中文的「搧风点火」所使用「搧动风使得火势变大」的概念和英文的 fan the flames 所使用「将火焰搧得更旺」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「搧动风使得火势变大」和「将火焰搧得更旺」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

13. kick up one's heels 比喻 to do things that you enjoy。这个英文惯用语使用 kick / one's heels 的概念来表达此一比喻。您觉得中文的「放松一下」所使用「放松身体和精神」的概念和英文的 kick up one's heels 所使用「将脚跟往后腾」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「放松身体和精神」和「将脚跟往后腾」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
14. call a meeting 比喻 to ask that people assemble for a meeting。这个英文惯用语使用 call / meeting 的概念来表达此一比喻。您觉得中文的「召开会议」所使用「召集众人来讨论」的概念和英文的 call a meeting 所使用「召唤一个会议」的概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「召集众人来讨论」和「召唤一个会议」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. burn the midnight oil 比喻 to work very late into the night。这个英文惯用语使用 burn / midnight oil 的概念来表达此一比喻。您觉得中文的「挑灯夜战」所使用「到了晚上打着灯继续作战」的概念和英文的 burn the midnight oil 所使用「到了晚上点着油灯继续做事」的概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「到了晚上打着灯继续作战」和「到了晚上点着油灯继续做事」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
16. by word of mouth 比喻 if you hear information by word of mouth, you hear it from other people and not from the radio or television or from reading newspapers。這個英文慣用語使用 word of mouth 的概念來表達此一比喻。您覺得中文的「口耳相傳」所使用「透過口語將事情傳出去」的概念和英文的 by word of mouth 所使用的「透過嘴巴」概念，相關程度為何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「透過口語將事情傳出去」和「透過嘴巴」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17. play with fire 比喻 to do something dangerous or risky。這個英文慣用語使用 play / fire 的概念來表達此一比喻。您覺得中文的「玩火自焚」所使用「玩火燒到自己」的概念和英文的 play with fire 所使用「玩火」的概念，相關程度為何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「玩火燒到自己」和「玩火」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18. break the bank 比喻 to use up all one's money。这个英文惯用语使用 break / bank 的概念来表达此一比喻。您觉得中文的「倾家荡产」所使用「倾出所有家产」的概念和英文的 break the bank 所使用「打破钱库」的概念，相关程度为何?

165
19. **climb the walls** 比喻 to be extremely nervous or upset。这个英文惯用语使用 climb / walls 的概念来表达此一比喻。您觉得中文的「坐立不安」所使用「无法好好的坐着或站着」的概念和英文的 climb the walls 所使用「爬墙」的概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「倾出所有家产」和「打破钱库」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

20. **follow the crowd** 比喻 to do what everyone else is doing。这个英文惯用语使用 follow / crowd 的概念来表达此一比喻。您觉得中文的「人云亦云」所使用「别人说什么，自己就跟着说什么」的概念和英文的 follow the crowd 所使用「跟随群众」的概念，相关程度为何?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「无法好好的坐着或站着」和「爬墙」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
「别人说什么，自己就跟着说什么」
和「跟随群众」这两个概念的相关程度

21. cool one's heels 比喻 to wait or to be kept waiting。这个英文惯用语使用 cool / one's heels 的概念来表达此一比喻。您觉得中文的「苦苦等候」所使用「花很长时间等待」的概念和英文的 cool one's heels 所使用「让某人的脚跟先冷却一下」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「花很长时间等待」和「让某人的脚跟先冷却一下」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

22. break the ice 比喻 to attempt to become friends with someone。这个英文惯用语使用 break / ice 的概念来表达此一比喻。您觉得中文的「破冰」所使用「打破冰冷的局面」的概念和英文的 break the ice 所使用「将冰块打破」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
23. pay through the nose 比喻 to pay too much for something。这个英文惯用语使用 pay / through the nose 的概念来表达此一比喻。您觉得中文的「被敲竹杠」所使用「查鸦片的官员在船杆上轻弹烟蒂,但走私鸦片的商人以为官员发现私运鸦片而赶紧塞钱行贿官员」的概念和英文的 pay through the nose 所使用「外来者占领土地后以数鼻子的方式查当地居民的人数,并要求以鼻子个数来付税,否则就割下鼻子」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>「塞钱行贿官员」和「以鼻子个数来付税」这两个概念的相关程度</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24. strike a bargain 比喻 to reach an agreement on negotiations。这个英文惯用语使用 strike / bargain 的概念来表达此一比喻。您觉得中文的「达成协议」所使用「双方商量后一致同意」的概念和英文的 strike a bargain 所使用「双方手牵手表示同意某事」的概念，相关程度为何？

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>A lot</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
「双方商量后一致同意」和「双方手牵手表示同意某事」这两个概念的相关程度

Thank you for taking our survey. Your response is very important to us. Should you have any questions or feedback, please contact the experimenter.
APPENDIX D
Idiom contexts

List 1

1. Alice baked a cake. Her boyfriend was eager to taste it. He kept telling her the cake was delicious. He praised the cake to the skies. Alice was very touched. Did Alice’s boyfriend hate the cake? N

2. John found out his wife had been dating another man. His wife was cheating on him. She tried to cover her tracks. John decided to divorce her. Did John’s wife try to hide her love affair? Y

3. Lucy and her friends were planning to travel in Europe. Her friends had many conflicting ideas about this trip. She hoped her friends could accept each other’s ideas. She made them split the difference. She hoped everyone was happy. Did Lucy try to find a solution to the problem? Y

4. Katie has an appointment with her dentist this morning. She also has a payment due at the bank this afternoon. She can kill two birds with one stone. She always plans ahead. Will Katie see her dentist and make a payment today? Y

5. Tom bought a new car. He loaned the car to his sister for four months. When his sister returned the car, the tires were worn out. She ran the car into the ground. He was very angry. Was Tom’s car used so badly? Y

6. Jenna asked Mike to raise his offer on her house, but he refused. Then she told him she wanted to pull out of the sale. She wanted to force his hand. In the end, Mike raised his offer.
Did Jenna force Mike to raise his offer? Y

7
Andy was late again, although he had a reason. His girlfriend scolded him for making excuses. He had to admit he was wrong. He had to swallow his pride. He hoped his girlfriend would forgive him.
Did Andy admit he was wrong? Y

8
Jessica was running for mayor. Her rivals brought up her husband’s alcoholism. She said that was hitting below the belt. She was very angry.
Did Jessica think her rivals’ comments were fair? N

9
Joan had a table in the living room. But she wanted to move it to her bedroom. She asked her brother for help. Her brother agreed to lend her a hand. Joan had the table moved finally.
Did Joan’s brother agree to help her? Y

10
Mark moved into his friend’s house. However, he found cockroaches in the bathroom. He made no bones about his dissatisfaction with the house. He was very disappointed.
Did Mark care his comment might be embarrassing? N

11
Patty was too poor to travel. She saw an advertisement about working in Australia during the summer. She signed up for the program. She decided to seize the day. She was very excited.
Did Patty give up the opportunity to go to Australia? N

12
Denny and Brian had an argument about how to run their restaurant. Denny’s wife didn’t like Brian. She fanned the flames of their argument. Denny decided to sue Brian.
Did Denny’s wife settle the dispute between Denny and Brian? N
13
Mary used to go to bars on Friday nights. However, she has been too busy to go to bars for years. One day, her friend asked her out to a bar to have some fun. Her friend told her she should kick up her heels a little. Mary agreed. Did Mary’s friend think Mary was too nervous in life? Y

14
Sam went bankrupt. He would have to close his factory. He called a meeting. He apologized to his employees. Did Sam gather his employees? Y

15
Sandy was not good at math. She had a math exam the next day. She decided to study all night. She had to burn the midnight oil. She wanted to get a good grade on the exam. Did Sandy give up on the exam? N

16
Tina said that a big department store was going to open in the neighborhood. Her father wondered where she got the news. She had learned about it by word of mouth. She was very excited. Did Tina hear the news on TV? N

17
Jack had lost a million dollars in the stock market. He decided to invest even more money. His wife said it was too dangerous. He was playing with fire. However, he did not care. Was Jack’s decision risky? Y

18
Kevin invited his girlfriend to a fancy restaurant. His girlfriend was worried that the restaurant was too expensive. He said it wouldn’t break the bank. His girlfriend was happy to hear that.
Did Kevin spend all his money at the restaurant? N

19
Lisa was supposed to pick up her child from school. Her boss did not let her leave work on
time. She was very worried. She was climbing the walls. She called the school.
Was Lisa very anxious? Y

20
Bill is a fashion designer. He is competing in a big show with many other designers. He does
not follow the crowd. He has confidence in himself.
Does Bill usually imitate other designers’ work? N

21
Daisy was late to a show. She was not allowed to enter. She was told to wait until the second
half began. She had to cool her heels outside. She was very upset.
Did Daisy decide to wait outside herself? N

22
Louis had his first date with Michelle tonight. He knew Michelle was a quiet person. He tried
hard to break the ice. He was exhausted after the date.
Did Louis try to start a conversation with Michelle? Y

23
Annie bought a new house. Her new house needed painting. She was introduced to a painter
who charged twice as much as other painters. She paid through the nose for that paint job.
She was angry when she found out.
Did Annie pay less than average for the paint job? N

24
Chris and his wife were arguing about where to go for lunch. He wanted to eat Japanese food,
but his wife wanted to eat Chinese food. He could not strike a bargain with her. He felt very
hungry.
Did Chris and his wife reach an agreement? N

filler 1
Sam did not like hiking, but his girlfriend did. She asked Sam to go hiking many times. However, Sam always turned her down. His girlfriend decided to keep asking until Sam agreed.
Did Sam like hiking? N

filler 2
Linda liked to eat chocolate and ice cream. She ate a lot when she was stressed out. Her doctor warned her that she would become overweight if she kept eating this way.
Did Linda hate chocolate? N

filler 3
Mary had a car accident. She stayed in the hospital for ten months. When the doctor told her that she could leave, she joked that after only ten months there, she did not expect to be out so soon.
Did Mary have a car accident? Y

filler 4
Ryan went to a concert yesterday. He saw his favorite singer there. That singer was his idol when he was a child. He was too excited to sleep well last night.
Did Ryan sleep well last night? N

filler 5
Mike liked swimming in the ocean. He had swum with hundreds of sharks. The sharks did not attack him. He said it was because he knew their language, so the sharks thought he was one of them.
Did Mike like swimming? Y

filler 6
Wendy was very good at badminton. She decided to enter a contest. However, she got sick for months and could not practice badminton as usual. She had nightmares in which she lost the game. Her father told her to take it easy.
Did Wendy sleep well at night? N

filler 7
Brian did not like to eat carrots. His girlfriend ground up carrots and secretly put them in his food. He was very upset when he found out. They had a serious fight and then broke up.
Did Brian’s girlfriend put carrots in his food? Y

filler 8
Antony liked watching movies very much. He had been planning to see the latest movie. However, it was pouring rain when he went out. He was the only person in the theater. He was happy to enjoy the movie all by himself.
Were there other people in the theater, too? N

filler 9
Sarah was from a rich family. She worked at a coffee shop, but sometimes skipped work. Her boss was very unhappy about this. He decided to cut her pay. Sarah never showed up again.
Did Sarah work for a coffee shop? Y

filler 10
Jonathon was a full-time novelist. He wrote many adventure stories. His stories were very engaging. Some people thought he must have experienced many adventures. However, he said that his good ideas were simply from playing online games.
Had Jonathon experienced many adventures? N

filler 11
Betty believed in romantic love. She went to bars to look for her soul mate. Her friends told her that it was ridiculous because she might end up with a beer belly before she found her Mr. Right.
Did Betty believe in romantic love? Y
Andy had a crush on a girl. He lost a lot of weight to make himself more attractive. However, the girl liked him better the way he was before. Andy regretted that he had not talked to her first.
Did the girl like what Andy looked like after he lost weight? N

Frank was ten years older than his wife. He and his mother-in-law had never met before. She was impressed by him. She saw how patient Frank was. She finally understood why her daughter loved this man.
Did the mother-in-law like Frank? Y

George wanted to be a director. He tried very hard to promote his work for years. But no production company liked his ideas. His wife comforted him by telling him to go for his dream. George felt very touched.
Did George promote his work to the production companies? Y

Janet went jogging in the park every morning. One day when she was jogging, a dog began barking and chasing her. She fell into a pond all of a sudden. When the dog was about to catch her, she woke up and realized she had had a nightmare.
Did Janet have a nightmare? Y

Nana wanted to open a cupcake shop. She was trying hundreds of recipes but could not find a satisfying one. Her brother suggested mixing sweet and spicy flavors. Nana thought it was a bold but interesting idea. She decided to give it a try.
Did Nana agree with her brother's idea? Y
Jay had been a street artist for years. He was famous for playing cards. He created thousands of card games. He eventually got invited to perform on TV. This was something he had never dreamed of.

Did Jay know little about card games? N

filler 18

Carson was a medical student. He went to Africa for an internship. He found that many villages did not have a doctor. People usually died if they got an infection. He decided to come back to Africa after he graduated.

Did Carson go to Latin America for an internship? N

filler 19

Helen liked to watch horror movies at night. She thought her boyfriend would like horror movies, too. Her boyfriend promised to come over and stay with her. However, her boyfriend was even more scared than Helen.

Did Helen hate horror movies? N

filler 20

Tiffany was scared of dogs because she had been bitten in her childhood. She went to see a doctor for help. The doctor suggested that she should raise a dog herself. The suggestion made her even more scared. She never came back.

Did the doctor suggest that Tiffany raise a dog? Y

filler 21

Nick used to be a baseball player. However, he was seriously injured in an accident. He could not play anymore, so he decided to retire. He had dreamed about opening a theme restaurant. Now he is a successful businessman.

Does Nick have a restaurant now? Y

filler 22
Jill wanted to buy the latest designer handbag. However, she could not afford it. She tried to borrow from her friends. None of them wanted to help her because she often forgot to pay money back.
Did Jill have money to buy a designer handbag? N

diller 23
Albert was an amateur photographer. He took pictures during a trip to an African village. The villagers were looking for a missing baby lion. They wanted to help the anxious mother tiger. Albert was very touched and decided to pass their stories on to other people.
Did the villagers kill the baby tiger? N

diller 24
Olivia was practicing how to drive. After she practiced several times, she decided to take the road test. She thought she was ready. However, she did not pass the test. She thought it was because the examiner was too charming to let her concentrate.
Was Olivia learning how to drive a car? Y

List 2
1
Alice baked a cake. Her boyfriend was eager to taste it. He praised the cake to the skies. Alice was very touched.
Did Alice’s boyfriend like the cake? Y

2
John found out his wife had been dating another man. His wife was cheating on him. She deleted all her text messages from the other man. She tried to cover her tracks. John decided to divorce her.
Did John’s wife keep her messages from the other man? N

3
Lucy and her friends were planning to travel in Europe. Her friends had many conflicting ideas about this trip. She made them split the difference. She hoped everyone was happy.
Did Lucy make her friends accept each other’s ideas? Y

4
Katie has an appointment with her dentist this morning. She also has a payment due at the bank this afternoon. She will go to the bank on the way to the doctor’s office. She can kill two birds with one stone. She always plans ahead.
Will Katie go to the bank before going to the dentist’s office? Y

5
Tom bought a new car. He loaned the car to his sister for four months. She ran the car into the ground. He was very angry.
Did Tom’s sister keep the car new? N

6
Jenna asked Mike to raise his offer on her house, but he refused. Then she told him she wanted to pull out of the sale. She believed she could push him into paying more. She wanted to force his hand. In the end, Mike raised his offer.
Was Mark forced to raise his offer? Y

7
Andy was late again, although he had a reason. His girlfriend scolded him for making excuses. He had to swallow his pride. He hoped his girlfriend would forgive him.
Did Andy argue with his girlfriend? N

8
Jessica was running for mayor. Her rivals brought up her husband’s alcoholism. She didn’t think that was right. She said that was hitting below the belt. She was very angry.
Did Jessica think her rivals should not mention her husband? Y
Joan had a table in the living room. But she wanted to move it to her bedroom. Her brother agreed to lend her a hand. Joan had the table moved finally.
Did Joan’s brother refuse to move the table? N

Mark moved into his friend’s house. However, he found cockroaches in the bathroom. He complained a lot and then moved out. He made no bones about his dissatisfaction with the house. He was very disappointed.
Did Mark complain to his friend without hesitation? Y

Patty was too poor to travel. She saw an advertisement about working in Australia during the summer. She decided to seize the day. She was very excited.
Did Patty decide to go work in Australia? Y

Denny and Brian had an argument about how to run their restaurant. Denny’s wife didn’t like Brian. She told Denny that Brian was stealing money. She fanned the flames of their argument. Denny decided to sue Brian.
Did Denny’s wife make the situation worse? Y

Mary used to go to bars on Friday nights. However, she has been too busy to go to bars for years. One day, her friend told her she should kick up her heels a little. Mary agreed.
Did Mary’s friend want Mary to enjoy herself? Y

Sam went bankrupt. He would have to close his factory. He decided to announce the decision to all his employees at the same time. He called a meeting. He apologized to his employees.
Did Sam want to expand his factory? N
Sandy was not good at math. She had a math exam the next day. She had to burn the midnight oil. She wanted to get a good grade on the exam.
Did Sandy go to bed early for the exam? N

16
Tina said that a big department store was going to open in the neighborhood. Her father wondered where she got the news. There was nothing in the newspapers about a department store. She had learned about it by word of mouth. She was very excited.
Was Tina told by other people about the new department store? Y

17
Jack had lost a million dollars in the stock market. He decided to invest even more money. He was playing with fire. However, he did not care.
Was Jack safe by investing more money in the stock market? N

18
Kevin invited his girlfriend to a fancy restaurant. His girlfriend was worried that the restaurant was too expensive. He promised the cost was reasonable. He said it wouldn’t break the bank. His girlfriend was happy to hear that.
Did Kevin have to spend all his money on the bill? N

19
Lisa was supposed to pick up her child from school. Her boss did not let her leave work on time. She was climbing the walls. She called the school.
Was Lisa calm? N

20
Bill is a fashion designer. He is competing in a big show with many other designers. He always thinks independently. He does not follow the crowd. He has confidence in himself.
Does Bill like copying other people’s work? N
Daisy was late to a show. She was not allowed to enter. She had to cool her heels outside. She was very upset.
Did Daisy have to wait outside? Y

22
Louis had his first date with Michelle tonight. He knew Michelle was a quiet person. So he told jokes and stories all evening. He tried hard to break the ice. He was exhausted after the date.
Did Louis keep quiet on his first date? N

23
Annie bought a new house. Her new house needed painting. She paid through the nose for that paint job. She was angry when she found out.
Did Annie pay too much money to have her house painted? Y

24
Chris and his wife were arguing about where to go for lunch. He wanted to eat Japanese food, but his wife wanted to eat Chinese food. He promised to take his wife to a Chinese restaurant tomorrow. His wife refused. He could not strike a bargain with her. He felt very hungry.
Did Chris’s wife accept his idea? N

filler 1
Sam did not like hiking, but his girlfriend did. She asked Sam to go hiking many times. However, Sam always turned her down. His girlfriend decided to keep asking until Sam agreed.
Did Sam like hiking? N

filler 2
Linda liked to eat chocolate and ice cream. She ate a lot when she was stressed out. Her doctor warned her that she would become overweight if she kept eating this way.
Did Linda hate chocolate? N
Mary had a car accident. She stayed in the hospital for ten months. When the doctor told her that she could leave, she joked that after only ten months there, she did not expect to be out so soon.

Did Mary have a car accident? Y

Ryan went to a concert yesterday. He saw his favorite singer there. That singer was his idol when he was a child. He was too excited to sleep well last night.

Did Ryan sleep well last night? N

Mike liked swimming in the ocean. He had swum with hundreds of sharks. The sharks did not attack him. He said it was because he knew their language, so the sharks thought he was one of them.

Did Mike like swimming? Y

Wendy was very good at badminton. She decided to enter a contest. However, she got sick for months and could not practice badminton as usual. She had nightmares in which she lost the game. Her father told her to take it easy.

Did Wendy sleep well at night? N

Brian did not like to eat carrots. His girlfriend ground up carrots and secretly put them in his food. He was very upset when he found out. They had a serious fight and then broke up.

Did the girlfriend put carrots in his food? Y

Antony liked watching movies very much. He had been planning to see the latest movie. However, it was pouring rain when he went out. He was the only person in the theater. He was happy to enjoy the movie all by himself.
Were there other people in the theater, too? N

filler 9
Sarah was from a rich family. She worked at a coffee shop, but sometimes skipped work. Her boss was very unhappy about this. He decided to cut her pay. Sarah never showed up again.
Did Sarah work for a coffee shop? Y

filler 10
Jonathon was a full-time novelist. He wrote many adventure stories. His stories were very engaging. Some people thought he must have experienced many adventures. However, he said that his good ideas were simply from playing online games.
Had Jonathon experienced many adventures? N

filler 11
Betty believed in romantic love. She went to bars to look for her soul mate. Her friends told her that it was ridiculous because she might end up with a beer belly before she found her Mr. Right.
Did Betty believe in romantic love? Y

filler 12
Andy had a crush on a girl. He lost a lot of weight to make himself more attractive. However, the girl liked him better the way he was before. Andy regretted that he had not talked to her first.
Did the girl like what Andy looked like after he lost weight? N

filler 13
Frank was ten years older than his wife. His mother-in-law had not liked him at first. But she knew her daughter had a bad temper and she saw how patient Frank was. She finally understood why her daughter loved this man.
Was Frank ten years younger than his wife? N

filler 14
George wanted to be a director. He had been working very hard and believed that he would succeed. His wife encouraged him to go for his dream. George felt very touched.

Did George want to be a teacher? N

filler 15

Janet went jogging in the park every morning. One day when she was jogging, a dog began barking and chasing her. She was very scared and kept running. She found there was no road in front of her. When the dog was about to catch her she woke up and realized she had had a nightmare.

Did Janet go swimming every morning? N

filler 16

Nana wanted to open a cupcake shop. She was trying hundreds of recipes and asked around for some ideas. Her friend told her to put garlic in a cupcake. Nana thought it was a bold but interesting idea. She decided to give it a try.

Did Nana hate garlic? N

filler 17

Jay had been a street artist for years. He played instruments and performed magic. He combined music and magic in interesting ways. He eventually got invited to perform on TV. This was something he had never dreamed of.

Had Jay dreamed about performing on TV? N

filler 18

Carson was a medical student. He went to Africa for an internship. He found that many cities did not have clean water. People got stomachaches after drinking dirty water. He decided to come back to Africa after he graduated.

Did every city in Africa have clean water? N

filler 19
Helen liked to watch horror movies at night. However, she usually got scared and could not go to sleep afterwards. She asked her boyfriend to come over and stay with her. However, her boyfriend was even more scared than Helen.
Did Helen like to watch horror movies? Y

filler 20
Tiffany was scared of dogs because she had seen a gang of dogs fighting. She went to see a doctor for help. The doctor suggested that she should play with dogs herself. The suggestion made her even more scared. She never came back.
Did the doctor suggest that Tiffany play with dogs? Y

filler 21
Nick used to be a baseball player. He had been the best pitcher of the year. However, he decided to retire at the summit of his career. He had dreamed about opening a theme restaurant. Now he is a successful businessman.
Is Nick a baseball player now? N

filler 22
Jill wanted to buy the latest designer handbag. She did not want to pay for it herself. She told her friends that she liked the handbag. None of them wanted to help her because she often forgot to pay money back.
Did Jill buy the designer handbag by herself? N

filler 23
Albert was an amateur photographer. He took pictures during a trip to an African village. The villagers were poor, but never complained. They were happy with everything they had. Albert was very touched and decided to pass their stories on to other people.
Did the villagers complain about their life? N

filler 24
Olivia was practicing how to drive. She drove well while practicing. She stopped practicing because she thought it was a waste of her time. She did not pass the test, but she thought it was because the examiner was too charming to let her concentrate.

Did Olivia pass the test? N
APPENDIX E

Distribution of familiarity and semantic decomposability ratings

NS Group

NNS Group
### APPENDIX F

Normality check

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Raw data</th>
<th>squared data</th>
<th>Log10 data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skewness</td>
<td>Kurtosis</td>
<td>Skewness</td>
</tr>
<tr>
<td>Idiom1f</td>
<td>-.581</td>
<td>-.358</td>
<td>-.1082</td>
</tr>
<tr>
<td>Idiom2f</td>
<td>-.852</td>
<td>-.533</td>
<td>-.1163</td>
</tr>
<tr>
<td>Idiom3f</td>
<td>.363</td>
<td>-.1225</td>
<td>.058</td>
</tr>
<tr>
<td>Idiom4f</td>
<td><strong>-2.190</strong></td>
<td><strong>4.210</strong></td>
<td><strong>-2.305</strong></td>
</tr>
<tr>
<td>Idiom5f</td>
<td>-.214</td>
<td><strong>-1.734</strong></td>
<td>-.347</td>
</tr>
<tr>
<td>Idiom6f</td>
<td>-.543</td>
<td>-.759</td>
<td>-.927</td>
</tr>
<tr>
<td>Idiom7f</td>
<td>-1.090</td>
<td>-.097</td>
<td>-1.380</td>
</tr>
<tr>
<td>Idiom8f</td>
<td>-.337</td>
<td><strong>-1.635</strong></td>
<td>-.477</td>
</tr>
<tr>
<td>Idiom9f</td>
<td><strong>-2.849</strong></td>
<td>7.537</td>
<td><strong>-2.962</strong></td>
</tr>
<tr>
<td>Idiom10f</td>
<td>.777</td>
<td>-.904</td>
<td>.566</td>
</tr>
<tr>
<td>Idiom11f</td>
<td>-.639</td>
<td>-.1278</td>
<td>-.815</td>
</tr>
<tr>
<td>Idiom12f</td>
<td>-.945</td>
<td>-.608</td>
<td>-1.169</td>
</tr>
<tr>
<td>Idiom13f</td>
<td>.043</td>
<td><strong>-1.646</strong></td>
<td>-1.142</td>
</tr>
<tr>
<td>Idiom14f</td>
<td>-.1360</td>
<td>.797</td>
<td><strong>-1.641</strong></td>
</tr>
<tr>
<td>Idiom15f</td>
<td>-.173</td>
<td>.368</td>
<td><strong>-1.547</strong></td>
</tr>
<tr>
<td>Idiom16f</td>
<td>-.1359</td>
<td>.347</td>
<td><strong>-1.556</strong></td>
</tr>
<tr>
<td>Idiom18f</td>
<td><strong>-1.834</strong></td>
<td>2.978</td>
<td><strong>-2.328</strong></td>
</tr>
<tr>
<td>Idiom19f</td>
<td>.159</td>
<td>-.1436</td>
<td>-.091</td>
</tr>
<tr>
<td>Idiom20f</td>
<td><strong>-1.771</strong></td>
<td>2.564</td>
<td><strong>-2.154</strong></td>
</tr>
<tr>
<td>Idiom21f</td>
<td>1.038</td>
<td>-.055</td>
<td>.692</td>
</tr>
<tr>
<td>Idiom22f</td>
<td><strong>-2.650</strong></td>
<td>6.560</td>
<td><strong>-2.868</strong></td>
</tr>
<tr>
<td>Idiom23f</td>
<td>.104</td>
<td><strong>-1.608</strong></td>
<td>-.083</td>
</tr>
<tr>
<td>Idiom24f</td>
<td>-.271</td>
<td><strong>-1.673</strong></td>
<td>-.432</td>
</tr>
<tr>
<td>Idiom1d</td>
<td><strong>-1.511</strong></td>
<td>1.244</td>
<td><strong>-1.622</strong></td>
</tr>
<tr>
<td>Idiom2d</td>
<td><strong>-1.748</strong></td>
<td>2.524</td>
<td><strong>-2.000</strong></td>
</tr>
<tr>
<td>Idiom3d</td>
<td>.014</td>
<td>-.813</td>
<td>-.404</td>
</tr>
<tr>
<td>Idiom4d</td>
<td><strong>-2.080</strong></td>
<td>4.219</td>
<td><strong>-2.346</strong></td>
</tr>
<tr>
<td>Idiom5d</td>
<td>-.199</td>
<td><strong>-1.338</strong></td>
<td>-.469</td>
</tr>
<tr>
<td>Idiom6d</td>
<td>-.370</td>
<td>-.763</td>
<td>-.738</td>
</tr>
<tr>
<td>Idiom7d</td>
<td>-.799</td>
<td>-.577</td>
<td>-.947</td>
</tr>
<tr>
<td>Idiom8d</td>
<td>-.512</td>
<td><strong>-1.019</strong></td>
<td>-.810</td>
</tr>
<tr>
<td>Idiom9d</td>
<td><strong>-2.957</strong></td>
<td>8.803</td>
<td><strong>-3.098</strong></td>
</tr>
<tr>
<td>Idiom10d</td>
<td>.441</td>
<td><strong>-1.196</strong></td>
<td>.172</td>
</tr>
<tr>
<td>Idiom11d</td>
<td>-.948</td>
<td><strong>-3.67</strong></td>
<td><strong>-1.237</strong></td>
</tr>
<tr>
<td>Idiom12d</td>
<td><strong>-1.581</strong></td>
<td>1.943</td>
<td><strong>-2.015</strong></td>
</tr>
<tr>
<td>Idiom13d</td>
<td>-0.064</td>
<td>-1.410</td>
<td>-0.325</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Idiom15d</td>
<td>-1.488</td>
<td>1.480</td>
<td>-1.720</td>
</tr>
<tr>
<td>Idiom16d</td>
<td>-1.299</td>
<td>0.885</td>
<td>-1.683</td>
</tr>
<tr>
<td>Idiom18d</td>
<td>-1.368</td>
<td>1.027</td>
<td>-1.560</td>
</tr>
<tr>
<td>Idiom19d</td>
<td>-0.363</td>
<td>-0.972</td>
<td>-0.714</td>
</tr>
<tr>
<td>Idiom20d</td>
<td>-2.497</td>
<td>5.843</td>
<td>-2.630</td>
</tr>
<tr>
<td>Idiom21d</td>
<td>-0.027</td>
<td>-0.899</td>
<td>-0.403</td>
</tr>
<tr>
<td>Idiom23d</td>
<td>-1.143</td>
<td>-1.514</td>
<td>-0.379</td>
</tr>
<tr>
<td>Idiom24d</td>
<td>-1.366</td>
<td>1.081</td>
<td>-1.738</td>
</tr>
</tbody>
</table>

*Note.* Idioms in the familiarity survey are marked with “f,” and idioms in the semantic decomposability survey are marked with “d.”

**Linearity check**

[Graph showing linearity check]
APPENDIX G

Reaction time distribution (before data transformation)

Priming Context:

Neutral Context:
REFERENCES


Raatz, U., & Klein-Braley, C. (2002). Introduction to language testing and to C-tests. In J. A. Coleman, R. Grotjahn, & U. Raatz (Eds.), *University language testing and the C-test* (pp. 75–91). Bochum, Germany: AKS-Verlag.


