The Pennsylvania State University
The Graduate School
Department of Psychology

LEXICAL SEMANTIC VARIATION AND CROSS-LANGUAGE INTERACTION IN CATEGORIZATION

A Dissertation in Psychology by Benjamin D. Zinszer

© 2014 Benjamin D. Zinszer

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

December 2014
The dissertation of Benjamin Zinszer was reviewed and approved* by the following:

Ping Li  
Professor of Psychology, Linguistics, & Information Sciences & Technology  
Dissertation Advisor  
Chair of Committee

Peter Molenaar  
Distinguished Professor of Human Development and Family Studies

Janet Van Hell  
Professor of Psychology and Linguistics

Daniel J. Weiss  
Associate Professor of Psychology and Linguistics

Melvin M. Mark  
Professor of Psychology  
Head of the Department of Department or Graduate Program

*Signatures are on file in the Graduate School
ABSTRACT

Language provides a finite set of labels (words) for an infinite set of possible objects that a speaker may encounter. Previous picture naming studies of language production have focused on highly familiar object stimuli that elicit uniform responses among native-speaker participants. By contrast, this dissertation explores picture naming responses to a broad range of object stimuli, from typical prototypes to unusual or unfamiliar examples of the same name. Drawing on research in lexical categorization, I document variation in native-speaker picture naming behavior and examine how this naming behavior and the underlying neural responses change as a result of language interaction in bilinguals.

Behavioral studies (Chapter 2) measure differences in lexical categorization patterns (picture naming responses over many different examples of an object name) among monolingual and bilingual speakers of Chinese and English. Further, this inter-personal variation is explained in terms of speakers' unique language histories and norms of each linguistic community (native speakers of English and Chinese). I propose a statistical model which describes the role of these variables in each language in predicting categorization patterns in both the native language (L1) and second language (L2), identifying significant effects of cross-language interaction for bilinguals in both languages.

Next, I introduce a new stimulus set of 407 objects sampled from several semantic domains (e.g., clothing and vehicles) and normed by native, monolingual speakers of English and Chinese (Chapter 3). This stimulus set demonstrates the extensive variation in picture naming responses among native speakers of each language and between languages. I use these norms to examine a few specific variables relating to native speaker norms identified in the previous chapter.

In the final set of experiments (Chapters 4 and 5) I select a subset of 183 objects from the new stimulus set to test functional neural correlates of these categorization variables in native,
monolingual speakers of each language and in Chinese-English bilinguals. Each categorization variable is associated with brain regions that uniquely respond to its variation, and activity in these regions confirms functional involvement of both L1 and L2 variables in bilinguals' L1 picture naming behavior. These findings are situated in the broader context of neurocognitive models of language production and offer a refined view of lexical semantic retrieval and selection, accounting for the variety of potential objects that speakers may encounter and variation among native speakers of a language.
# TABLE OF CONTENTS

List of Figures .......................................................................................................................... viii
List of Tables ............................................................................................................................. x
Acknowledgements ...................................................................................................................... xii

Chapter 1 Background .............................................................................................................. 1
  Introduction ............................................................................................................................... 1
  Lexical Categorization ............................................................................................................. 3
  Models for Categorization ....................................................................................................... 8
  Neurocognitive Studies of Naming ....................................................................................... 13
  Bilingualism ............................................................................................................................. 19
  The Present Dissertation ......................................................................................................... 22

Chapter 2 Language History and Language Norms in Bilingual Categorization ....................... 24
  Rationale & Hypotheses .......................................................................................................... 24
  Method ................................................................................................................................... 26
    Participants ........................................................................................................................... 26
    Materials .............................................................................................................................. 27
    Procedure .............................................................................................................................. 29
  Results ................................................................................................................................... 31
    Overview .............................................................................................................................. 32
    Group-wise Analysis: Cross-Language Transfer and Convergence ..................................... 33
    English Category Native-Likeness ...................................................................................... 35
    Chinese Category Native-likeness ...................................................................................... 43
  Discussion ............................................................................................................................... 47
    Summary ............................................................................................................................... 47
    Name Agreement .................................................................................................................. 48
    Alternate Names ................................................................................................................... 50

Chapter 3 Variables in Chinese and English Lexical Categorization ........................................ 53
  Rationale & Hypotheses ......................................................................................................... 53
  Method ................................................................................................................................... 56
    Participants ........................................................................................................................... 56
    Materials .............................................................................................................................. 57
    Procedure .............................................................................................................................. 60
  Results ................................................................................................................................... 61
    English Categorization ........................................................................................................ 61
    Chinese ................................................................................................................................. 67
    English and Chinese Comparison ....................................................................................... 71
  Discussion ............................................................................................................................... 72
    Name Agreement & Alternate Names .................................................................................. 72
    Concept Familiarity & Exemplar Typicality ......................................................................... 75
    Cross Language Correlation & Linguistic Relativity .............................................................. 77
Chapter 4 Brain Correlates of Monolingual Lexical Categorization ........................................80

Rationale & Hypotheses ...........................................................................................................80
Method..................................................................................................................................84
  Participants ........................................................................................................................84
  Materials ..............................................................................................................................85
  Procedure ............................................................................................................................88
  Image Acquisition ................................................................................................................89
  Image Processing ..................................................................................................................90
Results.....................................................................................................................................90
  Overview ..............................................................................................................................91
  Second Level Analysis ........................................................................................................92
  ROI Analyses .......................................................................................................................93
Discussion...............................................................................................................................97
  ROIs and Categorization .....................................................................................................98
  Neurocognitive Models and Categorization .........................................................................100

Chapter 5 Correlates of Competition in Bilingual Lexical Categorization .................................103

Rationale & Hypotheses ........................................................................................................103
Method..................................................................................................................................106
  Participants ........................................................................................................................106
  Materials & Procedure ........................................................................................................107
  Image Acquisition & Processing .......................................................................................107
  First-level Analysis ............................................................................................................108
Results.....................................................................................................................................108
  Anatomically-Constrained Group Results .......................................................................109
  Functionally-Defined ROI Analyses ................................................................................110
  Monolingual and Bilingual Responses in Control Regions ...............................................113
  Connectivity Models ........................................................................................................114
Discussion...............................................................................................................................117
  Cross-Language Interaction in Categorization ................................................................117
  ROIs & Language Competition ........................................................................................118
  Monolingual and Bilingual Comparison .........................................................................119

Chapter 6 General Discussion .................................................................................................121

Summary.................................................................................................................................121
Monolingual Lexical Category Variation ...............................................................................122
Predictors of Bilingual Categorization ................................................................................124
  Native Speaker Norms ......................................................................................................124
  Language History Variables ..............................................................................................126
Lexical Categorization & the Brain .........................................................................................128
Lexical Semantic Theories & Categorization .........................................................................130
  Representation Models ......................................................................................................130
  Production Models ............................................................................................................131
Conclusion..............................................................................................................................135
References ........................................................................................................................................... 137

Appendix A Voxels Significantly Correlated with English Variables (Chapter 4) ............... 149

Appendix B Voxels Significantly Correlated with Chinese Variables (Chapter 4) .......... 150
LIST OF FIGURES

Figure 1-1. (A) English and Spanish names for a set of seats. (B) English and Chinese names for another set of seats. .................................................................4

Figure 1-2. The LRM model illustrated for cross-category competition and the intersecting-category competition elicited in lexical categorization. In the latter example, conceptualization is not obvious, reflected by disagreement among native speakers in lexical retrieval. .................................................................9

Figure 2-1. Sample displays for English (A) and Chinese (B) naming tasks depicting two example objects from the dishes stimulus set. (C) Additional example objects. Participants viewed all 67 objects in this format with the text prompt (according to task language) provided below each item. .................................................................29

Figure 2-2. Cross-language correlation matrices for each Immersion group, representing the correlation between lexical categorization patterns in Chinese (C) and English (E) for the bilinguals and native speaker norms. .................................................................33

Figure 2-3. Two-way interaction plots for Immersion and each of the language history variables. (A) Years of L2 study. (B) Age of earliest L2 exposure. (C) Self-reported frequency of code-switching. .................................................................37

Figure 2-4. Two-way interaction plots for Immersion and each of the linguistic community norm variables. (A) L2 native speaker agreement (percent of a norming sample who produced the dominant name) (B) L1 native speaker agreement. (C) Number of different names for an object produced by the L2 native speakers in a norming sample. (D) Number of different names produced by L1 norming sample. ...........39

Figure 2-5. Observed and estimated effects of language-specific variables: Dominant name agreement and number of alternate names. (A) The interaction between native speaker naming agreement in Chinese and English as predictors of English naming accuracy by Chinese-English bilinguals. Each point on this graph represents a set of objects with particular agreement values in Chinese and English. Color of each point indicates mean accuracy ratings for the participants, and size of each point indicates the number of objects represented. (B) The accuracy values estimated by the statistical model at varying levels of Chinese and English agreement, controlling for all language history variables and number of alternate names. (C) The interaction between the number of names generated by native speakers of Chinese and English as predictors of English naming accuracy by the Chinese-English bilinguals. (D) The accuracy values estimated by the statistical model at varying numbers of alternative names in Chinese and English, controlling for all language history variables and the level of native speaker agreement. ........................................................................41

Figure 2-6. Two-way interaction plots for Immersion and each of the categorization variables. (A) L2 native speaker agreement (percent of a norming sample who produced the dominant name) (B) L1 native speaker agreement. (C) Number of alternate names for an object produced by the L2 native speakers in a norming sample. (D) Number of alternate names produced by L1 norming sample. ........................................45
Figure 3-1. Participant prompts for (A) naming (“CUP” entered by participant), followed by either (B) exemplar typicality rating or (C) concept familiarity rating................................60

Figure 3-2. Box plots for four English categorization variables in each semantic domain. ....63

Figure 3-3. Box plots for four Chinese categorization variables in each semantic domain.....69

Figure 4-1. (A) An object stimulus. (B) A scrambled stimulus. .....................................................88

Figure 4-2. Locations of ROI spheres in four anatomical regions of interest. (A) Calcarine fissure in OL, (B) Pars triangularis in IFG, (C) Fusiform gyrus in vOT, (D) Middle temporal gyrus. See Table 4-6 for exact coordinates and details.................................95

Figure 4-3. Standardized beta values for each category variable by language in four regions of interest. Note: Error bars indicate one standard error of the mean. * denotes significant pairwise differences between category variables. ...............................96

Figure 5-1. SPM thresholded T-map of functional response to English Names variable. Cluster of 18 voxels located in left anterior cingulate cortex.................................................110

Figure 5-2. Standardized betas for Chinese and English Names variables in two regions of interest. Significance * p<0.05, **p<0.01 for $\beta \neq 0$. .........................................................112

Figure 5-3. Standardized beta values for Chinese and English category variables in control regions of monolingual and bilingual speakers during Chinese lexical categorization. Significance * p<0.05, **p<0.01 for $\beta \neq 0$. .........................................................114

Figure 5-4. Connectivity graphs estimated for Chinese Bilinguals and Chinese Monolinguals. All networks contain only those connections which were included in 75% or more of participants’ individual models. “Pic” represents appearance of an object stimulus. All ROIs are represented by spheres described in Table 4-6. See Table 5-3 for ACC details. Note: All ROIs significantly correlated with themselves in the time-lagged comparison. These auto-regressive terms are not reported..............116
# LIST OF TABLES

Table 2-1. Participant demographics before screening. ............................................................27

Table 2-2. Summary of each category variable predicting bilinguals' native-likeness in L1 and L2 lexical categorization (picture naming task). Direction of relationship denoted by + or –. .........................................................................................................................49

Table 3-1. Participant demographics (means and standard deviations) for native English and Chinese norming samples. .......................................................................................................................56

Table 3-2. Correlations between categorization variables in English norming data. All correlations significant at $p<0.001$. .........................................................................................................................66

Table 3-3. Correlations between categorization variables in Chinese norming data. All correlations significant at $p<0.001$. .........................................................................................................................71

Table 3-4. Correlations between Chinese and English values for each category variable.....72

Table 4-1. Neurocognitive models of functional brain regions in word production. Stages are adapted from Levelt and colleagues' (1999) LRM model of production to accommodate the different cognitive models proposed in each paper. See Table 4.2 for abbreviations of anatomical structures. All structures refer to left hemisphere. ........81

Table 4-2. Legend of abbreviations for anatomical structure names. ........................................82

Table 4-3. Hypothesized functional-anatomical correlates of category variables. See Table 4.2. for abbreviations of anatomical structures. All structures refer to left hemisphere. .....................................................................................83

Table 4-4. Descriptive statistics for category variables in neuroimaging stimulus set. Median values presented instead of mean because variables were highly skewed from normal distribution.................................................................86

Table 4-5. Within-language correlations (Pearson $R$) for category variables in neuroimaging stimulus set. **$p<0.001$, *$p<0.05$ ...........................................................................................................87

Table 4-6. Locations of spheres in anatomical and MNI space representing each ROI. Studies providing peak voxel coordinates for each sphere centroid cited under Source. ...............................................................................................................94

Table 4-7. Standardized beta values for each category variable in four regions of interest. Compare to Table 4-3 in Rationale section. ................................................................................................................97

Table 5-1. Hypothesized functional correlates of Names variable in Chinese and English. ...106

Table 5-2. Results of anatomically-constrained analysis for Chinese and English Names. ....109

Table 5-3. Locations of spheres in anatomical and MNI space representing each ROI. ........111
Table 5-4. Mean standardized beta values for Chinese and English Names in each ROI. Compare with Table 5.1 (Rational & Hypotheses). Significance ~ p<0.10, * p<0.05, ** p<0.01 for β≠0. .................................................................112

Table 5-5. Connectivity weights for monolingual and bilingual Chinese connectivity models, estimated by the euSEM using the GIMME group-wise search procedure. ROIs in columns direct influence on ROIs in rows. Connection strengths are reported (X/Y) where X is the time-lagged parameter and Y is the contemporaneous parameter. ........................................................................................................115

Table 6-1. Summary of neuroimaging and behavioral correlates of two category variables in native Chinese speakers' L1 lexical categorization. ......................................................124
ACKNOWLEDGEMENTS

The research described in this dissertation would have been impossible without the generous contributions of several organizations. Throughout my stay at Penn State, I’ve benefited from both the intellectual and financial support of the Brain, Language, and Computation Lab and the Center for Language Science. Grants from the National Science Foundation (BCS-1057855), NSF’s Partnership for International Research and Education (OISE-0968369), NSF’s East Asia and Pacific Summer Institutes (OISE-1210179), the Confucius Institute at Penn State, and the Pennsylvania Space Grant Consortium all provided funding towards my graduate work. I am also indebted to the State Key Laboratory for Cognitive Neuroscience and Learning at Beijing Normal University and the Wang Lab at South China Normal University who both provided laboratory resources for this research.

I am grateful to my thesis and dissertation committees who provided guidance and advice through the long process of crafting a coherent research project and to my collaborators who were directly involved with the formation of this work: Barbara Malt (Lehigh University), Eef Ameel (University of Leuven), and Ruiming Wang (South China Normal University). Nothing is accomplished in academics without the diligent labor of many undergraduate research assistants, and I am especially thankful to Anqi Li, Tianyang Zhang, and Patrick Clark from Penn State and Emily Fricke from State College High School for their tireless work and invaluable contributions. Several graduate students also made important and direct contributions of their time and skills toward this work, namely Peiyao Chen (Northwestern University) and Han Wu, Lijuan Zou, and Zhichao Xia (Beijing Normal University).

Finally, this dissertation is dedicated to KNS who weathered every sleepless night of graduate school right by my side and to CPY who fed and sheltered me in the final months of writing and didn’t complain as my sanity slowly unraveled into a mess of Matlab code.
Chapter 1

Background

Introduction

Language users employ a finite set of names to refer to both concrete and abstract concepts, but the application of these names must be extended to a nearly infinite set of potential referents that the language user could consider or encounter. Referents (such as objects) may deviate from their prototypes on multiple dimensions (e.g., for objects: size, color, texture, shape, and function). Research in lexical categorization describes the relatively small set of implicit, linguistic groups into which the world of possible referents is divided, defined by the names that members of each group or category share. For example, English speakers use the name bottle to refer to a great variety of storage vessels, resulting in a lexical category (that is, the category bottle) with a diverse array of members but also excluding many other similar vessels. Studies of lexical categorization (reviewed below) have demonstrated that these categories are not easily defined by a single feature or by similarity to an ideal prototype. Instead, researchers have identified a great deal of inter-language and inter-personal variation in these categories, adding graded uncertainty and variation in lexical semantics—the links between exemplars and lexical concepts, how these mappings may shift as a function of language experience, and how these changing mappings support simple word production tasks like picture naming.

In the present dissertation, I aim to disentangle the respective roles of individuals’ history of language experience and norms of the native speaker community for word use as factors in the representation of lexical semantic mappings, through the lens of lexical categorization. Beginning with a cross-section of Chinese-English bilinguals, I document inter-speaker variation in lexical
categorization as a consequence of bilingualism and as a function of language history and language norms. Using a statistical model of lexical category patterns for bilinguals with varying experience with English, I offer an interactive account of how various aspects of native language, second language, and language learning history jointly influence the lexical semantic representations that underlie object naming.

Next I explore the neurocognitive mechanisms supporting lexical categorization in monolinguals and bilinguals. While neuroimaging techniques have been broadly applied to word production in general, no imaging studies have attempted to target lexical categorization per se, examining the wide range of referents within a given lexical category and changes in that range as a result of language interaction. I investigate what functional brain processes support lexical categorization in Chinese and English monolinguals, relating these results to known neurocognitive models of word production. I then ask how these processes are altered in native Chinese speakers by the acquisition of an English lexicon which conflicts with the lexical categories of Chinese. I specifically elaborate on the role of cortical areas strongly associated with lexical semantic processing (such as left inferior frontal gyrus, IFG, and left middle temporal gyrus, MTG) as mechanisms of retrieval and selection, examining the relationship between these regions in lexical categorization behavior and in resolving conflicts between the lexical categories of two languages.

With two decades of neurocognitive research of word production in one hand and a newly elaborated neurocognitive account of lexical categorization in the other, I compare these complementary fields for their respective contributions towards understanding how new objects are conceptualized and named by language users. Theories of word production, well-situated in the broader context of language representation and processing, provide strong hypotheses for cognitive and neural mechanisms of object naming but lack a sufficiently ecological account of the ambiguity associated with naming new objects. Lexical categorization, by contrast, provides a
detailed perspective on the diversity and variability of object names among native and second language speakers of a language. A neurocognitive account of lexical categorization that can be reconciled with current word production theory stands to unify these two fields and significantly elaborate our understanding of object naming behavior.

**Lexical Categorization**

Early studies of lexical categorization focused on color categories among speakers of Native American languages (e.g., Landar, Ervin, & Horowitz, 1960; Ervin, 1961). Landar et al. (1960) noted that native Navajo and English speakers used different wavelength boundaries along the visible spectrum for naming colors, such as the shared Navajo category *dool 'iž* for English categories *blue* and *green*. Decades of follow-up research have identified a rich inter-language variation in color categories, including Vietnamese (Jameson & Alvarado, 2003), Russian (Winawer, Witthoft, Frank, Wu, Wade, & Boroditsky, 2007), Himba (Roberson, Davidoff, Davies, & Shapiro, 2005), and 110 unwritten languages from around the world (World Color Survey: Cook, Kay, & Regier, 2005). More recent functional MRI research has revealed that these language-specific category norms predict variation in language-related processing activity (specifically activity in superior temporal gyrus and inferior parietal lobule) in a non-linguistic color discrimination task (Tan, Chan, Kay, Khong, Yip, & Luke, 2008), highlighting the importance of lexical category representations in the processing of novel visual stimuli.

Lexical categorization research has gradually turned towards differences in the names of familiar concrete objects, such as furniture. Native speakers of Spanish and English, for example, discriminate between single- and multiple-occupancy seats differently, with the English category *stool* being subsumed under Spanish categories *silla* (translated *chair*) and *banco* (roughly, *bench*; Graham & Belnap, 1986), as illustrated in Figure 1-1. Similarly, Chinese and English
discriminate between single- and multiple-occupancy seats based on different principle features: English speakers mainly focus on the number of possible occupants (*chair* for one person, *couch* for more than one), but Chinese speakers focus mainly on the material (*yízi* for hard seats, *shāfā* for cushioned seats; Malt, Li, Ameel, Pavlenko, & Zhu, 2013), also illustrated in Figure 1-1. The growing literature on lexical categorization of common concrete objects highlights differences in furniture, clothing, and household storage and serving vessels in English, Spanish, Chinese, French, Dutch, and Russian (Graham & Belnap, 1986; Malt, Sloman, Gennari, Shi, & Wang, 1999; Malt, Sloman, & Gennari, 2003; Ameel, Storms, Malt, & Sloman, 2005; Pavlenko & Malt, 2011; see Malt & Majid, 2013 for review).

![Figure 1-1](image)

**Figure 1-1.** (A) English and Spanish names for a set of seats. (B) English and Chinese names for another set of seats.

Inferring these category boundaries requires considerable experience with a language, as evidenced by a long developmental period required for categories to organize towards adult-like patterns, years beyond the typical age of acquisition for a category name (Ameel, Malt, & Storms, 2008). Language learners must acquire a nuanced, non-obvious, and language-specific ways of using these names. Formation of lexical categories may begin in infancy, described in developmental literature as generalization, in which children infer which known names are best applied to new objects. However, this learning continues beyond childhood for native speakers, at least up to 14 years of age (Ameel *et al*., 2008), reflecting the significant challenge in language
acquisition that word learning poses, even for native monolinguals (see Bowerman & Levinson, 1996 for monolingual studies).

Developing these adult, native-like boundaries between close competitor names also requires attention to an increasing number of features of an object over time (Ameel et al., 2008). The English lexical category *bottle* provides an interesting example of the multi-dimensional definition for these categories, and consequently the difficulty they pose to language learners. No single concrete or abstract feature is sufficient for determining membership in *bottle* even when the set of possible objects is constrained to a particular domain. Among 60 household containers, Malt, Sloman, Gennari, Shi, and Wang (1999) identified a subset of 16 objects labeled *bottle* by native English speakers. Participants’ lexical categories were not sufficiently described by shape (typically cylindrical), material (plastic or glass), or function (containing fluid or solids) alone, but rather by the interplay within a network of such features. Malt et al. (1999) also found that objects which were judged to be more similar to objects in other lexical categories (such as *container*) could still be labeled *bottle* by the majority of participants based on chaining (see Lakoff, 1987 and Heit, 1992), an effect which depends on knowledge of several exemplars of gradually decreasing typicality and their shared name in order to infer the name of a highly atypical object. This dependence of category membership on knowing an elaborated network of similar objects and their names helps to explain the long period of development for adult-like patterns and emphasizes the importance of individuals’ language learning experiences in shaping the lexical semantic representations that produce these categories.

Inter-personal and inter-language variation in lexical categorization poses a unique problem for bilinguals: Learning only translation equivalents between languages will lead second language (L2) speakers to use non-native-like categories in many circumstances. Returning to the example of Chinese and English names for single-occupancy seats, the Chinese translation *yǐzi* for *chair* does not apply to large cushioned chairs, which belong to the Chinese category *shāfā*. 
An English speaker learning Chinese is likely to call such a cushioned chair *yǐzi* based on the most common translation, resulting in a lexical semantic error. Can proficient bilinguals maintain two separate and monolingual-like systems of categorization, must one language give way to the other (perhaps the second language adopting categories of the native language), or will the lexical categories of each language be merged resulting in non-native-like object naming in both languages? Ameel *et al.* (2005) demonstrated that simultaneous Dutch-French bilinguals in Belgium produced a novel set of lexical categories for storage containers and serving dishes, drawing upon monolingual norms in each language, but maintaining a greater internal consistency for the bilingual than otherwise occurring between monolinguals of the two languages.

A follow-up study of Dutch-French bilinguals demonstrated that cross-language consistency was increased through at least two general changes in the bilinguals’ lexical categories: category centroids shifted closer for analogous categories between languages (to a significantly greater degree than between monolinguals of each language) and category boundaries were defined by a smaller set of features and a greater adherence to perceived (nonlinguistic) similarity than monolinguals (Ameel, Malt, Storms, & Van Assche, 2009). This shift is highly suggestive that bilingual lexical category development aims at conflict reduction, relying on common ground between languages like roughly equivalent category names and overall similarity to adjust the representations of monolingual norms towards one another.

Sequential bilinguals (those who learn a native language before beginning acquisition of a second) may reasonably be predicted to follow a different trajectory of categorization. Importantly, the Dutch-French simultaneous bilinguals were immersed in a bilingual environment and thus likely to receive input from each language on a regular basis, potentially reinforcing the novel bilingual system of categories through interaction with other simultaneous bilinguals. By contrast, immigrant populations in the United States are often sequential learners and immersed in
a significantly more monolingual English environment. Native Russian speakers living in the United States show not just L1-to-L2 transfer of lexical categories but also L2-to-L1 transfer, particularly for earlier ages of immigration (ages 15 and below; Pavlenko & Malt, 2011). This L2-to-L1 transfer effect has also been observed among Chinese-English bilingual students who were significantly less native-like in L1 categorization than Chinese monolinguals, even given relatively late immersion in English at an average age of 21 years (Malt et al., 2013, under review).

Sequential bilingual studies provide strong evidence for the dynamic state of bilinguals’ lexical categorization patterns, even well into adulthood. The mechanisms underlying these changes are not well understood, but the study of sequential bilinguals allows researchers to observe these changes in action, either longitudinally or cross-sectionally. In a comparison of Chinese-English bilingual students immersed at Chinese and American universities, Zinszer, Malt, Ameel, and Li (2012, 2013) have identified both the number of competing labels for an object (e.g., cup, mug, and glass for a tall, transparent vessel with a handle and typically used to serve beer) and monolinguals’ level of naming agreement on that object in both languages as significant predictors of bilinguals’ likelihood of producing the object’s monolingual-like name in English. Further, bilinguals’ unique language histories (age of second language acquisition, years of language exposure, and frequency of code-switching behavior) predicted overall monolingual-likeness of their second language (English) lexical categories, suggesting usage-based variability in category formation.

The evidence from bilingual categorization strongly suggests the importance of variation among native speakers and inter-language competition in shaping the development of lexical categories in bilinguals. Does resolution of this competition lead to the patterns of L2-to-L1 transfer and convergence observed in more proficient bilinguals? Recent training studies have examined response times and error rates in English monolinguals trained to categorize objects in
a second language (Zinszer & Li, 2010; Malt et al., under review). These studies show effects of category differences in the new language on English picture naming responses after only brief training. Bilingual speakers, by this account, contend with highly nuanced and often conflicting lexical semantic representations from each language when naming pictures of common concrete objects. The cognitive and neural mechanisms that underlie these representations and regulate this competition are not yet known. Understanding these aspects of lexical categorization may help situate it in the broader theoretical context, which is discussed in the following section.

Models for Categorization

The foregoing review of lexical categorization covers primarily descriptive research, documenting variation within native speakers, between languages, and among bilinguals of various languages. While a handful of these recent studies have employed statistical and experimental techniques to identify factors in lexical category formation and change (Ameel et al., 2008, 2009; Pavlenko & Malt, 2011; Zinszer et al., 2011, 2013; Malt et al., 2013, under review), lexical categorization has not yet been directly tied to a theoretical model, with the notable exception of Pavlenko's (2009) Modified Hierarchical Model, discussed later in this chapter. This descriptive focus, however, is not for lack of theoretical relevance. First, lexical categorization is a paradigm for the study of word production, differing from other picture naming studies mainly in scope by analyzing broad patterns of naming rather than the association of names with single objects of pictures. In this respect, lexical categorization is properly subsumed under theories of word production that must, in turn, adequately account for the variation within lexical categories (to an extent not observed in other word production studies). Further, the latest work in lexical categorization draws strong analogies between categorization patterns and connectionist models of lexical semantic representation (Ameel et al., 2009; Malt et
al., under review). Connectionist models offer explicitly defined mechanisms of learning and processing that can be compared with current and future observations in lexical category development.

**Picture naming with only cross-category competitors**

![Diagram showing conceptualization, lexical retrieval, lemma, and lexeme stages for cat and dog.]

*Figure 1. The LRM model illustrated for cross-category competition and the intersecting-category competition elicited in lexical categorization. In the latter example, conceptualization is not obvious, reflected by disagreement among native speakers in lexical retrieval.*

How is a stimulus, such as a novel cup, associated with an existing semantic concept, linked to the lexicon, and output as a lemma? This sequence for pairing conceptual and linguistic representations is described by existing models of word production, most notably, Levelt, Roelofs, and Meyer’s (1999) LRM model of word production. The LRM model describes two early stages for (1) Conceptualization and (2) Lexical Selection that result in a winning lemma, but the details of the Conceptualization stage itself remain unelaborated. Some experimental paradigms have addressed inter-lemma competition in the selection phase. Semantic interference studies (e.g., Hermans, Bongaerts, De Bot, & Schreuder, 1998; Hermans, 2004) have revealed a naming latency due to the activation of semantic neighbors in picture naming, although this effect
has been exclusively tested in paradigms that offer only one acceptable target response and one or more incorrect distractors (e.g., the target cup and distractor plate). By contrast, lexical categorization involves competition among nearly interchangeable concepts denoted by the multiple acceptable names for an object (e.g., glass, mug, and cup). This aspect of Conceptualization is not well understood or explained in LRM theory, but it is empirically suggested by the prolonged development of monolingual lexical categories and by cross-language interaction effects in bilingual lexical categorization.

Native speaker norms for lexical categorization provide some possible measures of lexical semantic variation between exemplars of a single name: In one study, native speakers of English named specific drinking vessels glass at different levels of agreement, ranging from 65% to 100% (Pavlenko & Malt, 2011). All of these exemplars of glass share the same dominant name (determined by the majority response among native speakers) and have alternate, competing names (e.g., cup in the same study). These norming data might be interpreted as degrees of competition between glass and cup for each object proportional to the number of native speakers who finally produced each name. By another measure, when native speakers produce many names for an object, the number of these alternate names may indicate competitive processes analogous to the semantic interference between a target name and related (but unacceptable) names for an object (e.g., cup and plate). In the same norms collected by Pavlenko and Malt (2011), native Russian speakers produce different numbers of alternate names for drinking vessels called kruzhka (roughly, mug), even at the same levels of naming agreement. One kruzhka named with 75% agreement in their stimulus set elicited three alternate names (stakan, chashka, and bokal) while another kruzhka with 75% agreement elicited only two alternatives (stakan and chashka). Picture naming studies which present only stimuli with high agreement and analyze only trials in which the dominant name is produced implicitly exclude these competitive factors, drawing the direct one-to-one connection from the LRM model’s Conceptualization and Lexical
Selection stages with no account of the ambiguity encountered in ecologically valid object naming.

Although the LRM theory of word production has not addressed variation in lexical categorization, connectionist models of lexical semantics are well suited for explaining this phenomenon. Lexical categorization itself has seen only a few recent modeling attempts (Zinszer et al., 2011; Fang, Malt, Ameel, & Li, in prep), but connectionism has sought to explain lexical semantic representation since the rise of parallel distributed processing in cognitive science (Rumelhart & McClelland, 1986). These models use distributed feature representation and variable connection weights to simulate word learning based on the association of conceptual representations (such as McClelland & Rogers, 2003). These feature-based models reproduce semantic category hierarchies, such as plant and animal taxonomy (e.g., sunfish belongs to fish, which belongs to animals, all of which differ from plants) as a result of feature overlap between exemplars. Ameel et al.’s (2008) analysis of object feature importance in learners’ lexical categories provides the critical theoretical link between lexical categorization and the distributed feature representations applied to connectionist models of lexical semantics. Computational implementations of connectionist models rely on adjusting connection weights between names and features to learn taxonomic hierarchies for multiple objects with different names (McClelland & Rogers, 2003), and statistical accounts of lexical category development show that child and adolescent L1 learners similarly adjust the relative weights of features to learn adult-like lexical categories for multiple exemplars of a single name (Ameel et al., 2008).

Native speakers’ naming patterns in lexical categorization tasks naturally translate into connectionist modeling terms: A particular storage vessel, represented by its unique set of features, may be associated with the lexical node for bottle in 70% of learning cycles and with jug in 30% of cycles, thus operationalizing the native speaker agreement of 70% observed in a behavioral norming study. This inconsistent training paradigm adds competition to the model’s
associations, creating graded connectivity between feature sets and names so that both learned and new feature sets (describing known and novel objects, respectively) do not associate exclusively with a single lexical node. In fact, bilingual theories of lexical semantic representation have already taken this route to describe the shared, cross-language conceptual store (Van Hell & De Groot, 1998; Dong, Gui, & MacWhinney, 2005), and the most recent accounts of bilingual lexical categorization have also drawn on this model (Ameel et al., 2009; Malt et al., under review) to describe the relationship between category names and object features.

The Distributed Feature Model (DFM) proposes a set of underlying features whose combination may be used to define lexical concepts by linking these features to a lexical node (De Groot, 1992; Van Hell & De Groot 1998). The DFM accurately predicts a semantic similarity effect in word translation. Concrete nouns, Van Hell & de Groot (1998) found, were translated more quickly than abstract nouns because these concrete nouns share more representational features with their translation equivalents than the abstract nouns do. The DFM has since been further adapted to accommodate broader asymmetry between languages (both in concrete and abstract nouns) by varying the connection strength as a result of relative importance or salience of a given feature in each language (Dong, Gui, & MacWhinney, 2005). This adaptation is critical to lexical categorization of concrete objects because multiple features must be weighed in categorizing individual objects (Ameel et al., 2008) and because the relative salience of these features changes with bilingual status (Ameel et al., 2009).

Building on the connectionist framework for learning lexical semantic connections and the emergence of lexical categories, several factors are likely to bear on these mappings and define an individual speaker’s lexical categories, especially in the case of bilinguals who have conflicting input from each language. Temporal variables (age of L2 onset, years of study, and years of immersion in L2) are transparently analogous to training cycles for a model (that is, the amount
of input over a period of time) and have been implemented in previous developmental connectionist models (e.g., Li, Zhao, & MacWhinney, 2007; Zhao & Li, 2010). Language use and native speaker norms may also parallel training parameters in connectionist models of lexical semantic representation. These variables describe the frequency or pattern of input over time. When both L1 and L2 lexical semantic systems are active in the use of either language (Wolff & Ventura, 2009), associating a name in L2 (e.g., bottle in English) with a set of features also activates the closest Chinese equivalent (perhaps píngzi), resulting in adjustment of the connections between the lexical node píngzi and the active features associated with bottle. Further, the degree of this co-activation likely depends on the frequency of code-switching in a particular language environment (Green & Abutalebi, 2013).

The studies described in this dissertation do not directly implement any computational architecture, observing only the patterns of lexical categorization in Chinese and English monolinguals and Chinese-English bilinguals over their varying language learning experiences. However, word production theory and connectionism provide an important framework for interpreting the variables that may bear on lexical categorization and elaborating the relationship between lexical categorization and the broader field of word production. This framework may permit a coherent theory for lexical semantic representation and production in monolinguals and bilinguals, providing for the interactive and competitive processes between languages and within the individual as a function of her language input and unique language experience.

**Neurocognitive Studies of Naming**

Recent efforts to understand word production have drawn on neuroimaging techniques to draw parallels between the LRM model and functional brain activity over time and in specific cortical regions. Chronometric studies have estimated completion of LRM’s Conceptualization
and Selection stages at 175 and 250 ms respectively (Indefrey & Levelt, 2004). These estimates are subject to considerable trial-by-trial variability and more recent work has shown that conceptual processing is likely to continue well beyond the first 200 ms as evidenced by the delayed availability of information not ostensibly required for naming (Abdel Rahman & Sommer, 2003).

Complementing chronometric studies, a vast literature of functional neuroimaging studies have described networks of brain regions implicated in production. The left ventral stream, ranging from occipital regions along the temporal lobes and terminating in the inferior prefrontal cortex, appears to underlie lexical semantic processing (contrast with the dorsal stream for phonological and articulatory processing; Rodriguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009; Hickok, 2009). Within this broad subset of brain regions, neurocognitive models have attempted to assign specific brain regions to the cognitive processes underlying production. Notably, Indefrey (2011) directly translates Levelt and colleagues’ (1999) LRM model into functional anatomic regions. Several other reviews and meta-analyses have ventured to integrate brain imaging studies across multiple tasks (picture naming, word reading, semantic and phonetic judgment, verbal fluency, etc) into a coherent brain-based model of speech production (Price, 2000, 2012; Rodriguez-Fornells et al., 2009; Hickok, 2009). While these models differ modestly in the extent of the identified brain regions and their specific cognitive functions, major patterns of agreement emerge:

- lexical semantic retrieval relies on inferior, middle, and anterior temporal regions
- phonological, acoustic, and lexeme (word form) processes are centered around superior temporal and temporo-parietal regions
- and cognitive control mechanisms such as sequencing, selection, and inhibition are largely located in prefrontal regions
For the purpose of investigating lexical categorization, I highlight two anatomical regions of interest which are reliably implicated in neuroimaging studies of word production and correspond with the cognitive processes of conceptualization, retrieval, and selection: left middle temporal gyrus (BA 21) and left inferior prefrontal gyrus (*pars triangularis*, BA 45). Although additional temporal regions have been strongly implicated in conceptual processing (inferior temporal gyrus: Wilson, Isenberg, & Hickok, 2009; anterior temporal pole: Rogers *et al.*, 2006; Woollams, 2012), middle temporal gyrus is preferred because inferior temporal areas are also likely associated with non-linguistic object or image processing such as the apparent familiarity effect observed by Wilson *et al.* (2009). Analogously, additional prefrontal and subcortical regions have been associated with lexical selection and conflict resolution (MFG: Rodriguez-Fornells *et al.*, 2009; anterior cingulate cortex & basal ganglia: Abutalebi & Green, 2007; Abutalebi *et al.*, 2008), but many of these regions may be more involved in executive control activities than *pars triangularis* which has more direct evidence for its involvement in selection (Kan & Thompson-Schill, 2004).

Investigation of middle temporal lobe's involvement in word production has resulted in several similar approximations of function: conceptual information storage, supramodal semantic processing, and lexical semantic processing (Rodriguez-Fornells *et al.*, 2009). MTG involvement in conceptually-driven lemma retrieval was hypothesized by Indefrey and Levelt (2004) and has subsequently been represented in most neurofunctional models (e.g., Hickok & Poeppel, 2007; Poeppel, Emmorey, Hickok, & Pylkkänen, 2012). Evidence for left MTG's role as a lexical semantic processor is supported by neuroimaging studies of word-picture semantic interference (de Zubicaray, Wilson, McMahon, & Muthiah, 2001; Maess, Friederici, Damian, Meyer, & Levelt, 2002), semantic fluency (Vitali, Abutalebi, Tettamanti, Rowe, Scifo, Fazio, Cappa, & Perani, 2005), and word association (Blumenfeld, Booth, & Burman, 2006). This word association study, in particular, highlighted individual differences in the representation of
semantic relationships in left MTG, showing that activity in this region was positively correlated with individual participants' overall accuracy in the judgment task, suggesting that greater activity in MTG corresponded with the more elaborated lexical semantic network needed to associate distant or unfamiliar word pairs.

Further confirmation of left MTG's value as an index of lexical semantic retrieval activity is obtained in regional transcranial magnetic stimulation research (rTMS), which reverses the experimental manipulation used by fMRI to test causal hypotheses about brain-behavior associations. By targeting left MTG for temporary, localized rTMS disruption, Schuhmann, Schiller, Goebel, & Sack (2012) demonstrated that disruption of picture naming occurred only when rTMS was applied at the 225 and 400 ms time points, the former corresponding to lemma retrieval (conceptualization and selection) in Indefrey & Levelt's (2004) chronometric model. Schuhmann and colleagues hypothesized that the latter time window demonstrates a self-monitoring process that verifies the articulated name is the same lexical concept that was initially selected.

None of these studies, however, have contrasted within-lemma variables, confounding word and exemplar effects in participant ratings such as familiarity, for which Wilson, Isenberg, and Hickok (2009) found significant effects in left fusiform gyrus, but did not contrast between more and less familiar exemplars of the same name. Based on the semantic interference research, one might speculate that an object which evokes multiple candidate names may show proportionally greater left MTG activity as a function of the number of name competitors, but again, studies which have examined naming of objects with low name agreement have not contrasted these objects against high agreement objects in the same lexical category (see Kan & Thompson-Schill, 2004, reviewed below). Consequently, it is unsurprising that effects of naming agreement, name competitors, and overall lexeme-level effects remain undifferentiated for this region.
The function of inferior prefrontal gyrus in production has been considerably less clear, owing largely to the number of processes in which it appears to be involved and the finer anatomical divisions that differentiate its functional regions. Left IFG has been associated with inhibitory control (Abutelabi & Green, 2007), lexical selection (Kan & Thompson-Schill, 2004), phoneme or syllable sequencing (Schuhmann et al., 2012), and all of the above (Sahin, Pinker, Cash, Schomer, & Halgren, 2009). Not all of these functional roles are mutually exclusive, as Abutalebi and Green (2007; reviewed below) describe the role of inhibition in selecting lexical candidates. Distinguishing between *pars opercularis* (BA 44) and *pars triangularis* (BA 45) clarifies these roles somewhat, attributing *pars opercularis* with sequencing tasks and *pars triangularis* with selection and inhibitory control (Parker-Jones, Green, Grogan, Pliatsikas, Filippopolitis, Ali, Lee, Ramsden, Gazarian, Prejawa, Seghier, & Price, 2012).

As an anterior structure, the prefrontal cortex is the terminus of Rodriguez-Fornells and colleagues’ (2009) ventral integration stream, along which visual stimuli are processed, conceptual representations activated, and lexical semantic processing results in lemma selection. Under their model, left IFG structures, namely *pars triangularis* and *pars orbitalis* (BA 45 and 47 respectively), participate in elaborated lexical semantic processing and selection. This model is supported by evidence from semantic interference studies in which left IFG activation increased when semantic competitors were presented, raising selection demands for a picture naming task (de Zubicaray, McMahon, Eastburn, & Pringle, 2006) and from semantic fluency research which similarly induces a participant-driven selection demand by asking the participants to generate a set of semantic related names as quickly as possible (Vitali et al., 2005).

As in the middle temporal lobe research, rTMS has offered some further causal evidence to the case of IFG's role in production. Selective application of rTMS at 300 ms after stimulus presentation results in significant response delays for picture naming, a finding which Schuhmann and colleagues (2012) interpreted this effect as disrupting syllabification, based in part on their
effort to localize the rTMS superior to the vertical ascending ramus in *pars opercularis*. However the timing of the rTMS disruption to this area fits well within the lexical selection phase of Indefrey and Levelt’s (2004) chronometric model, and a visual inspection of probe placement on the Schuhmann *et al.*’s ten participants shows sufficient variation to suggest the possible involvement of *pars triangularis* in the disruption. Schuhmann *et al.* thus acknowledge that lexical semantic processing may be reflected in their IFG result.

As previous IFG research (e.g., de Zubicaray *et al.*, 2001) has focused on resolving conflict between semantically related lemmas, Kan and Thompson-Schill (2004) operationalized this competition using the name agreement measures in Snodgrass and Vanderwart's (1980) normed line-drawing stimulus set. Of specific interest to lexical categorization research, Kan and Thompson-Schill compared high- and low-agreement images with *pars triangularis* activity, finding that this region of IFG was significantly more active in the low-agreement condition. Further evidence from ERP studies using the same line-drawing stimuli indicates that the name agreement effect starts early, with low-agreement names evoking a significant positive-going wave around 100 ms (P1), linked with object recognition, and later a negative-going wave around 200 ms (N2), likely due to lexical selection (Cheng, Schaefer, & Akyürek, 2010). While name agreement and name competition are correlated, this ERP evidence is suggestive that agreement affects lexical semantic processing in earlier retrieval processes, before the involvement of IFG. Sufficient variability between name agreement and number of competing names has been observed within English and Chinese monolingual object-naming norms that these variables make separable and significant contributions towards predicting the categorization behavior of Chinese-English bilinguals (Zinszer *et al.*, 2012, 2013). Consequently the roles of each of these lexical category variables in lexical semantic processing warrant closer investigation.

In the foregoing review, I have presented evidence that left middle temporal gyrus (MTG) and left inferior frontal gyrus (IFG) are reliably and causally involved in the processes of
lexical semantic retrieval and selection, respectively. Lexical categorization, as an application of lexical semantic processing, entails the retrieval of multiple acceptable name candidates for a stimulus and the subsequent selection among those candidates for the best fitting name. Such a procedure describes any picture-naming task to some degree, but lexical categorization uniquely probes instances in which competition between names is especially pronounced, reflecting both linguistic and subjective relativity in conceptual organization. It stands to reason that, in these instances, both retrieval and selection are especially taxed. This observation is supported by evidence of increased MTG and IFG activity in response to between-lemma manipulations of such variables as competition, agreement, and familiarity and by some emerging research on the connectivity between these regions.

Bilingualism

As in monolinguals, bilingual word production in picture naming has been studied to great extent, but bilingual naming research has overwhelmingly focused on competition between translation equivalents, and to a lesser extent on cross-language activation of semantic neighbors (Hermans, 2004), cross-language competition between imperfect translations (Van Hell & De Groot, 1998; De Groot, 1992), and translation of polysemous words (Degani & Tokowicz, 2010). Cross-language lexical category conflicts are distinguished by their occurrence in some members of the category but not others, suggesting that study of the lemma itself is insufficient to reveal these differences.

Lexical categorization studies have demonstrated these subtle changes in lexical semantic representation in the form of category convergence and transfer (changes in lexical categorization patterns corresponding with acquisition of categories in another language) between multiple language pairs (e.g., Ervin, 1961; Ameel et al., 2005; Pavlenko & Malt, 2011). These cases are
highly suggestive of a shared and interactive lexical semantic store for bilinguals, a claim supported in the earliest and foremost theoretical models of bilingualism (Potter, So, Eckardt, & Feldman, 1984; Kroll & Stewart, 1994). The concept-mediation models of bilingual lexical access assert that word forms in each language attach to a shared conceptual representation and that (at least for highly proficient bilinguals) translation between languages occurs by accessing the concept referenced by the source language and naming that concept in the target language. However, even simultaneous bilinguals do not use perfectly converged patterns for object categorization (Ameel et al., 2005), indicating that the conceptual store likely contains both language-shared and language-specific information. While these models don't specifically address cross-language (in)congruence in conceptual representation, some revisions have attempted to elaborate on this issue.

Pavlenko's (2009) Modified Heirarchical Model adds language-shared and language-specific elements of the Revised Hierarchical Model's (Kroll & Stewart, 1994) conceptual store. This modification allows for a mixture of unique lexical category patterns in each language and shared patterns, consistent with the general observation that bilinguals maintain some especially salient distinctions between languages while showing a broader pattern of convergence, particularly between category centroids (Ameel et al., 2009). However, the Modified Hierarchical Model offers no explicit mechanism for how shared and unique concepts are organized or distinguished in a single conceptual store.

Van Hell & De Groot's (1998) Distributed Feature Model seems to underestimate the considerable variation in feature composition even for approximate translation equivalents of concrete objects (Malt et al., 2003), but there is no requirement in the model that concrete objects be represented with strictly congruent feature sets between languages. Dong, Gui, and MacWhinney (2005) further propose that connections to features may be weighted and those
weights changed over time, an adaption that agrees with the re-weighting of features for bilinguals in object naming (Ameel et al., 2009).

Green's (1998) Inhibitory Control model describes a system of executive control functions that manage language competition arising from a shared lexical semantic store. A neurocognitive implementation of this model proposes the Single Network Hypothesis, claiming that first and second languages are (at sufficient levels of proficiency) processed in the same brain networks (Abutalebi & Green, 2007). This model also cites the prefrontal cortex as the locus of response selection and inhibition (in concert with several cortical and sub-cortical structures) for cross-language lexical competition. That aspect of the model is consistent with left IFG's role in lexical selection for monolinguals. Bilinguals, in fact, show marked increases in left IFG activity over monolinguals, even when both groups are using the same language in a “monolingual mode” (absent any language switching demands; Parker-Jones et al., 2012), and individual differences in the increase of this IFG activity were linked to performance on the Stroop task, lending further evidence for IFG's role as a source of inhibitory control in bilingual lexical selection.

Among temporal regions, Parker-Jones and colleagues' (2012) study did not identify significant increases in MTG activity, but did see increases in STG. Drawing on the principle of a shared conceptual store, one might hypothesize that MTG activity is approximately equivalent between monolingual and bilinguals insofar as the languages are roughly congruent in lexical semantic organization (i.e., when lemmas have direct translation equivalence between the languages) but still undergo competition at the lexeme level. In semantic interference studies, direct translations have been demonstrated to prime (rather than compete) with one another at the semantic level (Hermans et al., 1998, 2004), an effect that would be expected to result in decreased MTG activity. However, returning to the special problem of bilingual lexical categorization, this hypothesis must be revised in instances where direct translation fails. Namely, objects belonging to the best (or most frequent) translations of the category name in each
language may experience this translation facilitation. However objects for which the name in one language is *not* the most frequent translation in the other language may lack such a facilitation effect and could even induce additional name competition. Consequently, a bilingual naming study that uses more diverse representation of membership in each lexical category may substantively modify our current neurocognitive models of cross-language lexical interaction.

**The Present Dissertation**

In the first behavioral study (Chapter 2), I explore the lexical categorization of Chinese-English bilinguals at various stages of English immersion in China and the United States by gathering picture naming responses for a set of 67 serving dishes, named in both English and Chinese. Variables describing the bilinguals’ individual language histories and variables describing the diverse native Chinese and English speaker norms for these objects are entered in statistical models to estimate the bilinguals’ idiosyncratic lexical categories. The resulting models test the role of language history and language norms in L1 and L2 naming, and describe the interactive relationship L1 and L2 lexical semantics as a function of individual language experience and language norm variables.

The evidence for cross-language interaction in bilingual categorization and the role of language norms in this interaction motivate a functional neuroimaging study (described below) to test the neurocognitive correlates of variation in native speaker norms (naming agreement and number of alternate names). No existing picture naming stimulus set, however, provides the necessary range of native speaker agreement or documents the full set of alternate names available for object stimuli. To this end, in Chapter 3, I collect lexical category norms for Chinese and English monolinguals over a broad sample of 407 photographs of objects across multiple semantic domains, gathered from images publicly available on the Internet (via Google Image
search). These norms allow parameterization of variables that describe category membership of each photographic stimulus: name agreement, alternate names, name typicality, and concept familiarity. The Chinese and English native speaker norming data collected for these stimuli help to differentiate these variables and provide measures for the neuroimaging studies to follow.

In the first neuroimaging study (Chapter 4), a subset of 183 stimuli from the previous normed stimulus set are presented to Chinese and English monolinguals undergoing functional MRI during a picture naming task. Each of the four normed categorization variables are entered as predictors for functional brain responses and, after controlling for potentially confounding variables (visual complexity and word frequency), the differential involvement of cortical regions of interest is considered for each category variable. These results provide a baseline for neural correlates of monolingual lexical categorization and allow direct comparison of the processes supporting lexical categorization with established neurocognitive models of word production.

In the final neuroimaging study (Chapter 5), I repeat the picture naming task with a group of Chinese-English bilinguals, currently immersed in an English language environment. The bilinguals name objects in their native language (Chinese), and category variables are again associated with brain regions that uniquely respond to their variation. In this study, English variables are entered as predictors after accounting for Chinese variables, testing the behavioral finding that lexical categorization norms of both languages influence naming patterns of bilinguals in each language. These findings are situated in the broader models of bilingual language production and offer a refined view of lexical semantic retrieval and selection, accounting for the variety of potential objects that speakers may encounter and variation among native and bilingual speakers of a language.
Chapter 2

Language History and Language Norms in Bilingual Categorization

Rationale & Hypotheses

The present study uses a stimulus set of common household serving vessels that has previously been used in lexical categorization studies of Dutch monolinguals (Ameel et al., 2008), Dutch-French bilinguals (Ameel et al., 2005), and Chinese-English bilinguals (Malt et al., 2013). In contrast to the previous work, this study sampled Chinese-English bilinguals immersed in both L1 (Chinese) and L2 (English) language environments as well as gathering detailed language history data for each participant to help disentangle the respective roles of language history and language norm variables in predicting the bilinguals' lexical categorization patterns.

Four language history variables are measured: duration of L2 immersion (including zero years for L1 immersed participants), age of earliest exposure to English as a second language, years of English study, and self-reported frequency of code-switching. These variables have been highly confounded in the previous literature, and the present study simultaneously measures all four variables so that their respective effects may be compared.

Further, I draw upon existing monolingual norms from Chinese and English native speakers in China and the United States, respectively, to define language norms that describe the lexical categorization patterns used in each language environment from which the bilingual participants were recruited. These native-speaker language norms characterize objects by two category variables: name agreement (the proportion of the native-speaker norming sample who produced the dominant name), and number of alternate names for the object (the total count of different names produced for an object by the native-speaker norming sample). As native
speakers of each language differ in the way that they categorize these objects, a separate 
measurement of these category variables is used for the L1 and L2 norms. Several previous 
studies of bilingual lexical categorization have compared bilingual patterns to monolingual norms 
(Ameel et al., 2005; Pavlenko & Malt, 2011; Malt et al., 2013), but none of these studies have 
entered these categorization variables as item-wise regressors for predicting the performance of 
bilinguals. In the present study, both sets of native speaker norming data are used in predicting 
categorization patterns of the bilinguals for each language, providing a measure of the relative 
influence of each language over bilinguals' production of L1 and L2.

In this study, I put to test a number of hypotheses arising from the previous lexical 
categorization literature and from the application of existing theories of word production to 
lexical categorization. First, I ask whether cross-language lexical category transfer arises over 
time, as the result of inter-language competition in lexical semantic representation. Framed 
another way, this question distinguishes between sources of convergence in simultaneous 
bilinguals—norms of the bilingual community may confer some degree of lexical category 
convergence while individuals' cognitive resolution of inter-language conflict may result in 
unique convergence between languages. Sequential bilinguals in relatively monolingual (Chinese-
only and English-only) provide a natural test of the hypothesis that convergence may arise from 
competing monolingual input patterns (Pavlenko & Malt, 2011; Malt et al., 2013).

Second, if changes in lexical categorization indeed arise as a function of language 
interaction for bilinguals, I ask what aspects of the language learning experience and language 
norms might bear on the bilinguals’ lexical category patterns. Especially, I explore the role that 
L2 immersion plays in L1 and L2 lexical categorization patterns for bilinguals with consideration 
for the interaction of immersion with other commonly tested variables (e.g., age of acquisition) 
and with the L1 and L2 language norms.
Third, I test the extent of the hypothesized cross-language interaction by estimating the relative influence of L1 and L2 language norms in both L1 and L2 lexical categorization. In effect, I ask whether lexical categorization patterns in each language are sensitive to native speaker norms of the other language: Do L2 lexical categorization patterns depend on L1 norms, and (more interestingly) are bilinguals’ L1 lexical categorization patterns sensitive to L2 norms after L2 immersion experience?

Method

Participants

Two groups of bilingual students, one in the United States and one in China, participated in this study. In the U.S., Chinese-English bilingual undergraduate and graduate students were recruited from the Introduction to Psychology subject pool and through posters around the campus community at Penn State University (State College, PA, USA). In China, Chinese-English bilingual undergraduate and graduate students were recruited through an online campus message board and through word of mouth at Beijing Normal University (Beijing, China). Generally speaking, the students at Penn State were slightly younger than those at Beijing Normal, were first exposed to English at a slightly earlier age, and had higher self-rated proficiencies in English.

Although many of the bilingual participants reported some degree of training in a third language, most rated themselves at very low proficiency. Participants who self-reported a proficiency of 2.5 or greater in the third language on a 7-point scale (averaged across four ratings: reading, writing, speaking, listening) or failed to provide a proficiency rating in their third language were not included in the data. In total, 57 participants from Beijing Normal and 68
participants from Penn State met this inclusion criterion. Third languages included French, German, Russian, Mongolian, Japanese, Korean, Taiwanese, and Cantonese.

Penn State students ranged in age from 18 to 23 \((M=19.5, SD=1.2)\). They were first exposed to English between ages 1 and 16 \((M=8.2, SD=3.7)\), and self-rated their English proficiency between 2.5 and 7.0 \((M=4.7, SD=1.3)\). The Penn State students had resided in the United States for 0 to 19 years \((M=5.0, SD=5.9)\). Students at Beijing Normal University had ages ranging 18 to 28 \((M=22.8, SD=2.0)\), age of earliest English exposure was 5 to 15 \((M=11.4, SD=2.1)\). Their self-rated English proficiency varied between 1.3 and 5.5 \((M=3.9, SD=1.0)\), as some were studying English while others majored in different subjects. None of the participants at Beijing Normal reported living in or visiting an English-speaking country.

Table 2-1. Participant demographics before screening.

<table>
<thead>
<tr>
<th>Sample</th>
<th>(n_{\text{total}})</th>
<th>(n_{\text{included}})</th>
<th>Age ((sd))</th>
<th>L2 Prof. ((sd))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing Normal</td>
<td>57</td>
<td>30</td>
<td>22.8 (2.0)</td>
<td>3.9 (1.0)</td>
</tr>
<tr>
<td>Penn State</td>
<td>68</td>
<td>33</td>
<td>19.5 (1.2)</td>
<td>4.7 (3.7)</td>
</tr>
</tbody>
</table>

I also drew on a set of native-speaker norming data from functional monolinguals who had participated in a previous version of the lexical categorization task, using the same stimuli \((\text{Malt et al.}, 2013)\). These picture naming data for 25 native Chinese speakers in China and 28 English speakers in Pennsylvania provided linguistic community norms for the current analyses. Their choices represent the most likely input patterns for bilinguals immersed in their respective language environments.

**Materials**

All participants completed a language history questionnaire \((\text{Li, Sepanski, & Zhao, 2006; LHQ})\) to assess bilingual status, L2 proficiency, age of L2 acquisition, and behavioral predictors...
such as patterns of code-switching. The LHQ was available in both English and Chinese (simplified script) and administered according to the dominant language environment. With respect to code-switching, the LHQ allows participants to self-rate their frequency of code-switching in four contexts: Spouse & Family, Friends, Co-Workers, and Classmates. Participants rank their code-switching frequency in each context ordinally, ranging from “Rarely” to “Very Frequently.” These responses are transformed into a Likert score between 1 and 5 and averaged within context group to produce the CS scores.

Early trials at Penn State revealed that several participants failed to complete the code-switching section of the LHQ or claimed to never code-switch, a self-report that may (in some cases) be influenced by stigmatization of the behavior. An additional Code-Switching Questionnaire (CSQ) was added to subsequent sessions to specifically probe participants’ code-switching behaviors and screen for effects of negative attitudes about code-switching. The CSQ was administered according to the dominant language environment.

Sixty-seven photographs of common household objects were used to elicit category names from monolingual and bilingual participants. These objects were drawn from a stimulus set (called the dish set) used by Ameel and colleagues (2005) to reveal cross-language lexical categorization differences in Dutch-French bilinguals. Each photograph contained a single household serving vessel (e.g., a plate, cup, or bowl) on a neutral background and a centimeter ruler in the foreground for scale (see Figure 2-1). Photographs were displayed at 480x360 pixels on a personal computer equipped for digital recording. Each photograph was accompanied the written prompt: “What is this?” or “这是什么?” according to the task language (see Figure 2-1 on the following page for examples).

An Operation-Span (O-Span) test was also used to screen the bilingual participant groups for systematic differences in working memory, a cognitive factor that might be confounded with language proficiency or language transfer. The O-Span includes mathematical and verbal
components: participants judge the accuracy of math equations and are provided a word to remember after each judgment. After several math and word combinations, participants are prompted to recall the words they have seen. Arabic numerals were used for the math component (consistent with both Chinese and American math education) and Chinese characters were used for the verbal component. Participants entered their judgments using a computer keyboard and recorded their verbal responses on a paper worksheet. No significant difference was found between the two bilingual samples in their O-Span scores.

![Sample displays for English (A) and Chinese (B) naming tasks depicting two example objects from the dishes stimulus set. (C) Additional example objects. Participants viewed all 67 objects in this format with the text prompt (according to task language) provided below each item.](image)

**Procedure**

After giving informed consent in the local language, participants completed the language history questionnaire, also in the local language. They then completed an unrelated English receptive vocabulary task (results not presented here) to establish an English language mode to the extent possible in both the Chinese- and English-immersed participants. After the vocabulary
test, all participants performed the English picture naming task. The Chinese O-Span was then completed and used to shift participants into a Chinese language mode before naming the objects again in Chinese. Finally, the code-switching questionnaire was completed last. Participants in the US completed English and Chinese tasks on separate days at least one week apart, counter-balanced for order. Sessions in China could not be scheduled separately and all tasks were completed on the same day, with English first, followed by Chinese. I reasoned that the English task was less likely to influence Chinese naming in a Chinese immersion environment (as compared to effects on English production from completing the Chinese task first), and intervening Chinese tasks (namely, the O-Span) would help to reduce any language priming effects.

In the picture naming tasks, participants were instructed to name photographs of objects depicted on the computer out loud. They were asked not to name the objects’ contents, as illustrated by two photographic examples: a grocery bag full of vegetables (called bag) and a trash can full of paper (called trash can). These instructions were provided in written form on the computer screen according to the language of the task. Participants were verbally encouraged to name every object and to always make a guess if unsure. Participants were also provided two practice naming trials for photographs of unrelated bottle-like stimuli, followed by the most dominant monolingual name for each stimulus (bottle or 瓶子) to demonstrate the desired response type. Participants were permitted to take as long as they desired to name each picture to ensure that they selected what they considered the best name for each object. Due to storage constraints, only the first ten seconds (from the onset of the stimulus) of participants’ responses were digitally recorded by the computer for each stimulus.
Results

Participant responses were transcribed from audio recordings by high-proficiency Chinese-English bilinguals in the United States who were able to comprehend Chinese responses and phonetically-accented English responses. Transcribers were not able to view the objects during transcription to prevent bias on ambiguous recordings. Transcribed responses were subsequently reduced to head nouns (e.g., “a small blue bowl” is reduced to “bowl”) for comparison with the native norming data. Skipped trials, inaudible responses, and irrelevant responses (e.g., “I don’t know”) were entered as blanks and treated as incorrect responses.

Four biographical variables were included for each subject: Age of first exposure to English (AOEE), length of residence in the English immersion environment (LOR), self-reported frequency of code-switching between Chinese and English (CSFreq) and the total number of years spent learning English (current age minus the age of first exposure, YrsLearn). For participants who declined to complete some language history and code-switching questions, missing data for the CSFreq variable were replaced with the sample mean (3% of the participants included in the analysis). Participants who did not report AOEE were excluded from analysis, and an additional set of childhood bilinguals (AOEE < 5 years) were removed from the US participant data to maintain comparability with the sample in China (AOEE range 5-15 years).

Results from 20 participants in the US sample and 20 in the China sample were discarded due to recording equipment failure (no naming data recorded). Participants were encouraged to speak loudly and clearly directly into a desktop microphone; however additional participants (not excluded due to recording failure) periodically produced inaudible responses or no response, decreasing their total response rate. Two participants in the US and 4 in China were excluded for response rates below 50% on one or more of the naming tasks. Non-response rates were approximately the same between the English task (4 participants in Beijing) and the Chinese
task (3 participants in Beijing and 2 in State College), suggesting that most of these missing data were attributable to participant inattention. Some participants may have chosen not to respond to stimuli when uncertain about those objects' names, a possibility I tested by correlating response rate to self-rated English proficiency. Indeed, response rate in English was weakly correlated to English proficiency ($R=0.24, p=0.047$) while response rate in Chinese was not ($R=0.20, p=0.101$). Non-response as a predictor of name uncertainty is preserved in the remaining participants (50% or higher response rate) insofar as all trials are included for analysis, with non-response trials counted as incorrect names. Two participants in China were removed for naming accuracy scores greater than 2.5 standard deviations below the mean (see Subject-Wise Analysis). Final participant counts were 30 from China and 33 from State College.

Overview

In the following sections, I present a set of analyses that examine the lexical categorization patterns of the Chinese-English bilingual participants at three different levels. The Group-wise analysis compares the overall patterns of transfer and convergence between Chinese and English as spoken by the bilingual participants. This analysis looks at the overall trends in naming distributions generated by sub-groups (defined by their degrees of L2 immersion) and allows direct correlations of the bilinguals’ overall patterns with the monolingual norms. The Subject-wise analyses focuses on individual bilingual participants’ language histories and how these variables predict their individual differences in L1 and L2 naming patterns. Finally, the Item-wise analyses examine naming performance on each object of the stimulus set, controlling for variation in individuals’ language histories and examining the impact of native speaker norms on the bilinguals’ accuracy for producing dominant L1 and L2 names.
Group-wise Analysis: Cross-Language Transfer and Convergence

For group-wise comparison, participants were organized by three discrete values of LOR to describe three types of immersion conditions observed in our sample: No Immersion, Short-term, and Long-term. No Immersion was defined by LOR=0, describing participants who have never lived in an English immersion environment. English-immersed participants were divided into two groups by a median split (median non-zero LOR=1.3 years). Short- and Long-term Immersion were defined as the samples below and above the median, respectively.

A cross-language correlation matrix was calculated for each bilingual and monolingual group according to the method of Malt et al. (1999; see also Ameel et al., 2005) whereby the naming distribution for each object over all possible names is correlated with every other object (producing, in this case, a 67x67 correlation matrix with 2211 unique values). Sets of these inter-object correlations can then be correlated between languages or groups, representing the degree to which objects names are distributed similarly in the two samples (regardless of the actual names themselves). Figure 2-2 depicts these correlation matrices for each immersion group, compared with the monolingual speakers of Chinese and English and between the Chinese and English patterns produced by the bilinguals, according to the convention of Ameel and colleagues (2005).

![Cross-language correlation matrices for each Immersion group](image)

Figure 2-2. Cross-language correlation matrices for each Immersion group, representing the correlation between lexical categorization patterns in Chinese (C) and English (E) for the bilinguals and native speaker norms.

The cross-language correlations revealed that native, monolingual speakers of Chinese and English correlate in their categorization of this set of objects at about 0.64 (see the top row of
Figure 2-2). This value serves as the baseline correlation against which bilinguals’ Chinese and English categorization patterns can be compared. All of the bilinguals showed a highly convergent pattern of naming between languages, correlating their Chinese and English word use around 0.92-0.95 (the bottom row of Figure 2-2), strongly suggesting that they relied on a single set of mappings (with varying degrees of influence from each language). Correlation of the bilinguals’ English naming with the monolingual norms (the right-most vertical connection for each matrix in Figure 2-2) was highest in the Long Immersion group (0.81, compared to 0.78 and 0.77 in the No and Short Immersion groups respectively). The bilinguals’ English categorization also decreased in its dependence on Chinese norms with increased immersion (No Immersion: 0.80, Short Immersion: 0.78, Long Immersion: 0.74).

Curiously, the No Immersion group showed the highest convergence between their two languages (0.95), a relatively low correlation with the monolingual Chinese (0.80) compared to their recently immersed peers (Short Immersion, 0.85) and greater Chinese resemblance to the English patterns (0.77, compared to Short Immersion 0.75). This effect may be attributable to differences in the administration of the Chinese naming task, in which the No Immersion group (collected in Beijing) completed Chinese naming shortly after the English naming task. Henceforth, I will examine English naming for all three groups, as English names were not subject to priming across tasks because English naming occurring first, and Chinese naming analyses will be limited to the Short- and Long-term Immersion groups who were properly counter-balanced for order and tested with a week between each language.
English Category Native-Likeness

Subject-wise Analysis: The Role of Language History Variables

Participants' picture-naming responses were compared to a set of English native norms to generate a score for each participant describing the English native-likeness of their lexical categories. Each of a participant's responses was awarded a score based on the proportion of English natives in the monolingual norm who produced the same name as the participant (Malt & Sloman, 2003). Thus if an object was called *mug* by 75% of the norming group and *cup* by 25%, the bilingual participant would receive 0.75 points for responding *mug*, 0.25 for responding *cup*, and 0 for anything else. These point values were averaged across the 67 objects for each participant, rendering an agreement-weighted native-likeness score ranging between 0 and 0.68 (the mean of agreement level for native English speakers across all objects).

I estimated a linear regression model for the English native-likeness scores over participants' language histories to determine the relationships between language background and attained L2 lexical category proficiency. Previous analyses from smaller datasets showed several two-way interactions between the language history variables, and an inter-dependency of the significance of these interactions in the model (see Zinszer, Malt, Ameel, & Li, 2012, 2013 for examples). Consequently, the initial model was estimated with all possible interactions (up to four-way) and, indeed, yielded several highly significant three-way interactions (full model $F(15,46) = 2.14$, $p=0.02$, Adjusted $R^2=0.22$).

In an attempt to improve the parsimony of this model without discarding important interaction terms, an automatic Akaike Information Criterion ($AIC$) stepwise procedure was initiated which systematically excluded and re-included variables to find the best-fitting model with the lowest $AIC$ score (Venables & Ripley, 2002). This method produces a reduced model
while minimizing impact on the model's fitness to the data. Finally, the AIC search excluded only the four-way interaction term, resulting in a significant reduced model \( F(14,47) = 2.27, \ p=0.02, \ \text{Adj } R^2=0.23 \) with slightly improved parsimony (initial model: 16 terms, \( AIC=-103.9 \); reduced model: 15 terms, \( AIC=-105.0 \)). Ultimately, all predictors were included in one or more significant interactions.

To understand the highly interactive terms of the subject-wise model, I generated several estimated marginal means plots based on the model's predicted English native-likeness scores across a range of values for the two-way interactions between L2 immersion (LOR) and each remaining predictor (while holding other predictors constant at the mean value). These two-way interactions were all highly significant (LOR x YrsLearn: \( p=0.002 \); LOR x AOEE: \( p=0.014 \); LOR x CSFreq: \( p=0.007 \)). To further simplify the plots, I again used the three discrete values of LOR to describe three types of immersion conditions observed in our sample: No Immersion (LOR=0), Short- (LOR<1.3 y), and Long-term (LOR>1.3 y). Short and Long-term Immersion were represented by the mean LOR values for each of these two groups: 0.5 and 4.7 years, respectively. Figure 2-3 shows plots for the interactions between LOR and each of the remaining predictors: AOEE, CSFreq, and YrsLearn.

The first plot (Figure 2.3A) contrasts years of English study with the duration of English immersion, which are not independent predictors. That is, as the duration of immersion (LOR) increases, so too do the years of English study (YrsLearn). Conversely, however, YrsLearn may increase without immersion experience. Therefore these variables contrast the predicted English native-likeness associated with a varying durations of study when the amount of that time spent in an Immersion environment is held constant (in this case at 0, 0.5, or 4.7 years). In all three LOR conditions, the relationship between YrsLearn and English native-likeness is negatively sloped, indicating a relative disadvantage for years of English study after controlling for years of English immersion. In other words, every additional year of English study in China beyond a participant's
immersion experience reduced the native-likeness of their English categorization patterns. For example, a bilingual with 15 years of study and almost 5 years of immersion (LOR=4.7) has had over 10 years of English study in China, and they are predicted to perform worse (on average) than somebody with fewer years of English study and the same amount of (or even less) immersion.

![Figure 2-3. Two-way interaction plots for Immersion and each of the language history variables. (A) Years of L2 study. (B) Age of earliest L2 exposure. (C) Self-reported frequency of code-switching.](image)

Age of earliest English exposure (AOEE; Figure 2-3B) also displayed a negative relationship with English native-likeness. This finding is more expected. When controlling for the other variables, later ages of English onset generally result in poorer performance in English categorization. Interestingly, however, the interaction with LOR did not appear to be large (the lines are roughly parallel) indicating a largely additive effect of the two variables. The relative weight of each variable was approximately balanced such that the negative effect of being exposed to English one year later is offset by the benefit of one year of immersion experience.

Participants' self-reported code-switching frequency (CSFreq) was also a significant predictor in the model, and significantly interacted with immersion. As Figure 2-3C indicates, the effects of CFSFreq were relatively small for unimmersed bilinguals and bilinguals with relatively little immersion experience, and greater code-switching frequency was associated with less L2 achievement. However, CFSFreq was a much stronger predictor for bilinguals with longer immersion.
immersion experiences, and the direction of the influence was opposite, showing significant gains in English native-likeness with greater frequency of switching between languages.

*Item-wise Analysis: The Impact of Native Speaker Norms*

In this analysis looking at how native naming consensus for objects impacts the likelihood of naming objects correctly, I compared each response by the participants to the single dominant name produced by the English native norm for each given object. Trials in which the participant produced the native speaker norm's dominant name were scored as 1 (correct). All other trials were scored as 0 (incorrect). Next, I performed two binomial logistic regressions to estimate the probability that a participant would produce the dominant name for any given object.

In the first logistic regression, I entered the same language history variables used in the Subject-wise analysis to determine how adequate these variables were for identifying variation in native-like categorization for different objects. The logistic regression model including only participants' biographical information contained several statistically significant predictors, but offered a very poor fit to the data (Nagelkerke $R^2=0.02$), indicating that subjects' language backgrounds could account for overall trends in the native-likeness of their English categorization but not for most of the variation trial-to-trial. This result points to the importance of considering input variation (such as the native norms) as a predictor of success in naming individual objects.

In the next analysis, I added four language variables which described the native speaker norms for every given object: Name Agreement in Chinese (L1), Name Agreement in English (L2), number of Alternative Names produced by the Chinese norming group, and number of Alternative Names produced by the English norming group. Due to computational limitations, this model was estimated with up to four-way interactions and reduced using the same AIC stepwise search procedure described in the Subject-wise analysis. The resulting reduced model
improved the $AIC$ compared to the initial model and included 36% fewer terms than the initial model without a serious decrease in fitness (initial model: 163 terms, $AIC=5036.52$, Nagelkerke $R^2=0.25$; reduced model: 104 terms, $AIC=4951.7$, Nagelkerke $R^2=0.24$).

As in the subject-wise analysis, the model contained many interaction terms that impeded interpretation without isolating a few of the variables. Again, I sought to describe how immersion experience affected the role of these language variables in predicting the bilinguals' success in producing native-like English names for objects. A binomial logistic regression predicts the probability that an outcome will occur, in this case the probability that the bilingual will produce the English native-like dominant name for a given object. Again, I estimated plots in which the individual variables (this time, each language's native speaker norms) interacted with three levels of immersion while holding all other variables constant at a mean value.

![Figure 2-4](image)

Figure 2-4. Two-way interaction plots for Immersion and each of the linguistic community norm variables. (A) L2 native speaker agreement (percent of a norming sample who produced the dominant name) (B) L1 native speaker agreement. (C) Number of different names for an object produced by the L2 native speakers in a norming sample. (D) Number of different names produced by L1 norming sample.
Figure 2-4 contains plots of each language variable against the three levels of English immersion (None, Short-term, and Long-term). These plots revealed that English native-likeness in the bilinguals was more likely at higher levels of English norm agreement (Figure 2-4A), while the inverse was true for Chinese: There was less English native-likeness with higher agreement in the Chinese norm (2-4B). An opposing relationship was also observed for the number of alternative names available from the norming sample. Having a greater number of English names available in the norm actually increased the predicted probability that the bilinguals would produce the dominant English name (2-4C), but having many possible Chinese names for an object decreased the predicted probability of the participants producing an English native-like name (2-4D). The apparent advantage for a greater number of English names is explored in the next section.

I also asked how L1 and L2 norms might interact with one another in predicting a bilingual's success in producing the L2 dominant name. Several interaction terms between these norming variables were highly significant, so I examined the cross-language relationships between L1 and L2 agreement and number of L1 and L2 names. This analysis offers a closer examination of two interesting effects from the preceding results: 1) native speaker agreement in each language appears to compete in predicting L2 native-likeness and 2) an increasing number of names in English seems to be associated with greater L2 native-likeness.

Figure 2-5 depicts both the observed item-wise accuracies (A and C) and the estimated marginal mean accuracy at varying levels of the predictors using the logistic regression model. In Figure 2-5A, the average response accuracy across participants is plotted over both English and Chinese norm agreement levels for each object in the stimulus set. This plot shows the empirical effects observed in the sample data and is generally similar to the competitive account of L1 and L2 agreement estimated by the regression model. Among the objects in the experimental stimulus set, best performance (depicted by blue-colored dots) is observed when both languages have high
agreement levels. However, this performance diminishes as English agreement decreases. In general, high levels of L2 (English) agreement are associated with successful learning across varying levels of L1 (Chinese agreement). However, the worst performance by the bilinguals occurs when Chinese agreement is high and English agreement is low, confirming that L1 patterns can have a strong negative effect on L2 native-likeess when L2 input is inconsistent.

Figure 2-5. Observed and estimated effects of language-specific variables: Dominant name agreement and number of alternate names. (A) The interaction between native speaker naming agreement in Chinese and English as predictors of English naming accuracy by Chinese-English bilinguals. Each point on this graph represents a set of objects with particular agreement values in Chinese and English. Color of each point indicates mean accuracy ratings for the participants, and size of each point indicates the number of objects represented. (B) The accuracy values estimated by the statistical model at varying levels of Chinese and English agreement, controlling for all
language history variables and number of alternate names. (C) The interaction between the number of names generated by native speakers of Chinese and English as predictors of English naming accuracy by the Chinese-English bilinguals. (D) The accuracy values estimated by the statistical model at varying numbers of alternative names in Chinese and English, controlling for all language history variables and the level of native speaker agreement.

For comparison, Figure 2-5B plots the model's estimated accuracy levels, generalized to all levels of agreement in each language. While 2-5B covers a broader range of potential values (such as low naming agreement in Chinese, which is under-represented in the actual stimulus set), it generally fits the patterns established by the empirical data. Further, the observed data (2-5A) do not control for confounding variables (such as the number of alternate names). The regression model estimates both predictors and thus isolates the effect of agreement while holding the number of names constant, resulting in the smoother contours along values of L1 and L2 agreement depicted in Figure 2-5B.

Participants' observed performance across the different numbers of alternative names from in the English and Chinese norms, however, differed significantly from the regression model's estimated accuracy rates. Figure 2-5C depicts the accuracy rate for each object in the stimulus set across varying numbers of names in each language. This plot indicates that when bilinguals had only one L1 name for an object in each language, performance was highest. The worst performance was observed when exactly two competing names for an object were available in either language. The model's prediction that increasing the number of competing names improved the probability of bilinguals producing the dominant name (Figure 2-5D) is consistent with the latter observation that objects with three or more competing names were named more accurately than those with two names, but it overlooks the advantage for objects with only one name. Again, in the observed data (2-5C), the number of names and agreement level are confounded, but the regression model isolates these effects and controls for agreement in its estimations of the effects of L1 and L2 names (2-5D). It's not clear that these representations disagree, per se, but rather that 2-5C represents the objects provided in the stimulus set, while 2-
5D provides a controlled, parametric representation over many values of each variable. The discrepancy between these representations is further explored in the discussion.

**Chinese Category Native-likeness**

The same statistical models estimated for L2 naming were used again to test predictors of L1 native-likeness in lexical categorization. That is, I used regression models to find out which language history and linguistic community variables significantly predicted differences in lexical categorization between native Chinese monolinguals and Chinese-English bilinguals. As described in the Group-Level Analyses, unexpectedly high levels of convergence between L1 and L2 and unexpectedly low levels of L1 native-likeness were observed in the Chinese (L1) immersed group. However, it is presently impossible to determine whether these effects genuinely reflect cross-language convergence or the priming effect of the English naming task, which was completed before the Chinese naming task. Therefore, the following analyses of L1 lexical categorization are limited to the 33 participants tested in the United States, where at least one week intervened between Chinese and English picture naming tasks.

**Subject-wise Analysis: The Role of Language History Variables**

Participants' picture-naming responses were compared to the Chinese monolingual norms to generate a score for each participant describing the Chinese native-likeness of their lexical categories. As in the English native-likeness analysis, each of a participant was awarded a score based on the proportion of Chinese natives in the monolingual norm who produced the same name as the participant for each object (Malt & Sloman, 2003). I estimated a linear regression model for the Chinese native-likeness scores over participants' language histories to determine the
relationships between language background and change in L1 lexical categorization relative to the monolingual native norms. The initial model was estimated with all possible interactions (up to four-way) and yielded several highly significant three-way interactions, but the model was not significant overall \(F(15,17) = 1.27, p=0.323, \text{Adjusted } R^2=0.11\). The AIC search failed to identify any improvements in the model, and all lower-order models resulted in worse overall fitness. Finally, this analysis demonstrates that this set of language history variables is not suitable for explaining variation in L1 native-likeness. Although the model was only estimated among participants immersed in the second language, the same analysis was attempted with the combined set of participants and was similarly ill-fit \(F(13,48) = 0.86, p=0.596, \text{Adjusted } R^2=-0.03\). The following Item-Wise analysis tests whether a better account of L1 change may be found by examining individual responses to each object in the task.

**Item-wise Analysis: The Impact of Native Speaker Norms**

The first logistic regression identified each language history variable as a significant predictor, including all two-, three-, and four-way interactions. However, the Nagelkerke \(R^2\) for this model was very low (0.050), indicating that while the predictors were highly significant and did account for some overall trends in the item-wise data, they did not explain a large proportion of the trial-by-trial variation.

In the next analysis, the four categorization variables were added, and this model was estimated with up to four-way interactions and reduced using the same AIC stepwise search procedure described in the English analyses. The resulting reduced model improved the AIC compared to the initial model and included 39% fewer terms than the initial model without a serious decrease in fitness (initial model: 163 terms, \(AIC=2496.39, \text{Nagelkerke } R^2=0.41\); reduced model: 99 terms, \(AIC=2404.91, \text{Nagelkerke } R^2=0.39\)). Figure 2-6 illustrates the predicted L1
accuracy curves for each of the four categorization variables at two levels of Immersion. The No Immersion level is not included since data from the participants in Beijing were not included in the estimation this model.

Figure 2-6. Two-way interaction plots for Immersion and each of the categorization variables. (A) L2 native speaker agreement (percent of a norming sample who produced the dominant name) (B) L1 native speaker agreement. (C) Number of alternate names for an object produced by the L2 native speakers in a norming sample. (D) Number of alternate names produced by L1 norming sample.

With greater immersion duration, bilinguals were increasingly sensitive to L2 Agreement norms when naming objects in L1 (Figure 2-6A). Higher levels of English (L2) agreement supported more native-like Chinese (L1) naming, but lower English agreement resulted in lower accuracy for Chinese lexical categorization, demonstrating the influence of these L2 norms on the bilinguals’ categorization behavior in their native language. Conversely, sensitivity to Agreement
in Chinese was considerably attenuated with immersion (2-6B). Accuracy in Chinese naming appears to increase with immersion for the lowest levels of L1 Agreement, suggesting that the bilinguals tend to produce L1 dominant names more frequently with less of the naming diversity that characterizes monolingual categorization.

As in the English analyses, an unanticipated relationship between the number of alternate names and naming accuracy emerges. As the number of English names for an object increases, L1 accuracy decreases (Figure 2-6C), but the opposite relationship applies to the number of Chinese names (2-6D), which are positively associated with L1 accuracy. This pattern is a reversal of the effect observed in the L2 accuracy, such that the number of names in the target language (e.g., the number of Chinese names, when naming in Chinese) is positively associated with accuracy while the number of names in the non-target language (e.g., the number of English names when naming in Chinese) is negatively associated with accuracy. The statistical interpretation of these data must account for the correlation between these predictors, and as demonstrated in the L2 native-likeness analysis (see Figure 2-5) differing descriptive and statistical accounts may be presented: Objects with lower name agreement tend to also have a greater number of alternate names, confounding these two variables, but after controlling for variation in agreement, a greater number of alternate names in the produced language is actually associated with greater accuracy. This observation is consistent with both L2 production (wherein more English names are associated with greater native-likeness) and L1 production (wherein more Chinese names are associated with greater native-likeness), and possible causes underlying this relationship are explored in the discussion.
Discussion

Summary

In this study I examined the relative effects of four language history variables in predicting Chinese-English bilinguals' L1 and L2 lexical semantic native-likeness. Highly significant interactions were found between these variables in predicting L2 native-likeness, supporting the caution against evaluating any one factor of language history (e.g., age of L2 onset) in isolation from other variables. Significant age of L2 onset effects were observed, but these effects were tempered by the positive contribution of increased immersion experience. I made the surprising observation that increased experience with L2 prior to immersion was actually detrimental to L2 lexical semantic native-likeness in L2. I found that for bilinguals with long-term L2 immersion, patterns of language use were a significant predictor of L2 native-likeness, but for bilinguals with less immersion experience (including no immersion experience) language use was a less important predictor of L2 native-likeness. These language history variables did not significantly predict variation in L1 lexical semantic native-likeness between participants, and they accounted for a very small portion of variance item-by-item apart from interaction with L2 native speaker norms.

The naming norms of the native speakers of both languages influenced the bilinguals' success in acquiring native-like L2 lexical semantic mappings and the native-likeness of the L1 mappings for L2 immersed bilinguals. Both first language (Chinese) and second language (English) norms were significant predictors of the bilinguals' L1 and L2 native-likeness, consistent with previous findings in other domains of second language acquisition, such as phonology. Further, I identified unique effects for agreement among native speakers and the number of alternate names produced in the norming samples. The result of an item-wise analysis
revealed that a large amount of the between-object variation in naming was captured by these native speaker norms, indicating both the lasting importance of L1 mappings in L1 and L2 production and the sensitivity of bilinguals to the native speaker norms of the L2 in shaping their own L1 and L2 mappings.

While significant interactions between the bilinguals' language variables and language norm variables appeared in the item-wise analysis of the L1 data, the language history variables on their own (both main effects and interactions) were not significant predictors of the individual bilinguals' L1 native-likeness and accounted for very little of the item-wise variation in L1 naming. An extended discussion of the role of bilinguals' language history variables is, therefore, reserved for the General Discussion. The present discussion more closely examines the role of the four language norms (L1 and L2 agreement and names) in predicting L1 and L2 native-likeness and sets the stage for the chapters to follow.

**Name Agreement**

Overall, I found a competing relationship between the level of native-speaker agreement in L1 and L2 in predicting the native-likeness of bilinguals' L2 responses, but a mutually supportive relationship of L1 and L2 agreement norms in L1 native-likeness. In both languages, participants were sensitive to—and supported by—agreement among native speakers of that language for naming a given objects. The reversal occurs in the role of the non-target language: When naming objects in L2, bilinguals found greater L1 agreement problematic for finding the most native-like L2 name. When naming objects in L1, bilinguals were more likely to produce native-like L1 names when L2 native speakers also showed high agreement and decreased in performance as L2 agreement decreased.

The significance of L2 agreement norms in bilinguals' responses in both L1 and L2
production indicates that variation in L2 native speakers' lexical categories are an important influence in shaping the bilinguals' lexical semantic mappings in both languages. If bilinguals relied only on the majority or dominant name for objects, one should see little effect of the L2 agreement variable, but when naming objects in both languages, bilinguals respond proportionally to native speakers' level of naming agreement. The interaction between immersion and L2 agreement norms for naming in both languages demonstrates that over a period of prolonged L2 immersion, L2 native speaker norms (in this case agreement) become increasingly important in shaping the lexical semantic mappings of both L1 and L2.

These results are suggestive of an adaptive approach to managing lexical semantic mappings across languages. Sequential bilinguals begin the acquisition of L2 lexical categories with an established and relatively elaborate structure of lexical semantic representations that reflect the variability of L1 native speaker norms. As demonstrated in the unimmersed Chinese-English bilinguals, L2 mappings mainly reflect L1 norms in the absence of native-like L2 input (see Figure 2-2), and this reliance on L1 representations is further evidenced by the deleterious effect of L2 classroom training on L2 category native-likeness. These relationships are summarized in Table 2-2, describing the influence of each native-speaker norm on the native-likeness of the bilinguals in L1 and L2.

Table 2-2. Summary of each category variable predicting bilinguals' native-likeness in L1 and L2 lexical categorization (picture naming task). Direction of relationship denoted by + or –.

<table>
<thead>
<tr>
<th>Native Norm</th>
<th>Predictor</th>
<th>Native-likeness in: L1 (Chinese)</th>
<th>L2 (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 (Chinese)</td>
<td>Agreement</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Names</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>L2 (English)</td>
<td>Agreement</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Names</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>
The item-wise analyses of L1 and L2 production portray L1 agreement as a source of inertia (resistance to change) in lexical semantic mappings, supporting L1 native-likeness and decreasing L2 native-likeness. In this sense, L2-immersed bilinguals attempt to accommodate L2 input to their existing categories to some extent. However, with greater durations of immersion, more input from native speakers of L2 is added to the lexical semantic representations, which must eventually accommodate these novel mappings. This gradual trajectory of L2 accommodation is illustrated in the significant role of immersion in (1) increasing overall L2 native-likeness and decreasing overall L1 native-likeness, (2) the increasing sensitivity to L2 norms in L1 and L2 production, and (3) the decreasing influence of L1 norms in both L1 and L2 production.

Finally, the statistical models did not indicate a strong interaction between L1 and L2 agreement levels in L2 naming, instead identifying the unique main effects of each but only a very small (though significant) interaction term. This relationship is descriptive of two relatively independent inputs that additively compete in predicting bilinguals' naming responses (and likewise in shaping lexical semantic representations). It remains to be seen whether the observed small interactions are a byproduct of the objects in this particular task or whether the model underestimates the importance of this interaction. In either case, the important role of L1 and L2 agreement norms are apparent, either independently or interactively.

Alternate Names

The number of alternate names for an object produced by native speakers in L1 and L2 were also significant predictors of the bilinguals' native-likeness in each language. The direction of the relationship between the number of names in each language and the native-likeness of the bilinguals' responses depended on the target language. After accounting for agreement, an
increasing number of alternate names in the produced language was associated with greater native-likeness, while an increasing number of names in the non-target language was associated with diminished native-likeness. This relationship held for both L1 (Chinese) and L2 (English), suggesting that the effect was not limited to learning (L2) or attrition (L1) only. See Table 2-2 for a summary of these relationships.

One potential explanation for this effect is the proportion of input that each alternative name composes in the target language. Because the present model accounts for both dominant name agreement and the number of alternate names, the latter provides an indirect measure of the distribution of native speakers over the non-dominant names. If the number of alternate names in the native speaker sample increases, the proportion of the native speakers who produced each name decreases relative to the dominant name, thus reducing the salience of each alternate name as a competitor. Under this explanation, one would predict the lowest performance to occur when naming agreement is low and split with only one alternate name which shares all of the remaining native speaker agreement, e.g., 60% of native speakers name an object plate while 40% call the object dish. As new alternate names are introduced, the second- through nth-most dominant names fall off in agreement, e.g., 60% of native speakers call an object plate, while 20% call the object dish, and 20% call it platter. This account of agreement and alternate names emphasizes the competition between the alternate names. Successful native-like naming is supported not only by greater agreement but also by the relative dominance of the dominant name over the alternate names. This relationship was replicated within-language for both L1 and L2 production in the bilinguals, but was not found for non-target language names.

Even after accounting for agreement, a greater number of alternative names in the non-target language was still negatively related to native-likeness in the target language. If only observed in L2 production, this relationship might suggest that more names in the native language (L1) indicates a basic uncertainty about the identity of the object underlying the difficulty in
naming it. This explanation is inadequate, however, as the relationship reverses and in L1 production, the number of L2 names proves a significant negative predictor of L1 native-likeness. The fact that the predictor switches directions when applied to the target language (L1 names to L1 production) focuses any explanation on cross-language competition rather than basic conceptual uncertainty.

The interaction between L1 and L2 names in L2 naming may shed some light on this complex relationship. When bilinguals had only one name for an object in L1, they were equally likely to produce the dominant L2 name, regardless of the number of L2 alternatives. This relationship highlights the attraction of one-to-one translations, which may enable bilinguals to ignore competing names in favor of a single L1-to-L2 word-to-word mapping. However, the lowest probability of producing a native-like L2 name occurred when the target language (L2) provided two alternate names, with greater L2 native-likeness occurring when only one L2 name was available or when three or more L2 names were available. This pattern also suggests the attraction of the one-to-one translation in naming, and bilinguals struggle with increasing numbers of translation candidates activated by the single, dominant name in the target language.

To summarize, (1) more alternate names in the target language spreads the competition thinner and increases the relative salience of the dominant name, but (2) each additional name in the non-target language is activated as a translation candidate for the dominant name. Thus (after controlling for both L1 and L2 agreement) a greater number of target language names reduces competition and a greater number of non-target language names increases competition.
Chapter 3

Variables in Chinese and English Lexical Categorization

Rationale & Hypotheses

The foregoing study described a complex and interactive relationship between language experience and language norms of both L1 and L2 in predicting bilinguals' lexical categorization in both native and second languages. The effects of language experience and language norms were derived from a statistical model of bilingual participants' object naming responses to a relatively limited stimulus set of 67 photographs of common serving vessels such as cups and plates. This small stimulus set, drawn for a single semantic domain (dishes), may not represent the broader set of concrete objects that bilinguals encounter and name on a regular basis. For example, concrete objects may vary in complexity (e.g., a sophisticated aircraft compared to a tea cup) or familiarity (a specialized tool compared to a common shoe). In the present study, I assemble a large set of photographic stimuli across multiple semantic domains likely to represent a broad range of levels for these and other relevant category variables.

The stimulus set presented in this chapter uniquely attempts to sample a diverse set of semantic domains. Previous lexical categorization studies have identified cross-language variation in several types of concrete object, such as furniture (Graham & Belnap, 1986), containers (Malt et al., 1999; Ameel et al., 2005, 2009), and dishes (Ameel et al., 2005, 2009; Pavlenko & Malt, 2011; Chapter 2 of the present dissertation). Recent studies have compared apparent language transfer effects in Chinese-English bilinguals' categorization of bottle-like containers and dishes (Malt et al., 2013), but these studies have neither pooled the stimulus sets to observe broader patterns nor sampled outside the generally familiar domain of household objects.
As described below, the present stimulus set includes many examples of household objects as well as objects which are typically found in a workplace (tools) or outdoors (vehicles). Combining these semantic domains in a single stimulus set (1) provides a broader, more ecological perspective of lexical categorization (2) reduces category or name repetition, which are shown to affect participants’ responses in semantic blocking studies (e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006), and (3) provides a sufficient number of stimuli required for statistical inferences about neuroimaging data, in service of the neurocognitive studies to follow in Chapters 4 and 5.

As in the previous lexical categorization study, participants are shown photographic stimuli on a computer monitor and asked to produce the name they think best fits the presented object. In addition to the naming data, I solicit two additional ratings for each object based on previous picture naming studies: concept familiarity and exemplar typicality. In previous experiments, concept familiarity has been normed using individual exemplars of a given concept (e.g., one image of an owl used to evoke a familiarity rating of the concept owl; Wilson et al., 2009). It is unclear, however, to what degree this measure is dissociable from the respondent's judgment of the particular exemplar's typicality of the underlying category. In other words, would an unusual owl be rated differently for concept familiarity than a very typical owl, even if they were intended to evoke the same lexical concept? To clarify this relationship, norming participants are asked to rate each object in the set for either concept familiarity or exemplar typicality. In this way, I account for an important confounding variable (ostensibly, concept familiarity) and dissociate it from exemplar typicality, which operates at the lexical semantic level by measuring an object's fitness to the implicit set of same-named objects—that is, its lexical category.
While the present study is principally descriptive in nature, I propose a few important hypotheses that will motivate the analysis of the English and Chinese norming data and enable the use of this stimulus set in successive neuroimaging studies.

1. Each semantic domain should elicit a wide range of values across all four variables of interest: name agreement, alternate names, concept familiarity, and exemplar typicality. Semantic domains may differ from one another significantly over these variables, emphasizing the importance of taking a broad sample of objects for lexical categorization research.

2. In the previous behavioral study, name agreement and alternate names were significantly correlated with each other but produced differentiable effects in bilinguals' naming. A similar correlation is likely to emerge in the present stimulus set within each language group.

3. The present norming data should also provide a means of dissociating concept familiarity and exemplar typicality, determining whether they are unique variables and whether they operate at different levels of naming (conceptual vs. lexical). If concept familiarity indeed operates at the conceptual level, it should be more consistent between objects within the same lexical category while exemplar typicality varies from object to object within a category.

4. Finally, the three category variables predicted to measure lexical semantic representations (name agreement, alternate names, and exemplar typicality) should demonstrate greater linguistic relativity than the concept familiarity, which may vary somewhat by culture- or geographic-specific experience but not rely on linguistically influenced intuitions about object names.
Method

Participants

Native speakers of English and Chinese were recruited for this experiment. Twenty native English monolinguals (18 F / 2 M) at Penn State University and 24 native Chinese (Mandarin dialect) speakers (18 F / 6 M) at South China Normal University participated. Participants were comparable in age (Penn State: 19.4 y; SCNU: 21.0 y) and had modest experience with foreign language and no foreign language immersion experience.

Language status was verified by a brief version of the Language History Questionnaire (Li, Zhang, Tsai, & Puls, 2013) in which participants self-reported their histories of language exposure and proficiency. Different criteria were used for monolingual status in each sample as the United States and China differ in their typical second language education requirements. Both the Chinese and English samples had never lived or traveled for an extended period of time outside of China or the United States, respectively. See Table 3-1 for participant demographics.

Table 3-1. Participant demographics (means and standard deviations) for native English and Chinese norming samples.

<table>
<thead>
<tr>
<th>Native Language</th>
<th>n (M/F)</th>
<th>Age (sd)</th>
<th>L2 AOA (sd)</th>
<th>L2 Prof (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>20 (2/18)</td>
<td>19.4 y (1.6)</td>
<td>14.2 y (3.0)</td>
<td>2.2 / 7 (0.95)</td>
</tr>
<tr>
<td>Chinese</td>
<td>24 (6/18)</td>
<td>21.0 y (1.8)</td>
<td>9.2 y (2.4)</td>
<td>3.8 / 7 (0.78)</td>
</tr>
</tbody>
</table>

Participants at Penn State had two or fewer years of study in any foreign language. Average self-rated proficiency in their foreign languages ranged from 1 (Very Poor) to 4.25 (Functional) out of 7, with a mean rating of 2.21. All but one participant began studying a foreign language (if any) at age 12 or later, mainly or exclusively in a classroom setting. One participant began foreign language study at age 5 as part of primary education, but self-rated her proficiency as 2.5 out of 7. Of the 20 native English speaking participants, 15 had studied Spanish as their
foreign language and the remainder studied Arabic, French, German, Italian, and Latin. None of the participants reported any experience with Chinese.

By contrast, most of the Chinese participants began studying English at the primary education level. Participants at SCNU thus reported earlier exposure to English as a foreign language (mean age 9.2 years; earliest reported, 4 years), and none of the participants reported knowledge of any additional languages besides Chinese and English. Self-rated proficiency in English ranged from 2 (Poor) to 4.75 (Good), with a mean rating of 3.81 (Fair). Importantly, despite their classroom English experience, participants were unlikely to have much knowledge of English-specific lexical categorization norms according to the findings of Chapter 2.

Materials

Language History Questionnaire

As described in the foregoing section, participants completed the brief version of Li and colleagues' (2013) Language History Questionnaire (LHQ). The abbreviated form of this LHQ was administered in print in the participants’ native language (English or simplified Chinese script), and it required about 10 minutes to complete. The questionnaire asks for basic biographical information (e.g., age, sex, education), any experience living in a foreign country, native language, and details of any second languages they may have studied (age of onset, learning environment, and self-rated proficiency in all reported languages). Self-rated language proficiencies are reported on a scale of 1 (Very Poor) to 7 (Native-like)
Photographic Object Stimuli

Participants evaluated a set of 407 photographic stimuli depicting objects across approximately 40 lexical categories. The exact number of lexical categories would finally be determined by the participants’ responses, as some items that were initially categorized separately by the experimenter were given the same names by participants, and items that were categorized together by the experimenter were given different names by participants. Categories were drawn from six general semantic domains: clothing, dishes, tools, vehicles, food, and furniture. Within each semantic domain, I sampled lexical categories that would be familiar to native speakers but have enough within-category diversity and between-category similarity to be likely to evoke some naming differences within the monolinguals and between monolinguals and bilinguals. At this stage, the selection process was achieved using informed intuitions about dominant and alternate names for objects in English and Chinese in consultation with Chinese-English bilinguals.

Photographic stimuli were compiled from photographs publicly available via Google Images (http://www.google.com/imghp), and digitally modified to isolate the object of interest from background or other extraneous objects. Although norming data were not available for these images in advance, each object was selected with the intention of sampling a relatively broad range of exemplars for the lexical category (e.g., a number of different types of wrench of varying shape, size, and function). I also aimed to sample the category boundaries, seeking a few exemplars in each category that could have relatively ambiguous category membership, especially in relation to other sampled categories (e.g., a wrench that might be very similar to pliers). Finally all photographs were cropped to a square aspect ratio, resized to 200x200 pixels (small enough to identify each object clearly and preserve a moderate level of detail), and compressed at 95% quality in JPEG format.
**Familiarity & Typicality Scales**

After each naming trial, participants were asked to make a concept familiarity or exemplar typicality judgment relating to the same picture they had just named. The exact distinction between these measures is relatively unclear at present (see Rationale), and the instructions and prompts for each type of judgment were designed to help disentangle these measures (if they are indeed different).

Initial instructions for rating the typicality of each object asked about the suitability of the depicted exemplar for the name that the participant provided: “You will be asked to rate how typical you think this photograph is for the name that you entered.” At each trial, participants were prompted “How typical is this [NAME]?” where the name they entered previously was echoed back to them and the photograph of the object was again presented on the screen. Ratings were made on a 4-point Likert scale ranging from “1 - Very Unusual” to “4 - Very Typical” to prevent attraction to a central value and force participants to choose based on both valence (1-2 vs. 3-4) and magnitude (1 vs 2 and 4 vs 3).

Initial instructions for rating the concept familiarity were drawn from previous picture-naming studies (Wilson *et al.*, 2009; Graves, Grabowski, Mehta, & Gordon, 2007; Fiez & Tranel, 1997) and were intended to focus the participant's attention on the concept underlying a given object: “Familiarity is the degree to which you come in contact with or think about the concept.” Due to uncertainty about the effectiveness of this prompt at accessing a concept as opposed to the exemplar presented, I prompted participants to make this judgment on every exemplar, allowing comparison across exemplars that shared a concept. At each trial, participants were prompted “How familiar are you with [NAME] in your everyday experience?” where the name they entered previously was echoed back to them and the photograph of the object was again presented on the screen. Importantly, this prompt made no reference to the photograph (ostensibly inquiring about
the underlying concept) but still presented the photograph for consistency with the previous studies that have used Familiarity ratings. Ratings were made on a 4-point Likert scale ranging from “1 - Very Unfamiliar” to “4 - Very Familiar.”

Procedure

After obtaining informed consent, the experimenter seated the participant at a lab table and asked the participant to complete the LHQ. After completion of the LHQ, the experimenter confirmed that the participants were eligible for the respective norming sample (monolingual English or near-monolingual Chinese native) and informed the participant that he or she would be naming and rating pictures of many objects. Participants were instructed to respond by typing their best guess for the name of every depicted object, even if they were unsure. The task was not speeded, but participants were advised not to dwell too long on any one image as there would be many.

Figure 3-1. Participant prompts for (A) naming (“CUP” entered by participant), followed by either (B) exemplar typicality rating or (C) concept familiarity rating.

Eligible participants were seated at personal computers and advised to read the directions presented by the computer very carefully. After the participant read the text version of the instructions and completed a practice trial (using an unrelated photograph of a cat), the experimenter left the room and the experiment began. Participants viewed the 200 x 200 pixel
images on a 640 x 480 display accompanied by the text prompt “What is this?” (English version) or “这是什么?” (Chinese version). Participants entered their responses by typing into a small text box beneath the photograph. See Figure 3-1 for an illustration of the picture naming and rating prompts.

After a name was entered, the picture remained on the screen and the text prompt changed to either the familiarity or typicality prompt and scale. To rule out order and fatigue effects, images were randomly sequenced within four subsets of 102 for each participant. Order of the four subsets was counter-balanced across participants. In the first session, participants completed two subsets, and after at least 48 hours, participants returned to complete the remaining two subsets. All participants completed the same judgment task (typicality or familiarity) across all four subsets, such that no participant saw both types of prompt. Half of the participants in each language sample rated typicality and half rated familiarity.

Results

English Categorization

Name Distributions

English responses were coded by two native English speakers who first worked independently and then cooperated to resolve ambiguous trials. Name responses for each trial were first checked for spelling and typographic errors. Misspelled responses that could be reasonably corrected (in the judgment of the coder) were repaired, and unrecoverable responses were deleted. Next, head nouns were extracted from each response to isolate the lexical category indicated by the participant from further details. When the contents of a container were named
instead of the object (e.g., “paint” instead of “can”), the response was recoded as a single category “CONTENT” so that variation in the content-type responses would not inflate the number of names. In most cases, head nouns were isolated by removing adjectives (e.g., “large bowl” becomes “bowl”) and prepositional phrases (e.g., “bowl with stripes” becomes “bowl”). In some cases, two plausible nouns would appear in a response. If one noun appeared in the prepositional phrase (e.g., “bottle of wine”), the first noun (“bottle”) was selected as head noun and the object of the proposition (“wine”) was removed. When compound responses were given (e.g., “table and chairs” or “cup and saucer”), the first noun (“table” and “cup”, respectively) was selected. In a few cases, responses were a compound of two suitable nouns (e.g., “jetplane” or “breadpie”). In these cases, the first noun was treated as a modifier and removed, leaving the second noun as head noun (“plane” and “pie”). Finally, a few lexicalized noun compounds were observed (e.g., “airplane” and “backpack”). These compounds were not reduced because developmental data (Bates et al., 2003) indicated that they were acquired as object names before the head noun element of the compound (“plane” or “pack”, respectively).

Using the object-by-participant matrix of head nouns for each language, name distributions were calculated for the images. These distributions report the proportion of participants in each norming sample who responded with any given name to each image. Dominant names are those that account for the greatest number of the responses to each image (excluding blank or unintelligible responses). The name agreement variable for each object was defined as this proportion for the object’s dominant name. Alternate names were computed by counting the number of unique head nouns produced for each image.
Ratings

Two variable values were calculated for each type of rating: Average Typicality (averaged across all typicality ratings), Conditional Typicality (averaged across only the subset of participants who provided the dominant name), Average Familiarity, and Conditional Familiarity (same method as Conditional Typicality). These ratings are compared to each other below in an effort to determine their relationships to one another and to the lexical categories produced by the norming samples.

Figure 3-2. Box plots for four English categorization variables in each semantic domain.

Distributions of the four categorization variables are illustrated in Figure 3-2. Agreement levels for each object in English ranged from 0.15 to 1.00, strongly weighted towards the upper end of the distribution with a median value of 0.80. The number of alternate names provided for
each object was distributed in the opposite direction, ranging from one name only (the dominant name) to 10 names, with a median of 3. Average Typicality ratings were distributed between 1.2 and 3.9 with a median rating of 3.0, and Conditional Typicality was slightly higher, ranging from 1.0 to 4.0 with a median of 3.2 (difference between the two scales was highly significant, paired \( t(406)=5.42, p<0.001 \)). Average Familiarity ratings were distributed between 1.1 and 4.0 with a median rating of 2.8, and Conditional Familiarity was also slightly higher, ranging from 1.0 to 4.0 with a median of 2.9 (the difference between the two scales was also significant, paired \( t(406)=2.02, p=0.044 \)).

**Semantic Domains**

These categorization variables were compared across the six semantic domains to examine how lexical categorization patterns may differ for different domains of objects, before within-domain variation is considered. For each variable, a one-way ANOVA was performed across domains. Dominant Name Agreement and alternate names did not significantly differ between domains (Agreement: \( F(5,401)=2.05, p=0.071 \); Names: \( F(5,401)=1.53, p=0.179 \)). Average Typicality and Average Familiarity, however, were highly significant between semantic domains (Typicality: \( F(5,401)=9.04, p<0.001 \); Familiarity: \( F(5,401)=75.60, p<0.001 \)).

Pair-wise \( t \)-tests (with Bonferroni correction for multiple comparisons) were used to identify which semantic domains differed from the others on Typicality and Familiarity. The Furniture domain produced the lowest Average Typicality scores, significantly lower than all other domains except Vehicles. Vehicles, in turn, was not significantly different than the next lowest domains (Dishes and Tools) but was significantly lower than Clothing and Food. No significant differences in Average Typicality were observed between Clothing, Food, Dishes, and Tools. The same pattern was observed for Conditional Typicality.
The Clothing domain was scored significantly higher than all other domains in Average Familiarity. Dishes, Food, and Furniture did not differ significantly from one another, but were both significantly lower than Clothing and higher than Vehicles and Tools. Finally, Vehicles and Tools were scored the lowest on Average Familiarity, significantly lower than all other domains but not significantly different from each other. The same pattern was observed for Conditional Familiarity, although the differences between some contrasts decreased below significance thresholds after Bonferroni correction.

**Correlations Between Variables**

All of the categorization variables were significantly correlated with one another, as depicted in Table 3-2 (all \( p < 0.001 \)). The number of alternate names for an object was strongly negatively correlated with the Dominant Name Agreement (Pearson's \( R = -0.74 \)). This correlation describes the trend that greater naming agreement for an object is associated with fewer total names, which is necessarily true at the extreme values. Objects with Name Agreement of 1.00 can have only one name, and for a sample of 20 participants, objects with a Name Agreement of 0.95 can have only two names or one name and one non-response. On the other end of the distribution, objects with a Name Agreement of 0.20 must have five or more names. However, the majority of objects lie somewhere in between these extreme values. When computed only for objects in the middle 50% of both distributions (ranging from the 25th to 75th percentiles), the correlation between Name Agreement and alternate names decreased to \( -0.42 \) (\( p < 0.001 \)). While the two lexical category variables are clearly related, they are not completely interchangeable.

The Average Familiarity and Typicality ratings were highly correlated with their respective Conditional scores (Pearson’s \( R = 0.95 \) in both cases), indicating that while limiting responses by the dominant name yielded small increases in each scale (Familiarity: 0.02,
Typicality: 0.05), it did change the overall distribution of scores such that the raw Average scores accounted for 90% of the variance in Conditional scores. Further, if Familiarity and Typicality sampled participants' judgments at the respective conceptual- and object-levels as intended, I would expect significant variation of Typicality among objects within each lexical category, as some objects are better suited (that is, more typical) to the category than others. By contrast, Familiarity should not vary within a lexical category if participants are actually rating the named concept rather than the depicted exemplar.

Table 3-2. Correlations between categorization variables in English norming data. All correlations significant at \( p < 0.001 \).

<table>
<thead>
<tr>
<th>Name Agreement</th>
<th>Alternate Names</th>
<th>Average Familiarity</th>
<th>Average Typicality</th>
<th>Conditional Familiarity</th>
<th>Conditional Typicality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name Agreement</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Alternate Names</td>
<td>-0.76</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average Familiarity</td>
<td>0.32</td>
<td>-0.38</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average Typicality</td>
<td>0.41</td>
<td>-0.54</td>
<td>0.47</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Conditional Familiarity</td>
<td>0.30</td>
<td>-0.33</td>
<td>0.95</td>
<td>0.42</td>
<td>1.00</td>
</tr>
<tr>
<td>Conditional Typicality</td>
<td>0.37</td>
<td>-0.49</td>
<td>0.42</td>
<td>0.95</td>
<td>0.38</td>
</tr>
</tbody>
</table>

**Typicality vs. Familiarity**

To test the contrast between the measures of Typicality and Familiarity, I compared the average variation in the ratings within each lexical category (as defined by the dominant names assigned to each object). The Conditional values of Typicality and Familiarity were compared in this test to limit data to those participants who indeed categorized the object according to its dominant name, and the standard deviation of Conditional Typicality and the standard deviation of Conditional Familiarity across objects within each lexical category were estimated. Only lexical categories with three or more exemplars were used (\( n = 43 \) categories), and a paired-\( t \) test
was performed to test whether the standard deviation of Conditional Typicality was higher in each category than the standard deviation of Conditional Familiarity. One should expect the standard deviation of Typicality to exceed the standard deviation of Familiarity if Typicality was a measure that varied from object to object within a category while Familiarity remained a stable measure across objects belonging to the same category. I found that Conditional Familiarity (mean \(SD=0.41\)) showed significantly less variation within lexical category than Conditional Typicality (mean \(SD=0.52\); paired \(t(42)=4.15, p<0.001\)). However, the variation in Familiarity was not negligible, indicating that this rating did vary considerably from object to object within a lexical category.

**Chinese**

**Naming Distributions**

Chinese responses were coded by a high proficiency Chinese-English bilingual (native Chinese) who was familiar with the experimental procedure and coding technique but had not been exposed to this stimulus set. As with the English data, name responses were corrected for spelling and typographic errors. Head nouns were extracted from each response according to the same rules used for English data. Compound words that resulted in low frequency head nouns after decomposition were treated as unique head nouns themselves. For example, \(fēijī\) (飞机) and \(zhīshēngjī\) (直升机) share the same root character (\(jī\) or 机), but this character (roughly translated, machine) rarely appears alone. Thus, \(fēijī\) and \(zhīshēngjī\) are treated as unique lexical concepts and are not further reduced. Name agreement and alternate names were calculated by the same method used for English data.
Ratings

Distributions of the four categorization variables are illustrated in Figure 3.3. Agreement levels in Chinese ranged from 0.23 to 1.00, strongly weighted towards the upper end of the distribution with a median value of 0.79. The number of alternate names provided for each object was distributed in the opposite direction, ranging from one name only (the dominant name) to 14 names, with a median of 3. Average Typicality ratings were distributed between 1.8 and 3.9 with a median rating of 3.3, and Conditional Typicality was slightly higher, ranging from 1.2 to 4.0 with a median of 3.4 (difference between the two scales was highly significant, paired $t(406)=4.93, p<0.001$). Average Familiarity ratings were distributed between 1.3 and 3.9 with a median rating of 3.3, and Conditional Familiarity was also slightly higher, ranging from 1.0 to 4.0 with a median of 3.3 (the difference between the two scales was also significant, paired $t(406)=4.81, p<0.001$).

Semantic Domains

Categorization variables were compared across the semantic domains to examine a priori differences in category variables, before within-domain variation is considered. One-way ANOVAs were performed across domains for each variable, revealing significant between-domain differences in each variable (Agreement: $F(5,401)=2.58, p=0.026$; Names: $F(5,401)=5.40, p<0.001$; Typicality: $F(5,401)=8.86, p<0.001$; Familiarity: $F(5,401)=35.40, p<0.001$).

Pair-wise $t$-tests (with Bonferroni correction for multiple comparisons) were used to identify which semantic domains differed from the others on each category variable. Although the ANOVA for Agreement produced a significant difference, none of the corrected pairwise
comparisons were significant. The Food domain had marginally less Agreement than the Vehicles domain ($p=0.051$). The Clothing domain produced fewest alternate names, significantly lower than the Food ($p<0.001$) and Tools ($p=0.045$) domains. The Food domain produced the greatest number of alternate names, significantly more than Furniture ($p=0.008$), Vehicles ($p=0.002$), and Clothing.

Figure 3-3. Box plots for four Chinese categorization variables in each semantic domain.

Significant differences were not observed between most of the domains for Average Typicality, except for Clothing that significantly exceeded all other domains ($p<0.01$ for all pairwise comparisons). Conditional Typicality ratings were similarly patterned, although
Clothing and Food did not significantly differ, and Food significantly exceeded Furniture ($p=0.047$).

The highest Average Familiarity ratings were also observed for Clothing, which was significantly greater than all other domains ($p=0.026$ vs. Food and $p<0.001$ for all others). Food, in turn, was significantly greater ($p<0.001$) than the remaining domains except Dishes, Dishes was significantly greater than Tools and Vehicles ($p<0.001$) but not Furniture, and Furniture was marginally greater than Tools ($p=0.068$) and significantly greater than Vehicles ($p=0.010$). Tools and Vehicles did not significantly differ. Conditional Familiarity ratings were similarly patterned with some variations in the significance of the pairwise comparisons.

**Correlations Between Variables**

As in English, all of the categorization variables were significantly correlated with one another, as depicted in Table 3-3 (all $p<0.001$). The number of alternate names for an object was strongly negatively correlated with the Dominant Name Agreement (Pearson's $R=-0.79$). When computed only for objects in the middle 50% of both distributions (ranging from the 25th to 75th percentiles), the correlation between Name Agreement and alternate names decreased to -0.45 ($p<0.001$), demonstrating that (as in English) these variables are not related but not interchangeable.

The Average Familiarity and Typicality ratings were highly correlated with their respective Conditional scores (Familiarity $R=0.91$, Typicality $R=0.86$) with only small differences in each scale (Familiarity: 0.05, Typicality: 0.06). While Familiarity and Typicality significantly correlated with one another, these correlations where considerably lower (0.57 for Conditional measures and 0.72 for Average).
Table 3-3. Correlations between categorization variables in Chinese norming data. All correlations significant at $p<0.001$.

<table>
<thead>
<tr>
<th></th>
<th>Name Agreement</th>
<th>Alternate Names</th>
<th>Average Familiarity</th>
<th>Average Typicality</th>
<th>Conditional Familiarity</th>
<th>Conditional Typicality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name Agreement</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Alternate Names</td>
<td>-0.79</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average Familiarity</td>
<td>0.37</td>
<td>-0.49</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average Typicality</td>
<td>0.44</td>
<td>-0.58</td>
<td>0.72</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Conditional Familiarity</td>
<td>0.30</td>
<td>-0.39</td>
<td>0.91</td>
<td>0.64</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Conditional Typicality</td>
<td>0.30</td>
<td>-0.39</td>
<td>0.61</td>
<td>0.86</td>
<td>0.57</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Typicality vs. Familiarity**

Standard deviations of Conditional Typicality and Conditional Familiarity ratings were again compared for each lexical category with three or more exemplars ($n=31$ categories) using a paired-$t$ test. Variability in the Conditional Familiarity measure (mean $SD=0.32$) did not significantly differ from variability of Conditional Typicality (mean $SD=0.33$; paired $t(30)=0.115$, $p=0.909$).

**English and Chinese Comparison**

The native norms from each sample are compared with one another to identify commonalities between the two languages in the overall patterns of these categorization variables. The present analysis correlates the English and Chinese norms across all 417 objects on the four categorization variables (with both methods of calculating Typicality and Familiarity). Highly significant correlations were found between English and Chinese ratings of all variables, but the sizes of these correlations varied considerably. Agreement and Names were the least
correlated between languages, with cross-language similarity accounting for 3% and 12% of the variance, respectively, for each measure. Typicality and Familiarity correlations, however, were more highly correlated between languages, accounting for 27% (Conditional Typicality) to 44% (Average Familiarity) of variance.

Table 3-4. Correlations between Chinese and English values for each category variable

<table>
<thead>
<tr>
<th>Category Variable</th>
<th>$R \ (p&lt;0.001)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name Agreement</td>
<td>0.18</td>
</tr>
<tr>
<td>Alternate Names</td>
<td>0.35</td>
</tr>
<tr>
<td>Average Typicality</td>
<td>0.61</td>
</tr>
<tr>
<td>Conditional Typicality</td>
<td>0.52</td>
</tr>
<tr>
<td>Average Familiarity</td>
<td>0.66</td>
</tr>
<tr>
<td>Conditional Familiarity</td>
<td>0.62</td>
</tr>
</tbody>
</table>

The cross-language correlations of the Average and Conditional calculations of the Typicality and Familiarity variables were compared according to Cohen and Cohen’s (1983) method for comparing correlation coefficients. Familiarity ratings were relatively independent of the produced name, such that cross-language correlations did not significantly differ ($p=0.34$) on the basis of whether all Familiarity ratings were used (Average) or only those from participants who produced the dominant name (Conditional). Cross-language correlation between Typicality ratings, on the other hand, decreased ($R_{\text{Average}} > R_{\text{Conditional}}$, $p=0.060$) when ratings were restricted to participants who produced the dominant name.

**Discussion**

**Name Agreement & Alternate Names**

The English and Chinese norming data collected for this stimulus set supported the hypothesis that category variables in each language would show considerable variability within each semantic domain. Name agreement was generally high for most stimuli (median values 0.79
and 0.80 in English and Chinese, respectively), but ranged from 1.00 (complete agreement among native speakers) to below 0.25 in both languages. There is some cause to question the lowest levels of name agreement due to this measure's sensitivity to the coding schema for each language which relied on some arbitrary rules about reduction of multi-word and compound responses to single head nouns. Nonetheless, as illustrated in Figures 3-2 and 3-3, the majority of name agreement scores exceeded 0.50 and thus indicated a majority agreement on the dominant name of a given object.

Alternate names were also expected to show a great deal of variability and be inversely related to name agreement, given that this pattern emerged in the more limited stimulus set of Chapter 2 and the general observation that lower name agreement indicates that a greater proportion of the native speaker norming sample produced one or more additional alternate names. As with name agreement, the range of alternate names was wide, from one to 10 (English) or 14 (Chinese), but with a median value of three in both languages. Again, the extreme values may be driven by idiosyncrasies of the coding schema and should not be interpreted as undue precision. Consequently, the most representative correlations between name agreement and alternate names should exclude these extreme values and the over-constrained condition where name agreement of 1.00 by definition elicits only one name. The correlations between the middle 50% of name agreement and alternate names data in each language were highly significant but only moderate in magnitude (Chinese: $R=-0.45$, English: $R=-0.42$).

The relationship between these variables posed an important statistical challenge in the previous behavioral study of bilingual lexical categorization. Simultaneously estimating the effects of name agreement and alternate names in the native-likeness (item-wise accuracy) of bilinguals' object naming responses yielded some counter-intuitive findings about the effect of having multiple alternate names for an object. In the previous chapter, I attributed some of these findings to the marginal effect of adding alternate names after controlling for name agreement,
specifically the increasing salience of a dominant name as the number of alternate names increases. The general pattern of correlation between name agreement and alternate names observed with the dishes dataset in Chapter 2 is still observed when sampling across a broader set of object stimuli from multiple semantic domains. On the one hand, this observation supports the assumption that these two variables will inevitably correlate as some level and that this relationship did not arise as a peculiar property of the previous stimulus set. On the other hand, the correlation between these two variables necessitates that subsequent object naming studies that use these variables account for their correlation and adjust analyses to adequately resolve the respective influence of each variable.

By way of example, Kan and Thompson-Schill (2004) present the only functional MRI account of name agreement effects to date, finding significant inferior frontal gyrus (IFG) activity in response to decreased agreement levels. Their study makes a significant contribution in casting a spotlight on IFG as a locus of control for lexical semantic ambiguity, but their stimulus set bears two important weaknesses relative to the present stimuli. First, they contrasted only “High” and “Low” agreement images which included a wide range of variation (0.33-1.00 but produced only a binomial comparison (Low > High) in which the mean name agreement value for Low (0.76) was not particularly low relative to my own stimulus set (median English name agreement 0.79). Second, Kan and Thompson-Schill did not also control for the number of alternate names available for each object. The study presented in Chapter 2 identified differentiable effects of these two variables, but it is not the earliest evidence that near semantic competitors have a measurable effect on the picture naming (see semantic interference: de Zubicaray et al., 2001). The neuroimaging studies which follow will draw upon this stimulus set and its elaborated measurement of category variables to improve control of the correlative relationship between name agreement and alternate names as well as the effects of additional variables described below.
**Concept Familiarity & Exemplar Typicality**

Measures of exemplar typicality and concept familiarity were also widely distributed between the anchor values (1.0 to 4.0) in both languages. Overall, concept familiarity and exemplar typicality tended to be somewhat high (median around 3 out of 4) in both languages. The distribution of concept familiarity also included the predicted differences between semantic domains, indicating that some domains (e.g., Clothing) were extremely familiar to participants while others (Tools and Vehicles) were significantly less familiar. This pattern held in both Chinese and English, suggesting that concept familiarity may be driven by shared experience with the depicted objects themselves and less by language-specific norms. Typicality ratings also varied within semantic domains, but significant differences in Typicality between semantic domains were fewer in both English and Chinese, suggesting that a sufficient cross-section of exemplars were available in the stimulus set, spanning comparable ranges of typicality in each domain.

Both measures (average and conditional) of concept familiarity and exemplar typicality were significantly correlated with each other and with name agreement and alternate names. Overall, the conditional measures of familiarity and typicality correlated less strongly with name agreement and alternate names. This observation is surprising because the conditional measures are specifically anchored to the naming responses (averaged only over participants who produced the dominant name), but the decreased correlation may be explained by instability in the average values for each object due to the smaller sample size (again, limited to those participants who produced the dominant name).

Alternately, removing variability due to the naming responses may reveal a unique predictor: Exemplar typicality ratings among only participants who produced the dominant name may be a better measure of an object's fitness to its lexical category than the overall average of...
that object's typicality ratings. Under this explanation, familiarity ratings should vary less between the average and conditional measures, if participants judge concept familiarity before associating the object with a lexical category. In general, the differences between average and conditional familiarity's correlations to name agreement and alternate names were less than the differences between average versus conditional typicality correlations with name agreement and alternate names. Further, in the Chinese norm, average and conditional familiarity are more highly correlated with one another than average and conditional typicality (although they were equal in the English norm). Overall, these observations are inconclusive but provide some suggestive evidence that (1) typicality ratings are more related to language norms than familiarity ratings, (2) conditional typicality ratings reveal something more than average typicality ratings about an object's relationship to its lexical category, and (3) average and conditional familiarity ratings should not, in principle, greatly differ but may modestly differ given that conditional familiarity is likely to have greater instability of object-wise estimates due to the reduced sample size.

The concept familiarity and exemplar typicality prompts in this study were also designed to elaborate on the measures taken by Wilson et al. (2009) in a similar study of picture naming. While their study purported to measure concept familiarity, the use of only a single exemplar of each concept (e.g., one image of an owl to represent the concept owl) makes it impossible to reliably distinguish between the participants' familiarity with owls as a category and their covert judgment of the particular owl exemplar's typicality of an unspecified category (possibly owls, birds, animals, etc.). In this study, two prompts were contrasted, one displaying the exemplar image but drawing participants' focus specifically to the lexical category by echoing their name response (“How familiar are you with OWL in your everyday experience?”) and the other explicitly drawing participants' attention to the exemplar image (“How typical is this OWL?”), eliciting the present measures of concept familiarity and exemplar typicality, respectively.
Comparisons of the English within-category variance demonstrated the greater sensitivity of exemplar typicality to object-by-object variation as compared to concept familiarity. This difference was not replicated in concept familiarity, where variance of the ratings within a lexical category did not significantly differ between typicality and familiarity. Cross-language comparisons, however, help to highlight the differences between these two variables and are discussed in the next section.

**Cross Language Correlation & Linguistic Relativity**

The Chinese and English native speaker norming data collected in the present study point to the varying degree of linguistic relativity (that is, dependence of certain perceptions or judgments on language norms) across the observed categorization variables (see Table 3-4 in Results). Previous cross-language categorization studies (e.g., Malt et al., 1999) indicate that there is a great deal of variation between languages in lexical category patterns, but that these patterns are not entirely linguistically relative (i.e., they are not arbitrarily and independently defined by each language). Malt et al. (1999) found that non-linguistic similarity judgments were highly consistent between speakers of different languages, pointing to largely language-independent perceptions of objects that can be marginally modified by linguistic information.

Cross-language correlation coefficients for each of the category variables serve as a rough measure of the degree of linguistic relativity contributing to each measure, and they offer a secondary test of the predictions about the relationship between explicitly language-based variables (name agreement, alternate names) and less clearly defined variables (concept familiarity and exemplary typicality). Lower correlation coefficients suggest independence of the two languages from one another or a greater degree of language dependence for the variable in
question. Using this information, one may again compare each of the category variables to draw some inference about their respective roles in picture naming, and thus lexical categorization.

Consistent with the overall findings of lexical categorization research, name agreement and alternate names proved to be highly language-specific. Name agreement had the lowest correlation, of negligible magnitude \( R=0.18 \), though statistically significant, accounting for less than 3% of variance. Besides establishing the cross-language differences in name agreement as a category variable, this low correlation also provides evidence to exclude alternate accounts of name agreement that would assert that name agreement effects are driven by objects' lack of fitness to intuitive (language independent) categories. Unpacking the cross-language correlation, this value indicates that some objects elicit a high level of name agreement among native speakers of one language while native speakers of the other language do not see the object as an obvious fit for a single agreed-upon category. The especially low correlation indicates that these levels of native speaker agreement vary widely and (relatively speaking) independent of one another in each language. See also the moderate (relatively low, compared to other variables) correlations between name agreement and the more language independent measure of concept familiarity in both languages (Tables 3-2 and 3-3). By this account, name agreement is a clearly language dependent variable minimally influenced by any universal or non-linguistic sources of information.

The number of alternate names for an object showed a more moderate degree of cross language correlation, which may be attributable, in part, to uncertainty about the identity of the object itself. As described in Chapter 1 of the present dissertation, Cheng et al. (2010) associated two distinct ERP components with different sources of participant uncertainty in a picture naming task: Alternate Names and Picture Uncertainty. While the first of these sources is clearly language specific (and may, in fact, represent aspects of both name agreement and alternate names in the present study), Picture Uncertainty may arise from either subjective uncertainty
about an object's purpose or identity (a likely correlate of concept familiarity) or the quality and detail of the image or presentation of that image. Images of insufficient resolution, contrast, or detail pose naming difficulty regardless of language, and these sources of uncertainty could inflate cross-language correlations. To point, the alternate names variable shows a great degree of language dependence (that is, a relatively small degree of cross-language correlation), and may be even more language dependent after accounting for the known language independent sources of variance described in previous studies (Cheng et al., 2010).

Finally, the low cross-language correlation values between name agreement and alternate names facilitates statistical differentiation between the influence of each language in a regression model of bilinguals' naming responses (as in Chapter 2) or functional brain activity (as will be discussed in Chapter 5). This statistical separability sets the stage for a step-wise regression procedure whereby effects of one language (that is, name agreement and alternate names variables in that language) are first estimated before entering variables of another language to test the second language's marginal influence beyond the specific effects of the first. Chapter 2 offered strong behavioral evidence that both languages bear on bilinguals' lexical categorization patterns in object naming. The statistical procedure described above allows a second test of this hypothesis by using the same variables in each language to predict bilinguals' functional brain activity during the naming task.
Chapter 4
Brain Correlates of Monolingual Lexical Categorization

Rationale & Hypotheses

The literature exploring functional brain correlates of the lexical semantic processes is vast, and several neurocognitive models of monolingual and bilingual word learning, representation, and production have been proposed in recent years. These models share many common assertions about the respective roles of spatially localized neuroanatomical structures in particular cognitive stages of word processing. As I have argued in this dissertation, however, the experimental work underlying these theories overlooks considerable sources of variation in lexical semantic representations. In the foregoing chapters, I have illustrated how category variables are both highly language- and object-specific (Chapter 3) and that these variables are important predictors of the dynamic cross-language interactions that shape lexical semantic representation in bilinguals (Chapter 2). In the following study, I set out to apply the predictions of current neurocognitive models to lexical categorization by identifying functional correlates of category variables in object naming and comparing these responses to the regions' proposed functions in the current theoretical framework.

Table 4-1 summarizes the anatomical regions involved in four rough stages of object naming starting from access of conceptual representations and proceeding to selection of a single name for production. These models differ in which underlying cognitive theories they use to compare with functional imaging data, and thus each model varies in the definition and sequence of various processes in naming. Price (2012), for example, adapts Petersen and colleagues' functional-anatomical model (Petersen, Fox, Posner, Mintun, & Raichle, 1988, 1989) to generate
a parallel set of cognitive and functional models for multi-modal word production. Price's model uniquely specifies visual form processing in conceptual access while other dominant theories attribute early conceptual activation to broadly distributed cortical activity. Indefrey (2011) draws direct parallels to the cognitive stages of the LRM model, and Hickok & Poeppel (2007) elaborate on the Geschwind model. Both of these models attribute conceptualization to a network of (unspecified) cortical regions and propose a subsequent link to the lexical semantic interface for word activation. Rodriguez-Fornells et al.'s (2009) model proposes a somewhat broader neurocognitive account by including learning and reasoning mechanisms, but it shares the dorsal versus ventral stream dissociations that underlie Geschwind-like models.

Table 4-1. Neurocognitive models of functional brain regions in word production. Stages are adapted from Levelt and colleagues' (1999) LRM model of production to accommodate the different cognitive models proposed in each paper. See Table 4.2 for abbreviations of anatomical structures. All structures refer to left hemisphere.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Association</td>
<td>vOT</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Conceptualization /</td>
<td>ANG, aMTG,</td>
<td>OL, TL</td>
<td>MTG, aITG, pITG</td>
<td>aMTG, aITS</td>
</tr>
<tr>
<td>Semantic Integration</td>
<td>aITG, aSTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Retrieval</td>
<td>pITG, pMTG,</td>
<td>mMTG</td>
<td>-</td>
<td>pMTG, pITS</td>
</tr>
<tr>
<td></td>
<td>MFG, POrb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Selection</td>
<td>MFG, POrb,</td>
<td>mMTG</td>
<td>vIFG</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PTri, ACC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A few generalizations about the functional correlates of lexical semantic processing can be drawn from the convergent views of these neurocognitive theories. First, pre-lexical (conceptual or semantic) processing is widely attributed to occipital and temporal structures with a particular role for the most anterior temporal regions in integration. Lexical semantic association, or lexical retrieval based on semantic/conceptual representations, is principally focused in middle or posterior temporal regions, with middle temporal gyrus (MTG) as a common denominator across many models. These models differ on the details of selection,
whether lexical concepts are narrowed at a semantic processing level (Price, 2012) or whether selection is postponed until after lexical retrieval (Indefrey, 2011; Rodriguez-Fornells et al., 2009). However, two models (Price 2012; Rodriguez-Fornells et al., 2009) point to inferior frontal structures in selection. Price's (2012) model elaborates on selection by specifying middle frontal gyrus (MFG) as the locus of control for lexical retrieval, pars triangularis (PTri) as a mechanism for semantic comparison and decision, and anterior cingulate cortex (ACC) specifically for suppressing competitor names.

Table 4-2. Legend of abbreviations for anatomical structure names.

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Subsections</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Anterior cingulate cortex</td>
<td>Anterior / middle / posterior / ventral subsections</td>
</tr>
<tr>
<td>ANG</td>
<td>Angular gyrus</td>
<td></td>
</tr>
<tr>
<td>IFG</td>
<td>Inferior frontal gyrus</td>
<td></td>
</tr>
<tr>
<td>ITG</td>
<td>Inferior temporal gyrus</td>
<td></td>
</tr>
<tr>
<td>ITS</td>
<td>Inferior temporal sulcus</td>
<td></td>
</tr>
<tr>
<td>MFG</td>
<td>Middle frontal gyrus</td>
<td></td>
</tr>
<tr>
<td>MTG</td>
<td>Middle temporal gyrus</td>
<td></td>
</tr>
<tr>
<td>OL</td>
<td>Occipital lobe</td>
<td></td>
</tr>
<tr>
<td>OT</td>
<td>Occipito-temporal areas</td>
<td></td>
</tr>
<tr>
<td>POrb</td>
<td>Pars orbitalis of the IFG</td>
<td></td>
</tr>
<tr>
<td>PTri</td>
<td>Pars triangularis of the IFG</td>
<td></td>
</tr>
<tr>
<td>STS</td>
<td>Superior temporal sulcus</td>
<td></td>
</tr>
<tr>
<td>TL</td>
<td>Temporal lobe</td>
<td></td>
</tr>
</tbody>
</table>

Extending the assertions of these models to lexical categorization, I offer a few basic predictions based on inference about the general role of category variables in the established processing stream for word production: First, functional correlates of the explicitly linguistic variables (name agreement and alternate names) should be consistent with the identified lexical semantic processing regions in neurocognitive models. Namely, name agreement and alternate names should elicit activity in left MTG and IFG regions for lexical semantic retrieval and
selection, respectively. The more precise roles of these anatomical regions may be explored by their unique correlations with variation in name agreement and alternate names. Although previous empirical work has specified IFG as a correlate of name agreement (Kan & Thompson-Schill, 2004), this work has not accounted for the strong correlation between name agreement and alternate names, the latter of which I propose is actually a better predictor of activity in IFG.

Second, the currently undifferentiated roles of exemplar typicality (linguistic) and concept familiarity (pre-lexical, possibly language independent) will be functionally dissociable. Concept familiarity, properly measured, should not predict variation within lexical categories and thus correspond to activity earlier in the ventral stream (occipital regions). Exemplar typicality, by contrast, is a language dependent measure will thus evoke more upstream processes in ventral occipito-temporal regions. If Wilson et al.'s (2009) measure of concept familiarity is actually a measure of exemplar typicality, the present study will replicate fusiform gyrus as a functional correlate of exemplar typicality.

Table 4-3. Hypothesized functional-anatomical correlates of category variables. See Table 4-2. for abbreviations of anatomical structures. All structures refer to left hemisphere.

<table>
<thead>
<tr>
<th>Category</th>
<th>OL</th>
<th>vOT</th>
<th>MTG</th>
<th>IFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept Familiarity</td>
<td>β &lt; 0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Exemplar Typicality</td>
<td>-</td>
<td>β &lt; 0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Name Agreement</td>
<td>-</td>
<td>-</td>
<td>β &lt; 0</td>
<td>-</td>
</tr>
<tr>
<td>Alternate Names</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>β &lt; 0</td>
</tr>
</tbody>
</table>

Table 4-3 outlines the four-way dissociation predicted for these category variables. In practice, most functional neuroimaging studies elicit multiple (and overlapping) functional correlates of each condition, but this representation provides a summary of the clearest hypothetical outcomes to be tested. Importantly, all of these predictions are directional for negative correlations. Wilson et al. (2009) found that their ostensible measure of concept familiarity was negatively related to ventral occipito-temporal activity, and this relationship has
thus been predicted for exemplar typicality in the present study. Similarly, low name agreement 
elicited greater activity in Kan & Thompson-Schill's (2004) study, and while I modify the 
localization of name agreement to MTG in the current hypotheses, the negative correlation is 
preserved. This negative relationship for MTG is further supported by semantic interference (de 
Zubicaray et al., 2001) and lexical semantic proficiency (Blumenfeld, Booth, & Burman, 2006) 
studies that find that greater MTG activity is associated with a broader or less proficient lexical 
semantic processing. Chapter 2 of the present dissertation suggested that the count of alternate 
names in the target language was actually an inverse measure of competition, and thus the greater 
number of alternate names is negatively correlated with activity in IFG.

Method

Participants

As in the previous experiment, native speakers of English and Chinese were recruited to 
participate. Each sample of native speakers was functionally monolingual, speaking only English 
(with minimal second language training and experience) or Chinese (Mandarin dialect, with 
minimal English training in their secondary education). Eight native English speakers (3 F, 5 M) 
at Penn State University and eight native Chinese speakers (4 F, 4 M) at South China Normal 
University participated in this MRI experiment. Participants were similar in age (Penn State: 23.9 
y; SCNU: 20.5 y).

All participants were right handed, assessed by Snyder & Harris’ (1993) Handedness 
Questionnaire, and scored 3.5 or higher out of 5 (English and Chinese group mean scores were 
both 4.5). An abbreviated form of the Language History Questionnaire (Li et al., 2013) was also 
used to document experience with foreign language (see Chapter 3). Native-English speaking
participants at Penn State were highly monolingual, with all participants reporting Spanish courses as their primary experience with foreign language. Mean age of onset for Spanish was 14 years old (minimum 8 years), and self-rated proficiencies were poor (mean 1.7 out of 7, maximum rating 2.3). Participants at SCNU began studying foreign language relatively earlier than their monolingual English counterparts (mean age of onset 9 years, minimum 5 years), and all studied English with no other reported foreign languages studied. Self-rated proficiencies were, nonetheless, only moderate (mean 3.4 out of 7, maximum rating 5), and Chinese participants reported learning English primarily or exclusively in classroom environments. Participants in both locations had never resided in a foreign language immersion environment.

Materials

Photographic Object Stimuli

An abbreviated set of 200 photographic stimuli was selected from the behaviorally normed set of 407 generated in the previous experiment (see Chapter 3). This stimulus set for the imaging study included objects from five of the semantic domains: clothing, dishes, tools, vehicles, and furniture. The food domain was excluded due to difficulty identifying dominant names in several lexical categories where the coding scheme used in Chapter 3 did not obviously result in a single head noun (“cane” for candy cane and “chocolate” appearing frequently as both adjective, “chocolate cake” and noun “chocolates”). Each of the included semantic domains contained five to seven lexical categories (as defined by the objects’ dominant names). For each of the 28 total lexical categories, there were between three and eleven objects in the stimulus set, depending on the availability of stimulus photographs meeting the criteria (see below for selection criteria),
such that no lexical category composed more than 6% of the entire stimulus set, minimizing name repetition in the task.

Objects for the imaging stimulus set were selected based on the English norming data gathered in the previous behavioral experiment. Semantic domains were approximately balanced for mean name agreement (hereafter, Agreement) and number of alternate names (hereafter, Names) in English (see Table 4-4) and to reduce the correlation between these two variables (see Table 4-5). Although Agreement and Names were not completely independent, a broad range of values was represented for each variable at each level of the other to improve estimation of each variable’s unique effects. Agreement values ranged from 0.40 to 1, with a single dominant name for each object. Names ranged from 1 to 6.

Table 4-4. Descriptive statistics for category variables in neuroimaging stimulus set. Median values presented instead of mean because variables were highly skewed from normal distribution.

<table>
<thead>
<tr>
<th></th>
<th>English Ratings</th>
<th></th>
<th>Chinese Ratings</th>
<th></th>
<th>Cross-Lang. Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min  Max  Median</td>
<td>Min  Max  Median</td>
<td></td>
<td></td>
<td>R   p</td>
</tr>
<tr>
<td>Agreement</td>
<td>0.40  1.00  0.85</td>
<td>0.42  1.00  0.88</td>
<td>-0.03</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Names</td>
<td>1    6     2</td>
<td>1    6     2</td>
<td>0.09</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Typicality</td>
<td>1.5  3.9   3.3</td>
<td>2.4  4.0   3.4</td>
<td>0.50</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>1.5  4.0   3.0</td>
<td>1.9  3.9   3.3</td>
<td>0.62</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
</tbody>
</table>

Seventeen objects in the imaging stimulus set did not produce a dominant name in Chinese, resulted in a Chinese name agreement below 0.40, or had greater than 6 alternate names. These items were removed from the imaging stimulus set, resulting in a final set of 183 objects with comparable ranges of Agreement and Names in English and Chinese, allowing for cross language comparison on naming of the same stimuli. Importantly, items were not selected for cross-variable balance in Chinese, but because the predictors of interest in the subsequent bilingual study would be the English variables, the stimuli were sufficient for controlling for effects of Chinese before estimating the effects of English.
The two remaining category variables, concept familiarity (hereafter, Familiarity) and exemplar typicality (hereafter, Typicality) were obtained from the norms measured in Chapter 3. Familiarity was measured using the complete set of ratings (average familiarity), minimizing language dependence and maximizing the number of observations included in the mean value. Typicality, on the other hand, was measured using the conditional typicality measure, only including ratings from participants who produced the dominant name, to better assess its relationship to the lexical category (that is, the typicality of the object to the dominant name).

Table 4-5. Within-language correlations (Pearson R) for category variables in neuroimaging stimulus set. **p<0.001, *p<0.05

<table>
<thead>
<tr>
<th>Chinese Correlations</th>
<th>Agreement</th>
<th>Names</th>
<th>Typicality</th>
<th>Familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td>1</td>
<td>-0.69**</td>
<td>0.15*</td>
<td>0.13</td>
</tr>
<tr>
<td>Names</td>
<td>-0.67**</td>
<td>1</td>
<td>-0.24**</td>
<td>-0.34**</td>
</tr>
<tr>
<td>Typicality</td>
<td>0.12</td>
<td>-0.15*</td>
<td>1</td>
<td>0.50**</td>
</tr>
<tr>
<td>Familiarity</td>
<td>0.19*</td>
<td>-0.22**</td>
<td>0.24</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English Correlations</th>
<th>Agreement</th>
<th>Names</th>
<th>Typicality</th>
<th>Familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Names</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typicality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Log-transformed word frequency norms for category names were entered as a control variable, available from Brysbaert and colleagues’ U.S. English and Chinese Film Subtitle Corpora (U.S. English: Brysbaert & New, 2009; Chinese: Cai & Brysbaert, 2010). Visual complexity was also entered as a control variable for each photograph, estimated by measuring the size of the 95% quality JPEG format image files in kilobytes (KB) after cropping and scaling all images to 200 x 200 pixels (see Bates et al., 2003 for rationale on measuring visual complexity by compressed image file size).
**Scrambled Image Stimuli**

Twenty-six scrambled images (two for the practice session and 24 for the MRI session) were used to help control for the effects of visual processing (without semantic content) in the functional brain response to photographs, following on the design of Wilson et al. (2009) who scrambled line drawings for line drawn picture stimuli. Scrambled images were generated by randomly sampling and assembling sixteen square tiles (50 x 50 pixels each) from a set of eight object photographs (also segmented into 50 x 50 pixel tiles) that were not included in the present study’s stimuli (see Figure 4-1 for example).

![Figure 4-1. (A) An object stimulus. (B) A scrambled stimulus.](image)

**Procedure**

Data collection was completed in two one-hour sessions, one day apart. In the first session, participants completed the screening forms (handedness and LHQ) and practiced naming photographic object stimuli in a simulated MRI environment. The practice task was composed of seventeen object stimuli and two scrambled stimuli, drawn from the items normed in Chapter 3 but excluded from the present study’s stimulus set. Participants were advised that they would be viewing a series of photographs of common objects while lying in the MRI and listening to the simulated noise of an EPI scan. Participants were instructed to name the objects aloud, quickly,
and with minimal head movement. Text instructions presented on the screen explained that participants should clearly say the name of each object, guessing if unsure, and that they should silently ignore any of the scrambled images. The practice task took about three minutes and was repeated two to three times until the participant was able to make verbal responses with minimal head motion.

In the second session, one day later, participants completed three more naming tasks while undergoing functional MRI. Each naming task was approximately 15 minutes long, composed of 61 object images and 8 scrambled images. Lexical categories were evenly distributed across the three sessions to minimize name repetition within each run. Images were presented for four seconds, but participants were encouraged to name the object within the first two seconds of presentation while the scanner paused to facilitate voice recording. After the initial two second silence, scanning would resume and voice recording would cease. Inter-stimulus interval was jittered between 0 to 32 seconds based on an optimized sequence generated by the optseq2 tool (Dale, 1999).

Instructions and photographic stimuli were presented using E-Prime 2.0 (Psychology Software Tools; Schneider, Eschman, & Zuccolotto, 2002) on a PC, projected onto a screen behind the MRI and visible to participants inside the MRI through a mirror mounted on the head coil. Voice responses were recorded for the first two seconds of each picture naming trial, during a programmed delay between scans. The present analyses, however, make use of all responses to include the production of non-dominant names for objects.

Image Acquisition

All MRI sessions in the US and in China were conducted on Siemens 3T MRI scanners. Structural T1-weighted images were composed of 160 sagittal slices, 1 mm thick with 50% over-
sampling (field of view: 25 6mm, TR: 1650 ms, TE: 2.01 ms, Flip angle: 9º). Following the structural scan, each participant completed three counterbalanced runs of picture naming while undergoing functional T2*-weighted EPI. Functional images were composed of 34 transverse slices interleaved (anterior to posterior), 3.8 mm thick (field of view: 240 mm, TR: 4000 ms, TE: 25 ms, Flip angle: 70º). Each functional measurement was completed in 2000 ms with an additional 2000 ms delay before the onset of the next measurement. This delay allowed naming trials to be synchronized with the scanner so that the first 2000 ms of each naming trial would be relatively noise-free for recording voice responses.

### Image Processing

All imaging data were processed using SPM 8 (Friston et al., 1995). Functional images were referenced to the middle slice in each scan, realigned to a mean image of the functional sequence with six motion parameters, and co-registered with the T1 structural image. Segmented structural images were regularized to European or East Asian templates for the monolingual English and Chinese samples, respectively, and normalized to an MNI template. Finally, normalized functional images were smoothed with an 8 mm FWHM Gaussian kernel.

### First-level Statistical Analysis

A general linear model was fit to each voxel’s activity to estimate the effect of object stimuli and scrambled stimuli over baseline in that region. In each of the three functional runs, object and scrambled stimuli were entered as separate predictors. Six motion parameters were also entered as nuisance regressors to reduce the effects of motion that were not removed by
motion correction. Separate regressors were estimated for each functional run and combined in the first-level effects estimates.

The two control variables (Visual Complexity and Frequency) and four category variables were entered as parametric modulators of the object stimuli, yielding an estimated beta value for each variable representing the change in signal as compared to the average (un-modulated) response for all object stimuli. Parametric modulators in SPM are entered by a step-wise orthogonalization procedure wherein the first modulator is estimated, and if the second modulator is correlated with the first, only the orthogonal component of the second modulator is estimated. This procedure reduces the statistical power for later-entered variables but is intended to improve the stability of estimates for the earlier-entered variables. All category variables were standardized by subtracting the mean value and dividing by standard deviation. Subsequent analyses made use of the first-level estimations and are therefore each described in the Results.

**Results**

**Overview**

In the following analyses, English and Chinese native speakers' results are examined in parallel, following the same procedures, and are not initially pooled for second level (group) analysis. In the first analysis, a second level model is estimated for each group to highlight anatomical regions of interest (ROI) whose functional responses to each variable pass a statistical threshold. This analysis is a standard procedure in many fMRI studies and produces useful information about the spatial distribution of functional responses, but it is sensitive to small sample sizes and large numbers of predictors (as in the present study) due to corrections for the large number of comparisons necessary for whole-brain analysis.
The second analysis targets *a priori* ROIs identified in the hypotheses. In this analysis, mean beta weights are extracted for regions of interest and compared for each category variable to depict the relative sensitivity of the ROI to each of the four variables of interest.

**Second Level Analysis**

A group-level T-map was estimated for each of the four category variables over the eight participants in each group. Because each variable is estimated as a modulator over the picture condition, coefficients describing these variables describe marginal effects over the average functional response for a picture naming trial. Thus, no subtractive contrast is necessary for identifying the main effect of each variable. Due to the relatively small sample sizes, none of these effects survived family-wise or false discovery rate corrections for multiple comparisons, but a more liberal threshold of $p<0.001$ (uncorrected) at each voxel revealed several clusters with significant relationships to the category variables (see Appendix A for complete list of significant clusters and peak voxels).

Qualitative comparison of these findings to the hypothesized relationships between ROIs and category variables yielded mixed results: No significant relationships were found between left MTG and any of the category variables. In Chinese, left IFG (*pars triangularis* and *pars opercularis*) was a negative correlate of Agreement, but this relationship was not observed in English.

Neither language group showed any negative correlates of Names, reversing compared to the behavioral finding that (after accounting for Agreement) decreasing the number of alternate names for an object would increase the difficulty of naming that object. In fact, many positive correlates of Names were identified in both languages. Left postcentral gyrus and occipital areas were positive correlates of Names in both Chinese and English. The anterior cingulate cortex was
also a positive correlate of Names, notable because activity in this region is predicted by two models of language production (monolingual: Price, 2012; bilingual: Abutalebi & Green, 2007; see Discussion for more detail).

Typicality effects were consistent with the hypothesized ROIs in Chinese, showing activity in the bilateral fusiform gyrii of in ventral occipito-temporal regions. English, however, elicited no negative correlates of Typicality. In both languages, right STG and left Cuneus were positive correlates of Typicality.

In both languages, occipital regions were positive correlates of Familiarity, although the observed clusters were not overlapping between languages. Negative correlates of Familiarity were not consistent with the hypothesized activity in occipital lobe. Several clusters in right MTG were negatively related to Familiarity in English, and the right insula was negatively related to Familiarity in both languages.

While these qualitative observations were only partly consistent with the hypothesized relationships, the lack of correction for multiple comparisons makes it difficult to discriminate between real correlates and false positives (that is, Type 1 errors) in the broad array of regions significant at the \( p < 0.001 \) (uncorrected) level. Further, this threshold may still exclude some weaker correlates that were statistically significant in this sample size. In the following section, however, the relative importance of each category variable is compared in only the four \( a \ priori \) hypothesized ROIs. Constraining the analysis to these regions reduces the need for statistical correction and allows a more direct comparison with previous empirical research.

**ROI Analyses**

Four spheres were defined in MNI space for the ROI analysis, based on the findings of previous research linking functional activity in picture naming to semantic or lexical cognitive
processes. Using ROIs identified in previous studies provides an objective, hypothesis-driven definition for identifying functional correlates of the category variables defined in the present study. Because these regions are defined *a priori*, the need for correction of Type 1 error is drastically reduced and the probability of observing significant effects is greatly increased. Four spheres in MNI space were thus defined to represent each of the ROIs, using peak voxels identified in previous literature as centroids and extended 5 mm in radius for all three dimensions.

Occipital and occipito-temporal regions, hypothesized to correspond to Familiarity and Typicality ratings, respectively, were drawn from Wilson et al.'s (2009) study of picture naming which recently described the Familiarity effect in picture naming. To characterize visual processing of object stimuli, I selected a voxel in the left calcarine fissure of the occipital lobe with peak response for picture trials as compared to scrambled trials in Wilson and colleagues' study. This voxel provided the centroid for a sphere in the occipital lobe. The peak voxel observed by Wilson *et al.* in left fusiform gyrus was the centroid of a sphere in the ventral occipito-temporal regions. See Table 4-6 for exact coordinates and Figure 4-2 for illustration.

Table 4-6. Locations of spheres in anatomical and MNI space representing each ROI. Studies providing peak voxel coordinates for each sphere centroid cited under Source.

<table>
<thead>
<tr>
<th>ROI</th>
<th>Anatomical Structure</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>radius</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>OL</td>
<td>L Calcarine Fissure</td>
<td>-6</td>
<td>-6</td>
<td>-6</td>
<td>5 mm</td>
<td>Wilson <em>et al.</em>, 2009</td>
</tr>
<tr>
<td>vOT</td>
<td>L Fusiform Gyrus</td>
<td>-38</td>
<td>-54</td>
<td>-22</td>
<td>5 mm</td>
<td>Wilson <em>et al.</em>, 2009</td>
</tr>
<tr>
<td>pMTG</td>
<td>L Middle Temporal Gyrus</td>
<td>-56</td>
<td>-20</td>
<td>-14</td>
<td>5 mm</td>
<td>de Zubicaray <em>et al.</em>, 2001</td>
</tr>
<tr>
<td>IFG</td>
<td>L Pars triangularis</td>
<td>-36</td>
<td>24</td>
<td>21</td>
<td>5 mm</td>
<td>Kan &amp; Thompson-Schill, 2004</td>
</tr>
</tbody>
</table>

Activity in left MTG is observed in many studies of lexical semantic processing, but a peak voxel from de Zubicaray *et al.*'s (2001) study of semantic interference was selected for being particularly associated with spreading activity through a semantic network. Importantly, this cluster falls close to the middle portion of MTG, selected to help isolate it from effects in
anterior temporal lobe (multi-modal semantic integration), posterior superior temporal regions (angular gyrus and pSTG associated with semantic integration and lexical phonology processing, respectively), and lateral occipito-temporal regions (with more superior regions not clearly on the ventral stream and more inferior regions intersecting the vOT).

Figure 4-2. Locations of ROI spheres in four anatomical regions of interest. (A) Calcarine fissure in OL, (B) Pars triangularis in IFG, (C) Fusiform gyrus in vOT, (D) Middle temporal gyrus. See Table 4-6 for exact coordinates and details.

Finally, Kan and Thompson-Schill's (2004) study of name agreement was selected to provide the centroid for IFG activity because the present study compares the effects of Agreement (measured in Kan & Thompson-Schill, 2004) and Names in this region with the purpose of disentangling these variables to a greater degree than attempted in prior research. Kan and Thompson-Schill identified PTri as the locus of the agreement in effect in naming photographic object stimuli in their second experiment (2004; in contrast to line drawing object stimuli in the preceding experiments of the same study).

Mean standardized beta values in the ROIs were extracted for each participant in the Chinese and English groups using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002) for SPM. Figure 4-3 depicts the mean betas in each ROI, for English and Chinese values of each variable. These values were entered in a 4 x 4 x 2 (Variable x ROI x Language) repeated-measures ANOVA to test for significant variation in the category variables' respective beta weights in each region. Variable and ROI were entered as within-subject factors.
Figure 4-3. Standardized beta values for each category variable by language in four regions of interest. Note: Error bars indicate one standard error of the mean. * denotes significant pairwise differences between category variables.

Because Names was significantly, negatively correlated with the other three category variables (see Method), all beta values for Names were reversed in sign (+/-) before being entered in the ANOVA to improve the simultaneous comparison of each variable's effect. For example, when Agreement is low for a particular object stimulus, Names tends to be high, and the magnitude of each oppositely-signed beta should be compared to infer the relative effect of these two variables.

A significant interaction between Variable and ROI \((F(9,126)=3.22, p=0.002)\) indicated that the category variables responded differently in the four hypothesized regions of interest. Table 4-7 reports these mean betas, in the same arrangement as the hypothesized results presented in Table 4-3 (Rationale & Hypotheses). The main effect of Language \((F(1,14) = 0.75, p=0.402)\) and the three-way Language x Variable x ROI interaction \((F(9,126) = 0.311, p=0.970)\) were not
significant, indicating that mean betas in Chinese and English did not significantly differ from
one another. Therefore, beta values were combined to estimate each Variable x ROI mean,
helping to offset the small sample sizes in each language group.

Table 4-7. Standardized beta values for each category variable in four regions of interest. Compare to Table 4-3 in Rationale section.

<table>
<thead>
<tr>
<th>Category Variable</th>
<th>OL</th>
<th>vOT</th>
<th>pMTG</th>
<th>IFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarity</td>
<td>1.10</td>
<td>-0.48</td>
<td>0.14</td>
<td>-0.75</td>
</tr>
<tr>
<td>Typicality</td>
<td>-0.45</td>
<td>0.51</td>
<td>0.01</td>
<td>-0.19</td>
</tr>
<tr>
<td>Agreement</td>
<td>0.77</td>
<td>0.19</td>
<td>0.31</td>
<td>-0.34</td>
</tr>
<tr>
<td>Names</td>
<td>1.02</td>
<td>0.62</td>
<td>0.21</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Two significant pairwise contrasts (after Bonferroni correction for multiple comparisons in each ROI) were observed: between Names and Typicality in L Calcarine Fissure (OL) and between Names and Agreement in L pars triangularis (IFG). Although other pairwise comparisons were not significant, an overall trend emerged whereby the category variable predicted to be significant in each ROI (according to Table 4-3, Rationale) had the largest mean beta (in magnitude, sign notwithstanding) in three of the four ROIs (Familiarity in OL, Agreement in pMTG, and Names in IFG).

Discussion

Based on the consensus of neurocognitive models of word production, I generated a four-way dissociation hypothesis about the functional-anatomical correlates of each category variable. Identifying brain regions associated with these category variables can provide confirmatory evidence of the variable's role in lexical semantic processing and elaborate on the roles of these brain regions in processing object stimuli for naming tasks. The group analysis did not yield significant results in this study, primarily for a lack of statistical power. However, a more targeted
ROI-based analysis provided valuable links between the present study and preceding empirical studies of picture naming and word production. These ROI results are discussed below in relation to the previous empirical research most closely related to lexical categorization and the existing neurocognitive theories of word production.

**ROIs and Categorization**

The ROI analysis identified significant variation in localized, functional responses to the category variables. Although the category variables were moderately to strongly correlated with one another, some of their effects were statistically dissociable in this picture naming task. In addition to the four-way Variable x ROI dissociation hypothesis, I proposed two specific comparisons to previous studies involving category variables: (1) The effects of Names and Agreement would be dissociable by their specific correlations with middle temporal gyrus and the *pars triangularis*. (2) Typicality would be a stronger correlate of activity than Familiarity in ventral occipito-temporal regions such as left fusiform gyrus.

In the first hypothesis, I aimed to dissociate two highly correlated and language dependent category variables. Names and Agreement were strongly correlated in both languages (English: $R=-0.67$, Chinese: $R=-0.69$, $p<0.001$ in both cases) but were predicted to have different functional correlates in monolingual lexical categorization based on their unique effects on bilingual lexical categorization in Chapter 2. In the present study, I also directly test Kan & Thompson-Schill’s (2004) Agreement effect in IFG by first accounting for Agreement and then entering the number of alternate names proposed by native speakers in the norming set (that is, Names). Testing an ROI at the same coordinates specified by Kan and Thompson-Schill, I found that Names was a significantly stronger predictor than Agreement to the functional brain response in this region. This finding supports the conclusion that left IFG operates as a selection
mechanism among alternate names, and it achieves this result with a more direct test of selection than the prior study.

The complementary test of Agreement as a unique correlate of activity in left MTG was not as successful. Although Agreement was more strongly related to MTG activity than the other three category variables, this difference was not statistically significant in the present study. Further, based on the role of MTG in resolving semantic interference (de Zubicaray et al., 2001) and the sensitivity of this region to lexical semantic proficiency (Blumenfeld, Booth, & Burman, 2006), I predicted that Agreement would be associated with broader or more effortful lexical semantic search and thus greater MTG activity. This prediction was not supported, as Agreement trended towards positive correlation in the specific region previously associated with inter-lemma semantic competition.

Several explanations could be proposed for the lack of significant findings in MTG and the apparent reversal of the Agreement effect. Competition between names for the same semantic concept may not elicit the same neural responses as competition between related but non-interchangeable concepts. Because the beta value for Agreement did not significantly differ from zero in the present study, its direction could change with additional measurements. The representative sphere used to test the ROI hypothesis for MTG may also have affected the observed local response to Agreement. Each of these cases may be tested by increasing the number of participants in the study to allow for a whole-brain analysis and more statistical power for differentiating the predictors. If, after further inquiry, MTG remains a weak or non-correlate of the category variables, Rodriguez-Fornell et al.’s (2009) account of MTG as a semantic integrative mechanism may better account for these results, with MTG activity being less sensitive to within-category variation and more responsive to across-category semantic contrasts.

In the second hypothesis, I predicted that Typicality would prove to be a better correlate of vOT activity than Familiarity. This finding would be a significant revision of Wilson et al.’s
(2009) observation that their measure of concept familiarity was significantly and negatively related to activity in left fusiform gyrus (among other regions). Chapter 3 of the present dissertation demonstrated that Typicality and Familiarity are unique measures of category norms, differing in their degree of language dependence, and possibly differing in their sensitivity to within-category variation. However, in the present experiment, Typicality and Familiarity did not significantly differ in their relationship to activity in the same region of left fusiform gyrus measured by Wilson and colleagues. While the lack of statistical difference may, again, be related to insufficient power, the overall direction of the relationships actually supports Wilson et al.'s finding that Familiarity is negatively related to vOT activity, while Typicality appeared to have a positive relationship with vOT activity. Consistency with previous research lends to the reliability of the Familiarity measurement, as the single exemplar approach used by Wilson et al. and the multiple exemplar approach used in my norming study (Chapter 3) yielded similar functional brain correlates.

What remains unclear is whether both Typicality and Familiarity may have been affected by the presence of the exemplar during the ratings elicited in Chapter 3. Exemplar typicality may now be ruled out as the underlying effect measured in concept familiarity, given the norms of Chapter 3 and the evident dissociation of these two variables in the present neuroimaging results. The construct validity of the Familiarity measure, however, remains to be tested. A second measure of Familiarity using words (category names) without example images would provide some validation if the word- and picture-based measures were highly similar.

**Neurocognitive Models and Categorization**

Identifying common predictions in four neurocognitive theories of word production, I hypothesized a four-way dissociation relationship to relate the four category variables to
established functional brain correlates of cognitive processes in picture naming. While very few neuroimaging studies have addressed the category variables used in the present study, some reasonable inferences could be drawn from the prior literature, outlined in Table 4-3 (Rationale & Hypotheses). Ultimately, none of these predictions were fully realized (see Table 4-7, Results), but a few important steps were made towards the integration of these variables relevant to lexical categorization with existing theory.

The ROI analyses discussed above yielded two important points of convergence between the present study and word production theories. In Chapter 3, Names and Familiarity were normed specifically as measures of lexical categorization by eliciting on exemplar-level variation within each category. The only significant result among the four-way dissociation was that Names was a better predictor than Agreement in IFG, but in the direction opposite to the hypothesis. Familiarity did not differ significantly from Typicality in any region, but they were qualitatively distinguished by the respective directions of their relationships in vOT and OT. The relationship of these Names and Familiarity variables to the a priori ROIs were, in fact, consistent with the previous literature and overarching roles of vOT in associating objects with semantic concepts and IFG in lexical selection, respectively. Although specific functional correlates of Typicality and Agreement have not yet been identified, each can now be posed as a unique predictor with specific relevance to lexical categorization as quantifiers of individual objects' relationships to dominant names.

While the ROI analyses generally support integration of category variables with word production models, the restriction of this neuroimaging study's analyses to the a priori ROIs presents some weaknesses. It cannot be assumed that the selected ROIs are the most descriptive regions for activity related categorization or the best correlates of the category variables. In the worst-case scenario, an entirely different set of ROIs could be better correlates of the category variables and thus present results entirely incompatible with word production models. This
explanation is improbable given that the ROIs were selected from studies with similar variables of interest, but it remains untested.

Many other functional anatomical regions were identified as possible correlates in the whole-brain analysis. The ROI analysis does not address their potential involvement in lexical categorization. Of particular interest, the Chinese data suggested involvement of the anterior cingulate cortex as a correlate of Names. This region was not included in the original hypotheses, but ACC is noted in at least two models of word production, monolingual (Price, 2012) and bilingual (Abutalebi & Green, 2007). In both models, ACC is proposed to be a control mechanism for detecting conflict or suppressing competing words. This role would be consistent with the present observation that ACC is positively related to the Names variable, and in the next Chapter ACC provides a test for cross-language interaction of lexical categories in bilingualism.
Chapter 5

Correlates of Competition in Bilingual Lexical Categorization

Rationale & Hypotheses

Much of the recent behavioral research in lexical categorization has focused on bilingualism and the convergence of bilinguals' lexical categories in response to conflict arising from cross-language category differences. Chapter 2 of this dissertation is one such example, tracking the shift in lexical categories in Chinese (L1) and English (L2) through a cross-sectional sample of Chinese-English bilinguals immersed in each language. Cross-language influence, according to this study, is bi-directional between L1 and L2 for bilinguals with L2 immersion experience. Changes in both languages (relative to native, monolingual norms) occur as a result of interaction with the other language. How, then, do the underlying brain processes supporting lexical categorization in native Chinese speakers change as a result of Chinese-English bilingualism? The present neuroimaging study explores functional correlates of Chinese-English bilinguals’ lexical categorization, drawing parallels with the behavioral effects measured in Chapter 2, and building on the monolingual baseline described in the previous neuroimaging study of monolingual categorization (Chapter 4).

As in Chapter 4, the present study begins with a few hypotheses drawn from the word production literature. Studies of bilingualism put a special emphasis on shared semantic storage and competition between words with very similar meanings—translation equivalents (see Chapter 1, section on Bilingualism for review). Recent behavioral findings (such as those in Chapter 2) provide convergent evidence of this cross-language lexical semantic interaction, particularly the high degree of category convergence observed in both simultaneous and sequential bilinguals.
Further, the behavioral evidence presented in this dissertation emphasizes that multiple names for an object may compete in lexical categorization. Behaviorally, this effect is quantified as a significant negative effect of the number of alternate names for an object in the non-target language on category native-likeness.

A neural signature of this competition between multiple names is identified in monolingual categorization as an increase of activity in pars triangularis (see Kan & Thompson-Schill, 2004 and Chapter 4). Monolingual neurocognitive models of word production largely concur on this selective function for inferior frontal gryus (see review in Chapter 4), as well as Abutalebi and Green's (2007) neurocognitive model of bilingualism. One of the few neuroimaging studies to simultaneously compare monolinguals, bilinguals, and multilinguals in the same language found that activity in PTri scales up with the number of languages spoken (Parker-Jones et al., 2012), suggesting that knowing the name for an object in each language increases the selective demands proportionally. In sum, monolingual and bilingual studies both provide evidence that multiple, competing names for an object (whether in one or two languages) result in increased selection demands and thus increased activity in PTri.

One monolingual model (Price, 2012) has also proposed a specific role for anterior cingulate cortex (ACC) in suppressing competing words in production, drawing an interesting parallel with Abutalebi & Green's (2007) model of bilingualism which makes a slightly modified prediction for ACC’s involvement in conflict detection. The whole-brain analysis in Chapter 4 presented some tentative evidence that this region may participate in the resolution of competition between Chinese names in lexical categorization, and at least one study of monolingual semantic interference (de Zubicaray et al., 2001) has found the same for across-category competitors.

Parker-Jones et al.'s (2012) comparison of mono-, bi-, and multilingual participants did not find evidence for increased ACC activity in bilinguals and multilinguals, which they attributed to the monolingual context of the experiment failing to require language control.
However, the behavioral evidence in bilingual categorization suggests that even in highly monolingual environments, cross-language influence may be observed in lexical categories. One possible explanation for this discrepancy is that cross-language influence in lexical categorization is an offline, representational effect. L2-to-L1 transfer may affect the underlying lexical semantic mappings in both languages but result in minimal online competition between names. On the other hand, if ACC activity in bilinguals scales with both L1 and L2 Names, then both languages could be construed as active in word production processes.

This neuroimaging study sets out to test online cross-language interaction in bilingual lexical categorization. In this task, identical to the monolingual neuroimaging study in Chapter 4, Chinese-English bilinguals living in an English immersion environment name photographs of common objects in their L1 (Chinese). The same Chinese and English category variables normed in Chapter 3 and used as to describe monolingual categorization in Chapter 4 are now entered as predictors of bilinguals’ functional brain activity. SPM’s step-wise orthogonalization procedure accounts for any variance due to control variables (Visual Complexity, Chinese Frequency, English Frequency) and Chinese category variables before estimating the effects of the English Agreement and Names, insuring that resulting functional correlates of English variables cannot be explained by cross-language similarities.

For this study, I propose a basic cross-language competition hypothesis about the functional correlates of L1 and L2 category variables in these two regions implicated as lexical selection or control mechanisms for word production. Table 5-1 outlines this prediction in detail, but it may be stated simply as the number of alternate names in both Chinese and English (after accounting for Chinese) are hypothesized to be significant predictors of increased activity in both PTr and anterior cingulate cortex. In contrast to Parker-Jones et al.’s (2012) study which found that the number of languages was a predictor of PTr activity, I assert that within-language variation in the number of names in both L1 and L2 will be significantly related to variation in
activity of these regions, confirming that multiple competing names in both languages are active in this strictly L1 task.

Table 5-1. Hypothesized functional correlates of Names variable in Chinese and English.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pars triangularis</td>
<td>Lexical selection (monolingual)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interference control (monolingual &amp; bilingual)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anterior cingulate cortex</td>
<td>Competitor suppression (monolingual)</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Conflict-monitoring (bilingual)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language control (bilingual)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method

Participants

Chinese-English bilinguals were recruited at Penn State University through campus and community flyers and by word-of-mouth. Seven (4 F, 3 M) participants in all completed the two-day study. Participants were mainly graduate students and advanced undergraduate students (mean age 25.7 years) with two or more years of English immersion (mean 3.4 years) in the United States. All participants were right handed, assessed by Snyder & Harris’ (1993) Handedness Questionnaire, and scored 3.5 or higher out of 5 (mean 4.4).

The abbreviated Language History Questionnaire (Li et al., 2013) was used to document experience with other languages, including English. All participants listed English as their second language with moderate to high proficiency (mean 4.7 / 7, minimum score 4). Two participants also reported experience learning French but rated their proficiency rather low (2.5 and 2), and neither had lived in a French immersion environment.
Materials & Procedure

All materials and procedures were identical to those used with the participants in the preceding monolingual neuroimaging study. All written instructions were printed in Chinese simplified script and verbal instructions during the fMRI session were provided in Chinese by a native speaker.

Image Acquisition & Processing

MRI sessions were conducted on the same Siemens 3T scanner in the United States used for the English monolingual participants in the preceding study. Structural T1-weighted images were composed of 160 sagittal slices, 1 mm thick with 50% over-sampling (field of view: 256 mm, TR: 1650 ms, TE: 2.01 ms, Flip angle: 9°). Following the structural scan, each participant completed three counterbalanced runs of picture-naming while undergoing functional T2*-weighted EPI. Functional images were composed of 34 transverse slices interleaved (anterior to posterior), 3.8 mm thick (field of view: 240 mm, TR: 4000 ms, TE: 25 ms, Flip angle: 70°). Each functional measurement was completed in 2000 ms with an additional 2000 ms delay before the onset of the next measurement. This delay allowed naming trials to be synchronized with the scanner so that the first 2000 ms of each naming trial would be relatively noise-free for recording voice responses.

All imaging data were processed using SPM 8 (Friston et al., 1995). Functional images were referenced to the middle slice in each scan, realigned to a mean image of the functional sequence with six motion parameters, and co-registered with the T1 structural image. Segmented structural images were regularized to the East Asian template and normalized to an MNI
template. Finally, normalized functional images were smoothed with an 8 mm FWHM Gaussian kernel.

**First-level Analysis**

A general linear model was fit to each voxel’s activity to estimate the effect of object picture stimuli and scrambled picture stimuli over baseline in that region. In each of the three functional runs, object and scrambled stimuli were entered as separate predictors. Six motion parameters were also entered as nuisance regressors to reduce the effects of motion that were not removed by motion correction. Separate regressors were estimated for each functional run and combined in the first-level effects estimates.

Visual Complexity, Chinese frequency, four Chinese category variables, English frequency, and finally English Agreement and English Names were entered as parametric modulators of the object stimuli, yielding an estimated beta value for each variable representing the change in signal as compared to the average (un-modulated) response for picture stimuli. Parametric modulators in SPM are entered by a step-wise orthogonalization procedure wherein the first modulator is estimated, and if the second modulator is correlated with the first, only the orthogonal component of the second modulator is estimated. This procedure reduces the statistical power for later-entered variables but is intended to improve the stability of estimates for the earlier-entered variables. All category variables were standardized by subtracting the mean value and dividing by standard deviation.
Results

Anatomically-Constrained Group Results

The first group analysis was performed over an anatomically specified region of the brain to reduce the number of voxels searched and thus lower the necessary correction for multiple comparisons. Based on the results of the monolingual neuroimaging study, a whole brain analysis was not expected to produce significant results due to the relatively low sample size. Instead, the left anterior cingulate cortex and the left pars triangularis were specified from the Automated Anatomic Labeling (AAL; Tzourio-Mazoyer et al., 2002) definitions for anatomic regions of interest to perform the search. Although the first level model included all Chinese category variables and English Agreement to control for language-specific effects of Chinese and inter-variable correlation, only the effects of Chinese and English Names variables are necessary to test the main hypotheses. Length of residence in an English immersion environment proved to be a very important predictor in the behavioral results and was thus entered as a covariate for each participant.

Table 5-2. Results of anatomically-constrained analysis for Chinese and English Names.

<table>
<thead>
<tr>
<th>Anatomical Region</th>
<th>MNI Coord.</th>
<th>Peak Voxel Statistics</th>
<th>Cluster size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x  y  z</td>
<td>T  p (uncorr)  p (FWE) (voxels)</td>
<td></td>
</tr>
<tr>
<td>Positive correlates of Chinese Names</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pars triangularis (IFG)</td>
<td>-32 30 6</td>
<td>7.58 &lt;0.001 0.404 1</td>
<td></td>
</tr>
<tr>
<td>Anterior cingulate cortex</td>
<td>-4 24 16</td>
<td>7.17 &lt;0.001 0.444 1</td>
<td></td>
</tr>
<tr>
<td>Anterior cingulate cortex</td>
<td>-2 26 14</td>
<td>7.13 &lt;0.001 0.447 1</td>
<td></td>
</tr>
<tr>
<td>Anterior cingulate cortex</td>
<td>0 34 6</td>
<td>7.11 &lt;0.001 0.450 1</td>
<td></td>
</tr>
<tr>
<td>Positive correlates of English Names</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anterior cingulate cortex</td>
<td>-6 20 24</td>
<td>20.49 &lt;0.001 0.010 18</td>
<td></td>
</tr>
</tbody>
</table>
Second-level results were initially thresholded at $p<0.001$ (uncorrected) to search for significant voxels. See Table 5-2 for detailed results. One cluster was significantly and positively related to English Names, even after correction for multiple comparisons ($p$(FWE)=0.010). The peak voxel was observed in the left anterior cingulate cortex (see Figure 5-1 for illustration). Four individual voxels were significantly and positively correlated with Chinese Names, but none represented a larger cluster (cluster thresholds of 5 or more voxels are customary) and none survived family-wise error correction. No voxels were significantly, negatively related to Chinese or English Names at $p<0.001$ (uncorrected).

Figure 5-1. SPM thresholded T-map of functional response to English Names variable. Cluster of 18 voxels located in left anterior cingulate cortex.

**Functionally-Defined ROI Analyses**

A more restricted ROI analysis was performed to test whether average responses in a set of ROIs were significantly greater than zero. Three spherical ROIs were selected for this analysis to represent PTri and ACC (see Table 5-3). The significant cluster of activity identified in the
prior group analysis served as the template for ACC, with the ROI sphere centered on the peak voxel and extending 5 mm in radius. The first ROI for PTri was drawn from Kan & Thompson-Schill's (2004) study, as in Chapter 4. This ROI showed significant differences between Agreement and Names in monolingual lexical categorization, and should thus be a useful region for measuring the functional response to L1 names. Another sphere in PTri was also defined based on the most significant voxel correlated with Chinese Names in the present study. This location is far enough from the other PTri ROI that the two spheres do not intersect.

Table 5-3. Locations of spheres in anatomical and MNI space representing each ROI.

<table>
<thead>
<tr>
<th>Anatomical Structure</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>radius</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>L Anterior Cingulate Cortex</td>
<td>-6</td>
<td>20</td>
<td>24</td>
<td>5 mm</td>
<td>Present study</td>
</tr>
<tr>
<td>L Pars triangularis</td>
<td>-36</td>
<td>24</td>
<td>21</td>
<td>5 mm</td>
<td>Kan &amp; Thompson-Schill, 2004</td>
</tr>
<tr>
<td>L Pars triangularis</td>
<td>-32</td>
<td>30</td>
<td>6</td>
<td>5 mm</td>
<td>Present study</td>
</tr>
</tbody>
</table>

Individual participants' standardized betas for Chinese and English Names were extracted and averaged within each ROI. Figure 5-2 shows the mean beta values for each predictor in ACC and PTri. English Names was a highly significant predictor of functional activity in ACC ($t(6)=11.85$, $p<0.001$), consistent with the previous group analysis. Chinese Names, however, were not a significant predictor of activity in this region, despite being entered before English Names ($t(6)=0.46$, $p=0.34$). Chinese Names was a significant predictor only in the PTri ROI defined from the voxel in this study ($t(6)=6.76$, $p<0.001$). Using the a priori ROI from Chapter 4, English Names was a significant predictor ($t(6)=2.20$, $p=0.040$) but was only marginal in the PTri ROI defined by the present study ($t(6)=1.96$, $p=0.053$).
Comparing the two spherical ROIs in pars triangularis highlights a functional dissociation between these regions and their relative responses to each language. The \textit{a priori} region defined by Kan & Thomspson-Schill's study did yield a significant relationship with English Names but no relationship with Chinese Names was found. The sphere based on an active voxel from the present study reproduced the significant effect of Chinese names suggested by the anatomically-constrained group analysis, and the effect of English in this region was apparent though marginally significant. Table 5-4 describes these differences in parallel to the format of the hypotheses presented in Rationale & Hypotheses.

Table 5-4. Mean standardized beta values for Chinese and English Names in each ROI. Compare with Table 5.1 (Rational & Hypotheses). Significance \( \sim p<0.10, * p<0.05, ** p<0.01 \) for \( \beta \neq 0 \).

<table>
<thead>
<tr>
<th>ROI</th>
<th>Source for location of centroid</th>
<th>Mean Standardized ( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pars triangularis</strong></td>
<td>Kan &amp; Thompson-Schill, 2004</td>
<td>(-0.21) *0.71</td>
</tr>
<tr>
<td>Present study</td>
<td></td>
<td><strong>0.36</strong> (~-0.47)</td>
</tr>
<tr>
<td><strong>Anterior cingulate cortex</strong></td>
<td>Present study</td>
<td>0.18 <strong>1.01</strong></td>
</tr>
</tbody>
</table>

Taking all three ROIs as representative of their broader anatomical structures, the positive relationship between Chinese Names and English Names in PTri was supported by
significant standardized betas, and English Names were a significant predictor in ACC. Chinese Names were not a significant correlate of activity in this specific ROI of ACC, although the group analysis did reveal some significant voxels that were not explored in this ROI analysis.

**Monolingual and Bilingual Responses in Control Regions**

Comparison of the ROI analysis results to other findings could clarify the roles of L1 and L2 in bilingual lexical categorization. The ROI analysis was therefore extended to include Chinese and English Agreement and the neuroimaging data for Chinese monolinguals collected in Chapter 4. Although Agreement was a significantly weaker predictor of functional activity in PTri for monolinguals, its successful dissociation from Names in both behavioral (Chapter 2) and neuroimaging (Chapter 4) measures makes it a useful contrast for interpreting the effect of Names in a given region. Mean standardized betas for each variable (Agreement and Names) were estimated for both languages in Chinese-English bilinguals and for Chinese monolinguals. Chinese monolinguals' functional correlates for one ROI in PTri was presented in Chapter 4, and two additional regions (PTri and ACC, defined by active voxels in the present study) are estimated for this comparison.

Figure 5-3 illustrates the respective means for standardized betas for Agreement and Names in Chinese monolinguals and bilinguals in each ROI. Significant differences for each mean compared to zero were tested for a preliminary comparison (see Figure 5-3). Next, a 2 x 2 x 3 (Group x Variable x ROI) repeated measures ANOVA was performed to identify significant differences in how Chinese monolingual and bilingual speakers responded to L1 norms in each ROI. Because Agreement and Names are strongly negatively correlated, the sign of one variable (Agreement) was reversed to allow comparison of their relative magnitudes, as was done in Chapter 4. Significant interactions were found for Variable x Group ($F(1,13)=7.37$, $p=0.018$) and
Variable x ROI x Group ($F(2,26)=5.19, p=0.013$). Overall, Chinese monolinguals showed a greater response to L1 Names than L1 Agreement while Chinese bilinguals had the opposite tendency. This reversal (captured by the Variable x Group interaction) was reflected in ACC ($F(1,13)=5.84, p=0.031$) and the *a priori* ROI for PTri ($F(1,13)=9.12, p=0.010$), but not in the ROI defined in the present study ($F(1,13)=0.110, p=0.745$).

![Figure 5-3](image-url)

**Connectivity Models**

Extended unified SEM connectivity models were computed for each group using the GIMME group model search procedure (ver. 8; Gates & Molenaar, 2012) to infer the relationships between ROIs in monolingual and bilingual speakers of Chinese. All five ROIs were included in these connectivity models to give the best estimate of directed and inter-dependent connections for comparison with neurocognitive models discussed in Chapter 4. The euSEM estimates both time-lagged and contemporaneous relationships between regions of interest, enabling inferences about directed connections even for contemporaneous activity (Gates,
Molenaar, Hillary, & Slobounov, 2011). Lagged connections, however, are limited by the time resolution of measurements, which are 4 seconds apart in the present study to accommodate picture naming responses. Thus lagged connections, although important for the estimation of the model, are difficult to interpret in the context of relatively fast cognitive processes (picture naming). Many lagged connections occur along the same paths as a contemporaneous connection, but with the opposite sign. These connections may indicate oscillatory activity which would constitute a single connection, varying over time, rather than two connections. When this condition occurs, lagged connections are superimposed over the contemporaneous connection to this behavior along the path. Table 5-5 provides a full list of connection strengths estimated for the Chinese monolingual and bilingual models.

Table 5-5. Connectivity weights for monolingual and bilingual Chinese connectivity models, estimated by the euSEM using the GIMME group-wise search procedure. ROIs in columns direct influence on ROIs in rows. Connection strengths are reported (X/Y) where X is the time-lagged parameter and Y is the contemporaneous parameter.

<table>
<thead>
<tr>
<th>Chinese Monolinguals (n=8)</th>
<th>vOT</th>
<th>OL</th>
<th>MTG</th>
<th>ACC</th>
<th>PTri</th>
<th>Pic</th>
</tr>
</thead>
<tbody>
<tr>
<td>vOT</td>
<td>0.39 / -</td>
<td>- / -0.12</td>
<td>- / 0.26</td>
<td>- / -0.24</td>
<td>-0.08 / -</td>
<td>- / 0.17</td>
</tr>
<tr>
<td>OL</td>
<td>-0.31 / 0.58</td>
<td>0.65 / -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MTG</td>
<td>-</td>
<td>0.57 / -</td>
<td>0.10 / -</td>
<td>0.29 / -</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ACC</td>
<td>-</td>
<td>-</td>
<td>0.36 / -</td>
<td>0.60 / -</td>
<td>-</td>
<td>- / 0.20</td>
</tr>
<tr>
<td>PTri</td>
<td>- / 0.59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.38 / -</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chinese-English Bilinguals (n=7)</th>
<th>vOT</th>
<th>OL</th>
<th>MTG</th>
<th>ACC</th>
<th>PTri</th>
<th>Pic</th>
</tr>
</thead>
<tbody>
<tr>
<td>vOT</td>
<td>0.43 / -</td>
<td>- / -0.01</td>
<td>-</td>
<td>- / -0.29</td>
<td>-</td>
<td>-0.41 / 0.63</td>
</tr>
<tr>
<td>OL</td>
<td>- / 0.48</td>
<td>0.48 / -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MTG</td>
<td>-</td>
<td>-0.28 / 0.33</td>
<td>0.60 / -</td>
<td>-</td>
<td>0.16 / -</td>
<td>- / -0.19</td>
</tr>
<tr>
<td>ACC</td>
<td>-0.29 / 0.57</td>
<td>-</td>
<td>-</td>
<td>0.49 / -</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PTri</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.49 / -</td>
<td>0.29 / -</td>
<td>- / 0.06</td>
</tr>
</tbody>
</table>

Connectivity models for Chinese monolingual and bilingual participants are illustrated in Figure 5-4. Overall, the models are structurally very similar. In both models, ventral occipito-temporal cortex responds to the onset of stimuli (Pic -> vOT connection) and maintains a
recurrent processing connection with the occipital lobe. In more anterior regions, the anterior cingulate cortex has a strong influence over activity in PTri and extends a connection back upstream to vOT. Four important differences also appear between the monolingual and bilingual models. (1) The effect of the picture stimulus on vOT greatly strengthens in the bilingual model. (2) At the same time, a new negative or inhibitory effect of the picture stimulus on MTG appears in the bilinguals, and (3) the recurrent processing connection between ACC and MTG in the monolingual model does not appear in the bilingual model. Finally, (4) the valence of ACC’s connection to vOT changes from positive (in monolinguals) to negative (in bilinguals).

Figure 5-4. Connectivity graphs estimated for Chinese Bilinguals and Chinese Monolinguals. All networks contain only those connections which were included in 75% or more of participants' individual models. “Pic” represents appearance of an object stimulus. All ROIs are represented by spheres described in Table 4-6. See Table 5-3 for ACC details. Note: All ROIs significantly correlated with themselves in the time-lagged comparison. These auto-regressive terms are not reported.
Discussion

Cross-Language Interaction in Categorization

This neuroimaging study of bilingual lexical categorization tested the basic hypothesis that anatomical regions previously associated with language control would show significant sensitivity to variation in L2 category variables during an L1 naming task. This hypothesis was strongly supported as both the number of alternate names for an object in English (the English Names variable) and dominant name agreement (English Agreement) significant predictors of functional activity in the specified regions of interest. English Names and Agreement had highly significant correlations with left anterior cingulate cortex. L1 category information was also strongly associated with activity in ACC, as L1 dominant name agreement (Chinese Agreement) was a significant negative correlate. While no single ROI in PTri significantly correlated both L1 and L2 variables, two regions of this anatomical structure separately varied with the number of alternate names in each language.

One way of framing this result might be to argue that L2 category norms gradually reshape L1 mappings to such an extent that within-category variation in L2 predicts the object-specific changes in L1 representation. For example, if a given drinking vessel is extremely difficult to categorize in L2 (two or more alternate names, low dominant name agreement), then L1 lexical semantic mappings for this object may be altered resulting in greater competition without assuming direct activation of any specific L2 names. The broader bilingual lexicon literature, however, generally concurs that words in both languages are active in production (see Kroll, Bobb, & Wodnieka, 2006), suggesting that this alternate explanation is inadequate.

Instead, the present findings provide strong evidence for online participation of both L1 and L2 lexical semantic information (as measured by native speaker norms) during lexical
categorization. Activity in both ACC and PTri significantly increased with the number of alternate English names, as opposed to the global increase for bilinguals over monolinguals reported in previous comparisons such as Parker-Jones et al.'s (2012) study. Further, activity related to L2 category variables represented a significant increase over activity explained by L1 and control variables, which were entered first in the orthogonalization procedure.

**ROIs & Language Competition**

According to recent neurocognitive models of word production, the anterior cingulate cortex is associated with language control in bilinguals (Abutalebi & Green, 2007) and with suppression of competing words in monolinguals (Price, 2012). The present study affirms both of these assertions in that ACC was associated with L2 during an L1 production task (language control), and its functional response specifically varied with the number of L2 words available for naming a given object (competitor suppression).

PTri has also been associated with these control mechanisms both within-language (lexical selection) and between-language (interference control). One ROI, correlating with English Names, is drawn from prior research on dominant name agreement effects in English monolingual speakers (Kan & Thompson-Schill, 2004). This region probably offers the best objective test of language effects in PTri given its *a priori* definition and known relationship to another categorization effect.

The alternate PTri ROI, defined by a peak voxel but not a significant cluster in the group analysis, significantly correlated with Chinese Names. There is some risk that this finding reflects a Type 1 error, given that testing the region's response to a variable that was used to define the region is logically circular. On the one hand, the ROI sphere is considerably larger (approximately 500 mm³) than the single voxel used to define it (~15 mm³) and thus reflects a
real trend across several voxels rather than a single outlier value. Smoothing procedures in the first-level analysis, however, use an 8 mm FWHM kernel, which covers a significant portion of the 10 mm diameter sphere, potentially “smearing” a single Type 1 error throughout the ROI.

From these data, it would be difficult to determine whether the dissociation of L1 and L2 sensitivity in PTri is a result of finer-grained differences in the spatially localized response or whether it is, indeed, a false positive. Nonetheless, the role of L2 in these theoretically established control regions is apparent and statistically sound. The connectivity model suggests that ACC may operate up-stream from PTri, as the directional and contemporaneous connection from ACC to PTri is strong in both groups. This additional information might suggest that ACC operates as the primary top-down control mechanism to which IFG selectively responds. Further data collection might help to clarify the localization of L1 and L2-related responses in PTri and clarify its role in control and/or inhibition.

**Monolingual and Bilingual Comparison**

One way to test the validity of the observed correlation in the peak voxel-defined PTri ROI using only the data on-hand would be testing the effect in another, comparable sample. Comparing the results in this ROI to results obtained from Chinese monolinguals in the same sample helped to distinguish between the two measures of PTri. The repeated measures ANOVA in the monolingual bilingual comparison revealed a significant interaction between groups in the magnitude of their respective sensitivities to Agreement and Names variables. This interaction was significant in ACC and in the *a priori* PTri ROI, but it was not present in the other PTri ROI. In fact, responses to L1 variables in monolinguals and both languages in bilinguals were small in magnitude and qualitatively very similar, and only the original observed relationship with Chinese Names significantly differed from zero. By contrast, significant monolingual correlates
were found in both of the other ROIs, which were not defined based on monolingual results. This pattern is suggestive that the ROI in question may not actually be a strong correlate of any variable.

The comparison of monolingual and bilingual responses presented another surprising finding. As suggested in the ROI discussion, a statistical comparison of the monolingual and bilingual Chinese participants' responses to Chinese category variables revealed a significant reversal in sensitivity to each variable, and this reversal was found in both PTri and ACC. Monolinguals showed a stronger response to Names than to Agreement in both regions. Bilinguals showed a stronger response to Agreement than Names in the same regions, for the same task and stimuli. This observation was not expected, and there is no precedent in the lexical categorization literature for drawing this contrast (much less predicting its direction). This preferential sensitivity effect coincides with differences in the monolingual and bilingual connectivity models as well. MTG involvement generally decreases for bilinguals, as MTG activity declines with stimulus presentation, the ACC/MTG recurrent processing loop disappears, and a new recurrent loop appears between vOT and ACC, bypassing MTG. If these changes describe a change in the processing stream for word production, then the relative influence of each category variable could also change as a result. This unexpected results may provide a link to the behavioral results for L1 categorization by Chinese-English monolinguals in Chapter 2, a relationship which will be explored at greater length in the General Discussion.
Chapter 6

General Discussion

Summary

In this dissertation, I describe four studies of lexical categorization focused on understanding the characteristics of languages (native speaker norms) and speakers (language history) that relate to individual and cross-language variation in lexical semantics. These language norm and language history variables serve to describe lexical category variation among native, monolingual speakers and help to reveal the processes underlying changes in bilinguals L1 and L2 lexical categorization patterns. To this end, a few major findings emerge throughout the described studies: (1) Lexical categories of native speakers are marked by significant interpersonal variation not typically measured by word production or picture naming studies. (2) Bilingual lexical categorization in both L1 and L2 reflects the combined influence of both language norms and language history, given L2 immersion experience. (3) Language-dependent lexical categorization variables significantly affect brain processes underlying picture naming tasks, as evidenced by localized functional brain activity. (4) Brain processes supporting lexical categorization in monolinguals and bilinguals converge with current neurocognitive theories of word production, illustrating the relevance of lexical category variables to understanding lexical semantic processing in general.
Monolingual Lexical Category Variation

Numerous descriptive studies have documented language-specific patterns of lexical categorization for individual semantic domains (see Malt & Majid, 2013 for review), including the Chinese and English native speaker norms used for the dishes stimulus set in Chapter 2. Two measurements of these norms were formalized for the dish stimuli in Chapter 2: dominant name agreement and number of alternate names. The significance of these variables in predicting bilingual lexical categorization (Chapter 2) motivated a much broader norming study of lexical categorization patterns in monolingual Chinese and English speakers across six semantic domains. The norming study presented in Chapter 3 measures lexical category variables in multiple domains simultaneously, demonstrating that lexical category variation within and between Chinese and English occurs over a broad array of concrete objects and varies widely within these categories.

The cross-domain norms identified in Chapter 3 are generally consistent with the findings of previous, smaller picture naming datasets that have measured name agreement or alternate names for objects among native speakers with respect to the range of values for both variables (e.g., Snodgrass & Vanderwort, 1980; Malt et al., 1999; Ameel et al., 2008; Pavlenko & Malt, 2011). While at least one study has simultaneously compared categorization in two similar domains (dishes and bottles, Ameel et al., 2005) only this dissertation norms name agreement and alternate names across several highly distinct semantic domains at once, generating a coherent set of category variables for subsequent experiments in lexical categorization.

Results in both the monolingual and bilingual lexical categorization experiments demonstrate individual speakers' sensitivity to these variables. The relevance of native speaker norms to monolingual language learning was suggested by developmental patterns observed in monolingual child learners (Ameel et al., 2008), demonstrating that the complex input from
native speaker norms are learned with increasing refinement over a long developmental period (into late childhood). The neuroimaging study of Chinese and English monolinguals in Chapter 4 showed that these variables are similarly dissociable in functional brain activity and involved in the processes supporting picture naming in native speakers of each language.

Quantifying these category variables is a novel approach in lexical categorization, as previous studies have often relied on comparisons that require several exemplars from each of a few similar categories (e.g., multi-dimensional scaling: Malt et al., 1999 and Pearson correlations for name distributions: Malt et al., 1999; Ameel et al., 2005; Chapter 2 of this dissertation). Correlations observed between name agreement and alternate names in Chapter 3 support the assertion in Chapter 2 that these variables are, though highly correlated, statistically distinguishable and unique predictors of categorization behavior. The convergence between measurements of these variables in Chapter 3 and previous single-domain studies lends validity to both approaches, demonstrating that lexical category variation is a normative state in language, not constrained to any particular sets of peculiar objects.

Besides the gains in ecological validity associated with a broader sample of objects, the stimulus set normed in Chapter 3 also presents a methodological control not previously available in lexical categorization. Blocked picture naming studies have demonstrated that cumulative semantic interference from naming several semantically related pictures in a brief time period results in significant performance deficits (see Howard et al., 2006 for review). To date, no lexical categorization study has accounted for this effect that may be elicited by single-domain stimulus sets. If lexical categorization research is to be extended into the field of cognitive neuroscience, mixed-domain stimulus sets like the one presented in Chapters 3, 4, and 5 will be necessary to minimize confounds of response slowing and neural satiation resulting from semantic interference. Defining an array of lexical category variables to describe the individual category relationships of given a object (rather than overall category trends across multiple
objects) enables experimenters to compose these mixed sets and even to norm and control other picture naming stimuli from existing sets.

Predictors of Bilingual Categorization

Two studies in this dissertation sampled Chinese-English bilinguals to examine cross-language interaction between lexical categories in sequential bilinguals. The behavioral study (Chapter 2) revealed significant effects of individual language history on the native-likeness of bilinguals' L2 lexical categories and strong bi-directional influence of native speaker norms from both languages in L1 and L2 lexical categorization. Although many of these language history and language norm predictors are highly correlated with each other, statistically significant effects for each were identifiable. These variables are explored below.

Native Speaker Norms

Table 6-1. Summary of neuroimaging and behavioral correlates of two category variables in native Chinese speakers' L1 lexical categorization.

<table>
<thead>
<tr>
<th>Native Norm</th>
<th>Predictor</th>
<th>Ch. 2 Biling.</th>
<th>Ch. 4 Chi. Monoling.</th>
<th>Ch. 6 Biling.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 (Chinese)</td>
<td>Agreement</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Names</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L2 (English)</td>
<td>Agreement</td>
<td>+</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Names</td>
<td>–</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Of specific relevance to the bilingual neuroimaging study (Chapter 5), Chapter 2 quantified how input from native speakers of each language resulted in changes to L1 lexical semantic mappings of these bilinguals. Table 6-1 summarizes the influence of L1 and L2 Agreement and Names variables over L1 performance in three samples of native Chinese
speakers. In Chapter 2, these variables are related to item-wise L1 native-likeness in
categorization. High name agreement in L1 and L2 was positively associated with native-like
categorization, corresponding to the facilitation effect for translation equivalents (Hermans,
2004). The surprising aspect of this finding, however, was the incongruent effect of L1 and L2
alternate names.

The specific roles of L1 and L2 lexical category variables in predicting bilinguals' native-likeness in Chapter 2 also shows that language speakers appear to be sensitive to the relative
frequency of many alternate names, with improved performance when the agreement for the
dominant name increases and decreased performance when alternate name competitors increase
in frequency (e.g., two names distributed 60%-40% versus three names distributed 60%-20%-20%). Cross-situational word learning research has experimentally tested exactly this
manipulation and found that learners of an artificial language, indeed, benefit from the relatively
decreased competition of alternate names (Vouloumanos, 2008).

Comparing these behavioral findings to the results of the neuroimaging studies, a
possible link between brain and behavior stands out. Functional activation in left anterior
cingulate cortex (ACC) of Chinese-English bilinguals shows the inverse pattern of response to L1
and L2 category variables. ACC activity significantly decreases as L1 and L2 Agreement
increase, and it significantly increases with the number of L2 names. ACC showed no significant
response, however, to the number of L1 Names for bilinguals, even though this predictor was
entered relatively early in the orthogonalization procedure. By contrast, L2 names was entered
last (of nine variables) and proved a highly significant predictor. ACC activity may well serve as
an index for lexical semantic conflict in future bilingual categorization studies, consistent with its
conflict monitoring role in Abutalebi & Green's (2007) neurocognitive model of bilingual
language control.
One open question is whether L1 variables bear the same relationship to native-likeness in Chinese monolinguals. Monolingual neuroimaging data (Chapter 5) revealed that ACC positively scaled with Chinese Names. If the hypothesis that this region of ACC is a neural index of lexical semantic conflict, then ACC activity in the Chinese monolinguals predicts a reversal in the L1 Names effect on L1 native-likeness. This prediction would be relatively straight-forward test in any set of monolingual lexical categorization data, resulting in a direct contrast with the observed L1 Names effect in bilinguals.

**Language History Variables**

The most surprising finding about language history was that years of English classroom learning (instead of L2 immersion) was negatively related to L2 native-likeness in lexical categorization. This contrast suggests that L2 training outside of an immersion environment may reinforce L1-to-L2 transfer native-likeness in L2 lexical semantic mappings. The more time that L2 learners spend learning lexical semantic mappings in a non-immersive environment, the more entrenched the L1-driven mappings in L2 may become. This observation meshes well with the strong evidence for cross-language interaction of native speaker norms (Chapter 5). Bilingual categorization seems particularly sensitive to fine-grained detail of language input (Chapters 2 and 5), so L2 acquisition in the absence of native-like input is just as effective in reinforcing L1-like mappings as it would be L2-like mappings in an L2 immersion environment. Thus, Chapter 2 adds the unique caveat to language education that L2 learning without immersion may, in fact, decrease native-likeness.

Second language lexical learning has also previously been seen as relatively insensitive to age effects, due to the theoretical position formalized in Ullman’s (2001) Declarative-Procedural (D-P) Model that associates phonology and syntax learning with procedural memory and lexical
semantic knowledge with declarative memory, a faculty that tends to improve throughout development. The present findings suggest that the dichotomy may not be so clear. When lexical semantics are measured as a complex system of mappings for making generalizations about lexical categories (rather than sets of word-object pairs) a weak pattern of age effects is replicated (Chapter 2). Although there were significant advantages for earlier learners over later learners, these advantages were limited in the sense that for every year of earlier acquisition, the same effects could be gained by an additional year of L2 immersion.

Whereas age effects and training effects focus specifically on the conditions under which bilinguals begin acquiring a new language, eventual native-likeness also depends on how that language is used at later stages, such as in an L2 immersion environment. Research in first language lexical attrition has highlighted the role of bilinguals' specific language use patterns in re-shaping L1 (De Leeuw, Schmid, & Mennen, 2009) and offered a cognitive explanation for how L2 structures are eventually encoded into L1 representations (Wolff & Ventura, 2009). In Chapter 2, increased code-switching is associated with greater L2 native-likeness. However, this effect interacts with L2 immersion such that it applies only after a significant period of immersion (illustrated at 4.7 years in Figure 2.2). For learners with significantly less immersion (including no immersion at all) code-switching behavior had no strong effect on L2 native-likeness. Taken with the effects of classroom learning, this advantage for immersed code-switchers emphasizes the importance of prolonged exposure to native-like L2 input for the acquisition of these lexical semantic mappings, and perhaps this finding may mitigate a popular belief that frequent switching between languages significantly impedes native-like acquisition of an L2. However, one important caveat is that the direction of causal relationship between code-switching frequency and L2 native-likeness cannot be decided from the present results.

Weighing the consequences of the modest age of onset advantage on one hand and the L2 classroom learning disadvantage on the other hand, it's worth asking whether earlier L2
classroom instruction is beneficial for lexical semantic native-likeness. As developmental trajectories and cognitive processes enter the focus of lexical categorization research, the details of these language history variables can be more closely examined. Chapter 5, for example, used length of residence in the United States as a covariate in the group and ROI analyses for L1 and L2 category variables. Although LOR was a significant covariate of ROI activity for bilinguals, it was not a significant predictor for functional activity in any particular voxel. Future studies may sample a broader cross-section of language history variables in bilinguals to determine whether the behavioral relationships between language history and native-likeness are also reflected in brain function.

**Lexical Categorization & the Brain**

Neuroimaging studies of lexical categorization provide information about the processes underlying individual judgments of lexical category in picture naming. One potential criticism of applying lexical category norms to the study of individual language learning is that these norms strictly describe inter-personal variation within a community and are not representative of conflict in individuals' lexical semantic mappings. For example, if a certain vehicle is named with moderate name agreement, the 30% of native speakers who call it “car” may be equally as confident about this name as they are for another vehicle named “car” with 100% agreement, unconflicted by their relatively unusual lexical category boundary. If this were the case, one might infer that lexical category norms do not measure individuals' lexical semantic mappings, but rather just the variation between individuals with different mappings.

The functional neuroimaging study of monolingual Chinese and English speakers in Chapter 4 demonstrates that these norms, in fact, represent significant within-speaker effects and not just community-level statistics. That is, functional brain activity of native, monolingual
speakers of Chinese and English was predicted by the variation in language-dependent lexical category norms. The replication of the behavioral dissociation between Agreement and Names with a corresponding dissociation in functional response is particularly important to this argument. The distinction between Agreement and Names (modifying the conclusion of Kan & Thompson-Schill, 2004) and the replication of Wilson et al.'s (2009) concept familiarity effect provide relatively specific neural signatures for lexical semantic processes that support categorization. The contrast for monolinguals between Agreement and Names in pars triangularis of the left IFG (observed in Chapter 4) and ACC (observed in Chapter 5) serves as a benchmark for lexical semantic competition arising from the multiple names associated with a given object.

Activity in PTri is a strong correlate of alternate names in the monolingual neuroimaging data and provides a significant contrast to name agreement. The estimated effects of exemplary typicality and concept familiarity on ventral occipito-temporal regions of the left hemisphere (specifically, left fusiform gyrus) and the occipital lobe (specifically calcarine fissure) were similarly opposing (if not statistically significant). At present, neither typicality nor familiarity have been directly related to lexical category acquisition or representation by monolinguals or bilinguals, but with convergent evidence from Wilson et al. (2009), both a behavioral rating and neural marker for concept familiarity has been established and may be applied in future categorization experiments. Exemplar typicality, for its part, is now strongly evidenced to measure an aspect of lexical semantics not captured by the other three category variable, and it has been significantly distinguished from alternate names in relation to occipital lobe responses (Chapter 4).
Lexical Semantic Theories & Categorization

Representation Models

In Chapter 1, I discussed how the connectionist view of lexical and semantic representation has been extended to lexical categorization through distributed feature models. Various iterations of these representational models set out to describe how fine-grained semantic information can be integrated to compose broader lexical concepts or taxonomic hierarchies and thus activate lexical candidates for production. This dissertation does not directly investigate how lexical semantic mappings are represented in the brain, but the present findings do support the predictions made by these models.

The shared conceptual store in bilingualism (see Potter et al., 1984 and Kroll & Stewart, 1994) is most evident in the high degrees of category convergence exhibited by Chinese-English bilinguals at all stages of immersion. Independent of their relative dependence on L1 or L2 norms in production, the idiosyncratic mappings of these bilinguals correlated between languages above 0.90. This high degree of convergence is in line with previous measurement of simultaneous Dutch-French bilinguals (Ameel et al., 2005) who showed the same degree of convergence for bottles and dishes (0.88 and 0.91, respectively). This convergence appears to be supported by the intense cross-language interaction evidenced in the behavioral and brain measures throughout this dissertation. Such convergence actually exceeds the predictions of Pavlenko's (2009) Modified Hierarchical model, which describes partly-shared and partly-independent conceptual stores for each language. With cross-language correlations in the order of 0.90, the bilingual conceptual store may be better described as wholly shared between languages with language-specific mappings representing the exceptional case. The Distributed Feature Model (De Groot, 1992; Van Hell & De Groot, 1998) and its derivatives (Dong, Gui, & MacWhinney, 2005; Ameel et al.,
2009) are explicitly shared conceptual store models, in that L1 and L2 mappings share access to underlying features. The effect of code-switching observed in Chapter 2 is, in fact, suggestive of such a model as it highlights the parallel activation and associative learning of L1 and L2 mappings during production of either language (Wolff & Ventura, 2009).

Several factors in connectionist training paradigm, such as amount, frequency, and consistency of input are readily translated into lexical categorization and language history variables and thus useful for future tests of lexical category convergence in a computational implementation of distributed feature representations. Amount of language experience, frequency of the dominant name relative to other names (naming agreement), and alternate names have all been quantified in the present study, and they are naturally translated into associative training regimen between objects and an array of names defined by native speaker norms. The emergence of meaningful categorical relationships from such training paradigms is a particular strength of connectionist models (e.g., McClelland & Rogers, 2003), ripe for application to this lexical categorization.

No computational implementation of lexical category learning has been published to date, although at least two recent attempts have been presented (Zinszer et al., 2011; Fang et al., in prep). Future research may test whether or not training parameters that simulate native-like lexical category input can produce analogous results in computational models, validating the comparison between lexical categorization in language and lexical semantic learning in connectionist models.

**Production Models**

Finally, one of the ultimate goals of this dissertation has been to demonstrate that the variation in picture naming responses studied in lexical categorization is, in fact, normative to
object naming in general and thus can be situated in broader theories of word production. One short-coming in the current theories is a failure to account for this within-category (that is, between objects of the same name) variability for lexical semantic mappings even for monolingual speakers and the differences in these mappings that arise in bilinguals as a product of language interaction. Because Levelt et al.'s (1999) LRM model takes lexical semantic association for granted, the studies of lexical categorization presented here have the potential to significantly elaborate upon these processes and connect with the existing functional brain regions associated with these stages of production.

At the ground floor of the LRM model, a stimulus (an object, a mental representation, printed word, etc.) is conceptualized into a discrete package for association with retrieved lexemes. The neurocognitive models of word production outlined in Chapters 1 and 4 generally attribute this associative process to left temporal lobe regions. Based on previous studies of lexical semantic interference and proficiency, I predicted that variation in the difficulty of conceptualizing an object may be represented by dominant name agreement, which would be correlated with activity in middle and posterior portions of left middle temporal gyrus. This prediction was not supported by the current results, which is actually consistent with the failure of at least one previous study of bilingualism to find increases in MTG activity for bilinguals (Parker-Jones et al., 2012). While direct translations prime (rather than compete with) one another (Hermans, 2004), an extended search of the lexicon for semantic associations might have reasonably elicited increases in MTG (Blumenfeld, Booth, & Burman, 2006).

In retrospect, exemplar typicality may be a better operational measure for ease of conceptualization, which would move the search for functional correlates of conceptualization to anatomical regions associated with typicality in Chapter 4, such as the left occipital lobe (measured at calcarine fissure) and ventral occipito-temporal lobe (measured at fusiform gyrus). In these regions, my predictions about the effect of typicality were directionally backwards, but in
future studies, functional responses to typicality may serve as a better analog for conceptualization processes.

The overall configuration of the Chinese monolingual connectivity model highlighted vOT as a central integrative region, insofar as it received both bottom-up input (the picture stimulus and a visual processing region of the occipital lobe) and top-down influence from control regions. Activity appeared to cluster in these respective sub-networks with recurrent loops between vOT and OL and between MTG and ACC. The bilingual model did not reflect this segregation of function. Stimulus-related activity shifted directly and indirectly (via vOT) to ACC, bypassed connections observed in monolinguals from MTG to other ROIs, and directly inhibited MTG activity in response to stimulus presentation. Bilingual processing was, in this way, largely a bottom-up constructive process. The recurrent language control connections between ACC and MTG were replaced by a new recurrent connection between ACC and vOT. This shift may explain the puzzling reversal of preference between monolinguals and bilinguals for Names and Agreement, respectively, if bilinguals rely more on the graded suitability of an object to the dominant name than on selection between competitors. In essence, bilinguals may weaken top-down linguistic expectations driven by the ACC ↔ MTG network and instead favor intuitions about similarity (or typicality) informed by more posterior processes in OL and vOT.

While this explanation is highly speculative, it fits well with some other empirical findings. Individual participants’ idiosyncratic object-similarity judgments are strongly predicted by the distribution of activity in ventral temporal lobe regions (Raizada & Connolly, 2012). Given that the connectivity models in Chapter 5 identified a stronger relationship between vOT and language control regions, one might infer that the importance of these similarity judgments has been elevated in the name selection process. Preference for bottom-up similarity has, in fact, already been documented for simultaneous bilinguals who shift their converged lexical categories towards language-independent similarity to minimize conflict (Ameel et al., 2009). Relatedly,
recent neuroimaging research in Chinese lexical tone processing by Chinese-English bilinguals has specifically highlighted the decrease of MTG specialization for categorical perception of phonemic tones (Zinszer, Chen, Wu, Shu, & Li, 2014). Whether bilingualism analogously predicts de-specialization of MTG responses for categorical perception of lexical semantic concepts is presently unknown.

Finally, the involvement of frontal lobe structures in lexical selection was clearly evidenced in the present lexical categorization studies. Most of the neurocognitive models cited in Chapters 1 and 4 attribute lexical selection or one of its sub-processes to the inferior frontal gyrus, and I argue that variation at stage is captured by the alternate names variable. This basic claim was supported in the monolingual study (Chapter 4) where alternate names significantly exceeded name agreement in its correlation with activity in PTri. Extending the search for correlates of selection to anterior cingulate cortex proved fruitful for both bilingual and monolingual speakers of Chinese (Chapter 5), although pars results in PTri were not as clear in this study.

Untangling the role of each region in lexical selection is still a matter of theoretical debate. Price (2012) directly links ACC with suppression of competitors, but Abutalebi & Green (2007; see also Abutalebi et al., 2011) limit ACC's function to conflict monitoring. Both accounts may be true, distinguished either by monolingual vs. bilingual status, or by an indirect role for ACC in evoking inhibitory responses for non-selected words. The euSEM connectivity model estimated in Chapter 5 revealed one important commonality between the monolingual and bilingual models to this end: a single, directed connection from ACC to PTri. This relationship may describe a top-down control mechanism in which ACC drives activity in PTri, which in turn exerts selective or inhibitory forces. No direct evidence for such a hierarchy is presented in this dissertation, but it is worth noting that ACC does act as a useful index for lexical category conflict. ACC’s responses to L1 and L2 category variables most closely matched behavioral
patterns while all significant correlates of PTri thus far have been positively correlated with Names. The detection or indexing role for ACC would fit Abutalebi and Green’s account of bilingualism and suggests further experiments to elaborate the range of PTri responses to lexical category variables and to activity in ACC.

**Conclusion**

For decades, descriptive research in lexical categorization has detailed wide variation in individuals’ lexical semantic mappings and languages’ lexical category norms. In more recent years, a concerted effort to understand the mechanisms of category learning and change has emerged, but this area of research remains relatively independent of the word production theories. With the extent and pervasiveness of idiosyncratic lexical semantic variation highlighted in lexical categorization research, it’s a wonder that modern neurocognitive accounts of language production have not attempted to incorporate these factors. In the present dissertation, I take first steps towards formalizing quantitative measures of lexical semantic variation with lexical category variables. I demonstrate that these measures are highly significant in predicting the naming behavior of bilinguals, and I identify some functional brain correlates of lexical category variables.

The behavioral and neuroimaging findings of this dissertation closely correspond with many dominant neurocognitive models of word production, particularly in the area of within- and cross-language competition and control. Lexical categorization doesn’t only conform to word production theories, but instead it offers significant elaborative opportunities for neurocognitive models. For example, few such models incorporate anterior cingulate cortex in monolingual lexical selection, but the present study provides compelling evidence for ACC’s involvement with both monolingual and bilingual conflict monitoring. As illustrated in these studies, only by
examining the full range of lexical semantic variability among native speakers of a language will we understand the neurocognitive underpinnings of word production.
References


doi:10.1017/S1366728903001020


## Appendix A

### Voxels Significantly Correlated with English Variables (Chapter 4)

<table>
<thead>
<tr>
<th>English Variables</th>
<th>MNI Coordinates</th>
<th>Cluster</th>
<th>Peak Voxel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td><strong>Positive correlates of Agreement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Superior Occipital</td>
<td>-12</td>
<td>-96</td>
<td>16</td>
</tr>
<tr>
<td>R Supramarginal Gyrus</td>
<td>64</td>
<td>-32</td>
<td>26</td>
</tr>
<tr>
<td>L Supramarginal Gyrus</td>
<td>-62</td>
<td>-24</td>
<td>30</td>
</tr>
<tr>
<td><strong>Negative correlates of Agreement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Supplementary Motor Area</td>
<td>-4</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>R Supplementary Motor Area</td>
<td>4</td>
<td>18</td>
<td>50</td>
</tr>
<tr>
<td>L Supplementary Motor Area</td>
<td>-8</td>
<td>16</td>
<td>48</td>
</tr>
<tr>
<td>L Supplementary Motor Area</td>
<td>-6</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td><strong>Positive correlates of Names</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Middle Temporal Gyrus</td>
<td>50</td>
<td>-62</td>
<td>2</td>
</tr>
<tr>
<td>R Supramarginal Gyrus</td>
<td>46</td>
<td>-34</td>
<td>24</td>
</tr>
<tr>
<td>R Rolandic Operculum</td>
<td>40</td>
<td>-30</td>
<td>18</td>
</tr>
<tr>
<td>L Rolandic Operculum</td>
<td>-46</td>
<td>-22</td>
<td>20</td>
</tr>
<tr>
<td>L Postcentral Gyrus</td>
<td>-52</td>
<td>-18</td>
<td>16</td>
</tr>
<tr>
<td>R Angular Gyrus</td>
<td>28</td>
<td>-66</td>
<td>46</td>
</tr>
<tr>
<td>L Middle Occipital</td>
<td>-44</td>
<td>-68</td>
<td>0</td>
</tr>
<tr>
<td>L Middle Occipital</td>
<td>-36</td>
<td>-82</td>
<td>4</td>
</tr>
<tr>
<td><strong>Negative correlates of Names</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Positive correlates of Typicality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Supramarginal Gyrus</td>
<td>64</td>
<td>-40</td>
<td>24</td>
</tr>
<tr>
<td>L Caudate</td>
<td>-6</td>
<td>12</td>
<td>-2</td>
</tr>
<tr>
<td>L Pallidum</td>
<td>-10</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>R Inferior Frontal Gyrus (pars orbitalis)</td>
<td>44</td>
<td>48</td>
<td>-2</td>
</tr>
<tr>
<td>R Middle Frontal Gyrus</td>
<td>42</td>
<td>16</td>
<td>48</td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus (pars triangularis)</td>
<td>-38</td>
<td>44</td>
<td>4</td>
</tr>
<tr>
<td>L Cuneus</td>
<td>2</td>
<td>-84</td>
<td>32</td>
</tr>
<tr>
<td>L Inferior Parietal</td>
<td>-40</td>
<td>-54</td>
<td>54</td>
</tr>
<tr>
<td>L Inferior Parietal</td>
<td>-48</td>
<td>-52</td>
<td>50</td>
</tr>
<tr>
<td>R Superior Temporal Gyrus</td>
<td>60</td>
<td>-30</td>
<td>20</td>
</tr>
<tr>
<td>R Superior Temporal Gyrus</td>
<td>50</td>
<td>-32</td>
<td>18</td>
</tr>
<tr>
<td><strong>Negative correlates of Typicality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Positive correlates of Familiarity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Calcarine</td>
<td>14</td>
<td>-100</td>
<td>4</td>
</tr>
<tr>
<td><strong>Negative correlates of Familiarity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Thalamus</td>
<td>-8</td>
<td>-28</td>
<td>8</td>
</tr>
<tr>
<td>R Thalamus</td>
<td>10</td>
<td>-30</td>
<td>6</td>
</tr>
<tr>
<td>R Middle Temporal Gyrus</td>
<td>46</td>
<td>-60</td>
<td>-4</td>
</tr>
<tr>
<td>L Heschl's Gyrus</td>
<td>-40</td>
<td>-16</td>
<td>8</td>
</tr>
<tr>
<td>R Middle Frontal Gyrus</td>
<td>34</td>
<td>50</td>
<td>24</td>
</tr>
<tr>
<td>R Insula</td>
<td>40</td>
<td>-18</td>
<td>10</td>
</tr>
<tr>
<td>R Insula</td>
<td>40</td>
<td>-8</td>
<td>2</td>
</tr>
<tr>
<td>R Supplementary Motor Area</td>
<td>8</td>
<td>6</td>
<td>54</td>
</tr>
<tr>
<td>R Cuneus</td>
<td>22</td>
<td>-68</td>
<td>22</td>
</tr>
<tr>
<td>L Inferior Occipital Lobe</td>
<td>-42</td>
<td>-66</td>
<td>-4</td>
</tr>
<tr>
<td>L Insula</td>
<td>-28</td>
<td>20</td>
<td>-6</td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus (pars orbitalis)</td>
<td>-34</td>
<td>30</td>
<td>-4</td>
</tr>
<tr>
<td>R Middle Temporal Gyrus</td>
<td>46</td>
<td>-74</td>
<td>20</td>
</tr>
<tr>
<td>R Fusiform Gyrus</td>
<td>32</td>
<td>-46</td>
<td>-6</td>
</tr>
</tbody>
</table>
## Appendix B

### Voxels Significantly Correlated with Chinese Variables (Chapter 4)

<table>
<thead>
<tr>
<th>Chinese Variables</th>
<th>MNI Coordinates</th>
<th>Clusters</th>
<th>Peak Voxel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td><strong>Positive correlates of Agreement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Superior Parietal Lobe</td>
<td>18</td>
<td>-60</td>
<td>64</td>
</tr>
<tr>
<td>R Superior Parietal Lobe</td>
<td>34</td>
<td>-58</td>
<td>62</td>
</tr>
<tr>
<td>R Precuneus</td>
<td>12</td>
<td>-62</td>
<td>54</td>
</tr>
<tr>
<td>R Superior Frontal Gyrus</td>
<td>30</td>
<td>0</td>
<td>66</td>
</tr>
<tr>
<td>R Postcentral Gyrus</td>
<td>48</td>
<td>-30</td>
<td>50</td>
</tr>
<tr>
<td>R Precentral Gyrus</td>
<td>48</td>
<td>-34</td>
<td>60</td>
</tr>
<tr>
<td>L Precentral Gyrus</td>
<td>-40</td>
<td>-12</td>
<td>60</td>
</tr>
<tr>
<td>L Postcentral Gyrus</td>
<td>-48</td>
<td>-12</td>
<td>56</td>
</tr>
<tr>
<td>R Precentral Gyrus</td>
<td>60</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>R Postcentral Gyrus</td>
<td>52</td>
<td>-10</td>
<td>42</td>
</tr>
<tr>
<td>R Postcentral Gyrus</td>
<td>56</td>
<td>-6</td>
<td>36</td>
</tr>
<tr>
<td><strong>Negative correlates of Agreement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus (pars opercularis)</td>
<td>-38</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus (pars triangularis)</td>
<td>-44</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus (pars triangularis)</td>
<td>-38</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td><strong>Positive correlates of Names</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Postcentral Gyrus</td>
<td>-52</td>
<td>-8</td>
<td>50</td>
</tr>
<tr>
<td>L Inferior Occipital Lobe</td>
<td>-34</td>
<td>-8</td>
<td>28</td>
</tr>
<tr>
<td>L Anterior Cingulum</td>
<td>-2</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>R Postcentral Gyrus</td>
<td>56</td>
<td>-6</td>
<td>38</td>
</tr>
<tr>
<td><strong>Negative correlates of Names</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Positive correlates of Typicality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Cuneus</td>
<td>-10</td>
<td>-92</td>
<td>16</td>
</tr>
<tr>
<td>L Cuneus</td>
<td>-6</td>
<td>-8</td>
<td>28</td>
</tr>
<tr>
<td>R Superior Temporal Gyrus</td>
<td>62</td>
<td>-38</td>
<td>8</td>
</tr>
<tr>
<td>R Superior Temporal Gyrus</td>
<td>54</td>
<td>-34</td>
<td>6</td>
</tr>
<tr>
<td>R Inferior Parietal Lobe</td>
<td>44</td>
<td>-56</td>
<td>50</td>
</tr>
<tr>
<td><strong>Negative correlates of Typicality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus (pars triangularis)</td>
<td>-42</td>
<td>34</td>
<td>14</td>
</tr>
<tr>
<td>R Fusiform Gyrus</td>
<td>34</td>
<td>-40</td>
<td>14</td>
</tr>
<tr>
<td>R Fusiform Gyrus</td>
<td>36</td>
<td>-46</td>
<td>-8</td>
</tr>
<tr>
<td>L Fusiform Gyrus</td>
<td>-30</td>
<td>-44</td>
<td>12</td>
</tr>
<tr>
<td><strong>Positive correlates of Familiarity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Calcarin</td>
<td>-6</td>
<td>-100</td>
<td>2</td>
</tr>
<tr>
<td>L Middle Occipital Gyrus</td>
<td>-20</td>
<td>-100</td>
<td>10</td>
</tr>
<tr>
<td>R Middle Occipital Gyrus</td>
<td>26</td>
<td>-98</td>
<td>6</td>
</tr>
<tr>
<td>R Superior Occipital Gyrus</td>
<td>20</td>
<td>-96</td>
<td>22</td>
</tr>
<tr>
<td>R Cuneus</td>
<td>14</td>
<td>-100</td>
<td>12</td>
</tr>
<tr>
<td><strong>Negative correlates of Familiarity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Middle Cingulum</td>
<td>4</td>
<td>-12</td>
<td>32</td>
</tr>
<tr>
<td>L Supplementary Motor Area</td>
<td>-8</td>
<td>-4</td>
<td>56</td>
</tr>
<tr>
<td>L Superior Parietal Lobe</td>
<td>-22</td>
<td>-58</td>
<td>48</td>
</tr>
<tr>
<td>R Precuneus</td>
<td>14</td>
<td>-70</td>
<td>46</td>
</tr>
<tr>
<td>L Precentral Gyrus</td>
<td>-50</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>R Insula</td>
<td>38</td>
<td>24</td>
<td>6</td>
</tr>
</tbody>
</table>
VITA
Benjamin D. Zinszer

Education
2014  Ph.D., Psychology, Pennsylvania State University
2012  M.S., Psychology, Pennsylvania State University (University Park, Pennsylvania)
2008  B.A., Psychology, Wheaton College (Wheaton, Illinois)

Fellowships & Honors
2013-14  Pennsylvania Space Grant Consortium Graduate Fellow
2013  Multi-modal Neuroimaging Training Program participant
2012  Delegate to 2nd Young Scientist Forum, US-China Consultation on People-to-
        People Exchange
2012  NSF East Asia and Pacific Summer Institutes Fellow
2012  Mu Sigma Rho, Statistics Honor Society, Penn State chapter
2011  Center for Language Science PIRE Graduate Fellow
2009  University Graduate Fellowship, Pennsylvania State University

Refereed Articles
        learning.
Zinszer, B. D., Rolotti, S. V., Li, P. (under revision). Bayesian word learning in multiple language
        environments.
        categorization reflects individual language history and linguistic community norms. Frontiers
Zinszer, B. D. & Weiss, D. J. (2013). When to hold and when to fold: Detecting structural
        changes in statistical learning. Proceedings of the 35th Annual Conference of the Cognitive
        experience modulates functional brain network for the native language production in bimodal
        R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science

Reviewer
Behavior Research Methods
Bilingualism: Language and Cognition
Workshop on Bilingualism and Cognitive Control (abstracts)
The Modern Language Journal