BILINGUAL LEXICAL MEMORY: TOWARDS A PSYCHOLINGUISTIC MODEL OF ADULT L2 LEXICAL ACQUISITION, REPRESENTATION, AND PROCESSING

by

Matthew Finkbeiner

A dissertation Submitted to the Faculty of the

GRADUATE INTERDISCIPLINARY PROGRAM IN SECOND LANGUAGE ACQUISITION AND TEACHING

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

In the Graduate College

THE UNIVERSITY OF ARIZONA

2002

UMI Number: 3073282

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Matthew Finkbeiner

entitled Bilingual Lexical Memory:
Towards a Psycholinguistic Model of Adult L2
Lexical Acquisition, Representation, and Processing

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

	11/13/02 Date
	11-15-02 Date
	11/13/02 Date
	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director	12/03/02 Date

STATEMENT BY AUTHOR

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Janet Nicol. By allowing me to establish a close and collaborative working relationship with her over the past four years, I have learned more about each facet of the research process than I thought possible. She has been a mentor in every sense of the word, and has provided me with the standard upon which I hope to model my future working relationships. I also want to thank Ken Forster. Ken has a tireless passion to teach the art and science of psycholinguistic research, and I have benefited from his open-door policy immensely.

There have been several additional people who have contributed to this work, some more directly than others. None of the work done with the Japanese-English bilinguals would have been possible without Kumiko Nakamura's assistance. She spent countless hours helping me develop materials and recruit and run subjects on four different occasions – I owe Chapter 4 to her. Merrill Garrett and the Cognitive Science Program have provided financial assistance on several occasions to pay for subjects and to travel to conferences. When I just had to get out of the lab, Sonya Bird was always willing to throw a Frisbee with me. And the University of Arizona Ultimate team provided the 'ultimate' opportunity to let off steam. Finally, I want to thank Tricia Stover for her patience, support and understanding – thank you!

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Present models of bilingual lexical processing assume common meaning representations between lexicons. The nature of these representations is such that a single meaning "node" or "set of nodes" is thought to subserve L1 and L2 translation-equivalent forms. Models of this type face two critical problems. First and foremost is the very real problem that there are very few true translation equivalents. Not only do translation "equivalents" frequently mean slightly different things, but quite often they can be used language specifically in ways the translation equivalent is unable to capture. The second problem facing these models is asymmetrical lexical performance between languages in translation priming tasks. For example, priming is robust in the L1→L2 direction, but not in the L2→L1 direction. Models assuming a symmetrical relationship between a common meaning node (or set of nodes) and translation-equivalent lexical forms cannot provide a straightforward account of these phenomena.

In the present thesis I propose the sense model, which holds that meaning representations are comprised of distinct semantic senses, some of which may be shared across languages. A representational asymmetry is assumed between lexicons, such that, on average, L1 forms are associated with more semantic senses than L2 forms. Initially, L2 forms are associated with a restricted number of semantic senses that have been extracted from the semantic entry of the L1 translation equivalent. Later on in L2 lexical development, semantic senses specific to the L2 are incorporated into the semantic entries of L2 lexical items. The value of the sense model comes in its ability to account straightforwardly for (one) how translation "equivalents" can be used language

specifically in ways not captured by its translation (the particular sense is not shared across languages); and (two) the patterns of asymmetrical lexical performance between languages. Because many of the senses represented in L2 entries are also represented in their L1 equivalent, the proportion of L2 senses activated by the L1 equivalent is large, if not complete. Conversely, because there are many senses represented in L1 entries that are not similarly represented in the L2 equivalent entry, the proportion of L1 senses activated by the L2 equivalent is very small. Hence, the translation priming asymmetry is argued to be the logical consequence of the representational asymmetry assumed by the sense model.

Chapter 1 Introduction

This thesis is concerned quite broadly with bilingual lexical memory. Bilingual lexical performance in several different tasks, including vocabulary learning, translation, picture naming, and masked translation priming, are all brought to bear on this issue. The central question motivating the research reported in this thesis is the following: can a single, parsimoniously constructed psycholinguistic model of bilingual lexical representation and processing account for bilingual lexical performance at all levels of proficiency?

Before discussing lexical processing, it is important to discuss how lexical information is thought to be represented. In this thesis, I will be assuming that a lexical item is a tripartite structure comprised of form-level information (phonological and/or orthographic), syntactic information, and conceptual information. For example, the information registered in long term memory for the lexical item "cat" would be represented in three distinct "levels" (see Figure 1-1). This hierarchical distinction is justified both logically and empirically (Garrett, 1975, 1988; Jackendoff, 1997; Levelt,

/kæt/	form-level information
+N, -V singular	syntactic information
[$_{\text{Thing}}$ CAT, TYPE OF ANIMAL, etc.]	conceptual information

Figure 1-1: Structure of a lexical item (adapted from Jackendoff, 1997, p. 84).

1989; Levelt, Roelofs and Meyer, 1999).

Given this hierarchical structure, at least three different logical possibilities present themselves in terms of how lexical information could be represented in the bilingual lexicon. For example, it may be that newly established L2 lexical forms map directly onto already-existing information at other levels of representation. Equally plausible is the notion that completely new information at all levels of representation is established during L2 vocabulary learning. A third logical possibility is that information at some level (or all levels) of representation is distributed such that the components in common between L1 and L2 lexical entries may be shared, while those components that are necessarily distinct would not be shared. A review of this literature reveals that the wide number of bilingual lexical processing models proposed over the years can be organized into one of three different classes of models; the classes are delineated by the constraints of these three logically plausible hierarchical architectures.

1.1 Three classes of models

When an adult speaker of one language begins to learn vocabulary items for a second language, there are several different possible ways in which new L2 lexical information can be represented and incorporated into the already-established lexico-semantic system. These different architectural possibilities can be grouped into the following classes of models: common meaning models, distinct meaning models and distributed meaning models. Common to all of these models is the notion that learning a new L2 word involves establishing a new L2 form (orthographic or phonological) in memory and then associating that form with appropriate meaning representations. Where the three classes of models diverge from one another is in determining how the new L2 lexical form comes to be associated with its appropriate meaning representation.

1.1.1 Distinct meaning models

The distinct meaning models assume that along with the establishment of new L2 form level representations, new L2 meaning-level representations are formed during L2 vocabulary learning as well. This class of models is best represented by the coordinate model (Weinreich, 1953). According to this model, two completely isolable systems develop, with no connections or overlap existing between them. Coordinate bilingualism was thought to emerge when there was a strict separation between environments in which the bilingual's two languages were learned and used. For instance, coordinate bilingualism was thought to emerge when one would use language A exclusively at home and language B exclusively at work or school, or when languages A and B were learned and used in distinct cultural and national settings. When this is the case, Ervin and Osgood (1954) argued, "...the set of linguistic signs and responses appropriate to one language come to be associated with one set of representational mediating processes [i.e.

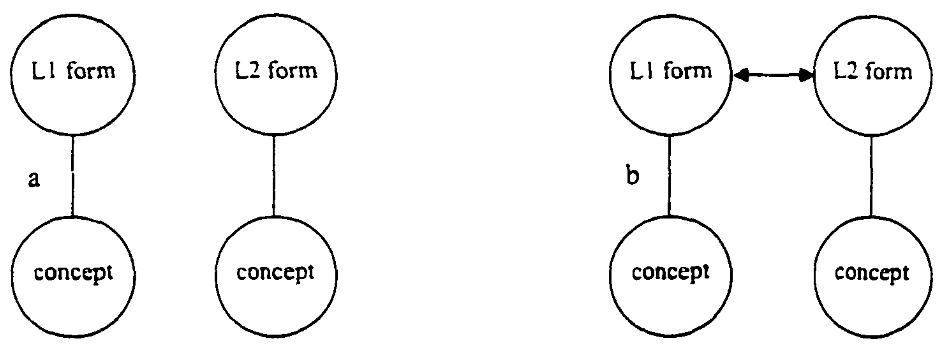

Figure 1-2: Distinct Meaning class of models: a) coordintate model; b) logically possible distinct meaning model with interlingual lexical associations

meaning representations] ...but the set of linguistic signs and responses appropriate to the other language become associated with a *somewhat different* set of representational processes..." (p. 140; emphasis in the original).

The coordinate model enjoyed some early success (e.g. Ervin and Osgood, 1954; Lambert, Havelka and Crosby, 1958) in accounting for the performance of bilinguals, but later studies found that the distinct meaning hypothesis did not stand up to empirical investigations very well. Consequently, the coordinate model soon fell out of favor with researchers. When Potter et al. (1984) conducted their seminal study that pit the predictions of the word association (subordinative) and concept mediation (compound) models against each other, they did not include a test of the predictions made by the coordinate model in their study.

Though never proposed before, a logically possible model within this class of models would be one in which interlingual lexical (form-level) associations are allowed (see Figure 1-2b).

1.1.2 Common meaning models

The common meaning class of models assumes that translation-equivalent L1 and L2 forms map onto the same meaning representation. In other words, new meanings are not thought to be created when learning new L2 vocabulary; rather, the new L2 form maps onto an old meaning representation, which comes to subserve both L1 and L2 equivalent forms. There are three commonly cited models within this class of models: the word association model, the concept mediation model and the revised hierarchical model (see Figure 1-4). The word association model (Potter, So, Von Eckardt &

Feldman, 1984), originally termed the subordinative model by Weinreich (1953), assumes that new L2 forms access meaning via their equivalent L1 form. That is, according to the architecture of this model, there are no direct L2 form-meaning connections. The concept mediation model (Potter et al., 1984), which was originally referred to as the compound model (Weinrich, 1953), assumes direct L2 form-meaning connections with no form-level connections mediating equivalent L1 and L2 lexical representations. The revised hierarchical model (Kroll & Stewart, 1994) attempts to combine the word association and concept mediation models into a single developmental model. According to the revised hierarchical model (RHM), the bilingual lexicon develops from a representational structure best captured by the word association model at low levels of proficiency to a structure best captured by the concept mediation model at high levels of proficiency.

1.1.3 Distributed meaning models

The distributed meaning class of models assumes that meaning representations are comprised of a distributed set of conceptual "features" or "nodes". A well known model

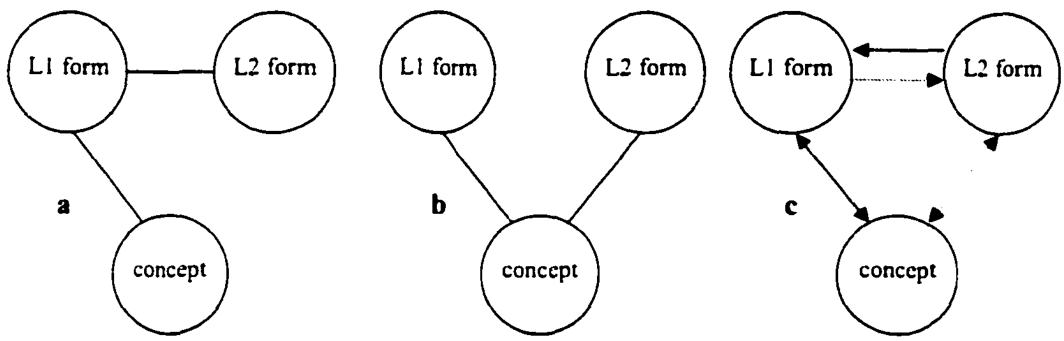

Figure 1-4: Common Meaning class of models: a) word association; b) concept mediation; c) revised hierarchical model.

within this class of models is De Groot's distributed feature model (1992). According to her model, there are varying degrees of overlap of these conceptual features between L1 and L2 meanings. The extent to which featural overlap occurs between L1 and L2 meanings depends upon what type of word is represented. De Groot (1992, 1993) has argued that there is more featural overlap between translation equivalents for concrete words than abstract words because concrete referents tend to have the same shape, size and function cross-linguistically. De Groot (1993) has referred to this as a "mixed" model because it is possible for the meaning representations to range from having complete featural overlap (essentially the architecture of a compound bilingual) to having no featural overlap (i.e. the architecture of a coordinate bilingual) depending upon the type of word in question.

A second possibility within this class of models, and one that to my knowledge

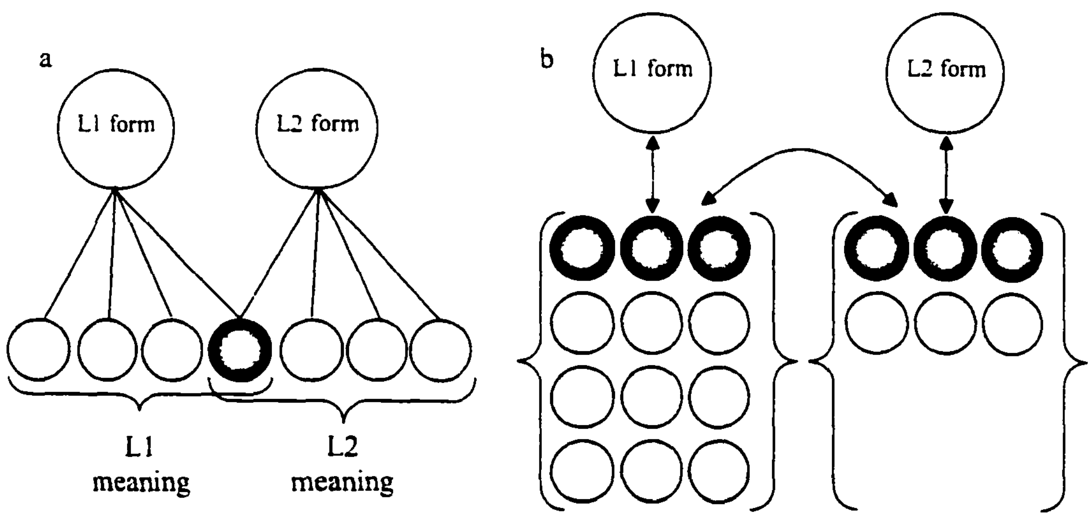

Figure 1-5: Distributed Meaning class of models: a) De Groot's distributed meaning model showing overlap of one feature; b) sense model depicting two translation equivalents that share one sense while having several other senses that are not in common between L1 and L2 lexical entries.

has not been articulated before, is the "sense" model (c.f. Figure 1-5b). According to this model, translation equivalents share a similar (or perhaps even identical) sense representation, but may still differ from each other if there are remaining senses represented within each lexical entry that are not shared between the "equivalents". For example, the English word "black" and the Japanese word "kuroi" are translation equivalents, despite having little besides the core sense (COLOR) in common. In English "black" is used to refer to a type of humor, as well as a calamitous day on Wall Street (*black Friday*). In Japanese, "black" can be used to refer to those who are evil-minded (*black belly*), as well as those who are well tanned or guilty of a crime.

The sense model differs from De Groot's distributed feature model in that the amount of meaning overlap between equivalents is determined by the number of senses in common between L1 and L2 equivalents, which is something that is determined on a word by word basis. This is an important difference from the distributed feature model (which assumes that the amount of meaning overlap between equivalents is determined by the type of word in question (e.g. concrete or abstract)), because of the very different predictions that the two models make in terms of cross-language lexical performance. Taking translation performance, for example, the distributed feature model predicts that concrete words will be translated faster than abstract words because concrete equivalents have more meaning features in common than do abstract equivalents. The sense model, on the other hand, assumes that when all else is held equal the ratio of senses in common between equivalents can affect translation times. That is, according to the sense model there will be faster translation times for equivalents with a high ratio of senses in common (e.g. 1:1) and slower translation times for equivalents with a low ratio of senses in common (e.g. 1:5). Not surprisingly, concrete translation equivalents tend to have fewer senses overall relative to abstract words, and concrete words tend to have more senses in common between equivalents across languages than abstract words do. Evidence relevant to the predictions of the sense model versus the predictions of the distributed feature model is discussed below.

1.2 Relevant evidence for each class of models

1.2.1 Distinct meaning models

The earliest in depth discussion of the coordinate model was provided by Ervin and Osgood (1954), whose largely theoretical paper is perhaps the most often cited paper on the similarities and differences of the Compound (common meaning) and Coordinate (distinct meaning) architectures. They argued that the structural difference between Compound and Coordinate bilingualism is the emergent consequence of differences between language learning contexts. For example, Compound bilingualism was thought to emerge from learning contexts in which the "signs" (i.e. forms) from two different languages shared the same referent. This was thought to occur when two languages were used simultaneously in the child's home, or when the L2 was learned at school. Ervin and Osgood (1954) argued that learning contexts such as these led to L1 and L2 forms being associated with the same set of "representational mediation processes" (i.e. meanings), resulting in a "compound" organization of the bilingual lexicon (p. 140).

Coordinate bilingualism, on the other hand, was thought to emerge when there was a strict separation between environments in which the bilingual's two languages were learned and used. Ervin and Osgood (1954) refer to this kind of development as being "typical of the 'true' bilingual ..." (p. 140). Because of the different contexts within which the two languages are learned and used, Ervin and Osgood reasoned that "The kinds of representational processes developed must then also be different and hence the meanings of the signs. This development can also characterize the second language learner, who, relying as little as possible on translation and immersing himself in the living culture of another language community, comes to speak a second tongue well" (p. 140). A test of the implicit prediction made in this quote by Ervin and Osgood, that second language learners in "sufficiently unique" learning conditions will come to represent new L2 information differently, is tested and reported in Chapter 3.

1.2.1.1 Word association tasks

Early on there was some tentative support for the hypothesis that the contexts in which the bilingual's two languages were learned and used led to performance differences predicted by the Compound/Coordinate distinction. Lambert, Havelka and Crosby (1958), for example, argued that for a French-English bilingual who lived in France for several years before moving to North America, the referent for the sign *eglise* would be a gothic cathedral while the referent for the sign "church" would be a tall wooden building used on Sundays. Using Osgood's (1952) "semantic differential", a method of scaling stimulus-words on a standard set of meaning dimensions, Lambert et al. (1958) found that bilinguals who acquired their languages in separated contexts showed a significantly greater difference in meanings of translated equivalents than did bilinguals who acquired their two languages in similar contexts. That is, those who were thought to be Coordinate bilinguals by virtue of distinct language learning contexts rated translation equivalents like *eglise* and "church" differently; those who were thought to be Compound bilinguals (because of identical language learning contexts for both of their languages) rated translation-equivalents like *eglise* and "church" similarly. This performance difference was predicted by the proposed structural distinction between the Compound and Coordinate architectures. If a bilingual's two languages are represented in the way proposed by the Compound architecture, then translation-equivalent signs will refer to the same "representational processes" and thus have identical meanings. It is rather straightforward, then, that a "compound bilingual" would rate translation-equivalent forms similarly and that a "coordinate bilingual" would rate translation-equivalent forms *differently* in a semantic differential task. The Coordinate architecture is thought to be such that signs (forms) in one language refer to one set of representational processes (concepts) and signs in the other language refer to a completely distinct set of representational processes. According to the predictions of the Coordinate model, translation-equivalent forms *cannot* have identical meanings.

Though the findings in the semantic differential task reported by Lambert et al. (1958) were taken as support for the Compound/Coordinate distinction, they did not find any difference in translation latencies between groups. This was very problematic for the proposed distinction between the Compound and Coordinate architectures. According to the predictions of the Compound and Coordinate architectures, translation between equivalent forms should be much easier and faster when those forms are represented in a Compound system than when they are represented in a Coordinate system. This is because in the Compound system equivalent forms are "connected" via a single concept that lies in common between the two equivalent forms. In the Coordinate system, on the other hand, translation-equivalent forms have no connections (direct or indirect) between them, which should lead to slow and difficult translation processes – requiring, perhaps, conscious recall on the bilingual's part in order to perform the task. The lack of a clear difference in translation latencies between "compound bilinguals" and "coordinate bilinguals" reported by Lambert et al. (1958) was problematic for the Compound/Coordinate distinction, despite the empirical support found for the distinction in the semantic differential task.

Gekoski (1980) reported a similar pattern of findings, with support for the Compound/Coordinate distinction when an association task was used, but no support for the distinction when response times were taken into account. Gekoski had both compound and coordinate bilinguals perform four different association tasks: two intralingual (in L1 and then in L2) and two interlingual. In a separate procedure, Gekoski had determined that the stimuli in one language (e.g. English) were translation equivalents of those presented in the other language (e.g. Spanish). Gekoski predicted that compound bilinguals would give higher proportions of equivalent responses to translation-equivalent stimuli than would the coordinate bilinguals across both the two intralingual association conditions as well as across the two interlingual conditions. The findings confirmed this prediction.

Though these findings support the proposed Compound/Coordinate distinction, the response latencies in the interlingual association conditions did not. Gekoski predicted that compound bilinguals would not only provide a higher percentage of equivalent L1 and L2 responses than coordinate bilinguals, but that they would do so faster than the coordinate bilinguals. This is a very straightforward prediction. Because it is presumed that equivalent forms are connected via a common conceptual representation in the Compound architecture but not in the Coordinate architecture, and because translation must be a major factor in the interlingual association conditions, "compound bilinguals" should perform the interlingual association task faster. They did not.

1.2.1.2 Repetition priming

A separate line of research, using cross-language repetition priming, provided early support for the notion that a bilingual's two languages are represented distinctly (this hypothesis has been referred to with a variety of different labels over the years, including: the *independence*, *separate storage*, and *dual-code* hypotheses). In studies with monolinguals, repetition priming has been observed in lexical decision tasks (Kirsner & Smith, 1974; Scarborough, Gerard, & Cortese, 1979) as well as in word-fragment completion tasks (Graf, Mandler, & Haden, 1982; Tulving, Schacter, & Stark, 1982). Typically, these studies involved participants first studying a list of words and then subsequently performing the lexical decision or word-fragment completion task. Because priming is argued to occur as a result of the prime and target both accessing the same lexical representation, once during the presentation of the prime and once during the presentation of the target, this experimental technique was ideal to test the "common" versus "distinct" hypotheses of bilingual lexical representation. The purpose of the cross-language repetition priming experiments was to test whether translation equivalent words were represented in the same language system or not. The occurrence of repetition priming between translation equivalents could be taken as evidence that the two words shared some of the same underlying representational specifications and, thus, belonged to the same "cognitive system". A failure to observe repetition priming across languages, though, despite observing strong effects within languages, would indicate that the two equivalent words were represented in distinct systems. Most studies failed to find any evidence of cross-language repetition priming in lexical decision (e.g. Gerard & Scarborough, 1989; Kirsner, Brown, Abrol, Chadha, & Sharma, 1980; Kirsner, Smith, Lockhart, King, & Jain, 1984; Scarborough, Gerard, & Cortese, 1984) or in word-fragment completion (Watkins & Peynircioglu, 1983; Durgunoglu & Roediger, 1987; Basden, Bonilla-Meeks & Basden, 1994).

It might be argued that repetition priming is only observed when the physical properties of the prime stimulus and target stimulus are highly similar, if not identical. Because the orthographies of translation equivalents are typically not similar, one may want to argue that this is the reason no priming was observed in these cross-language priming experiments. However, if repetition priming was an orthographic effect alone (as opposed to a lexical one), then similar nonwords should prime each other – but they do not (Friedrich, Henik, & Tzelgov, 1991). This suggests that the occurrence of repetition priming is due to the same lexical representations (instead of the same orthographic representations) being retrieved from memory. Consequently, because the locus of repetition priming effects is assumed to be at the level of lexical representations, and because cross language repetition priming effects were difficult (if not impossible) to observe, there came to be widespread agreement among researchers that lexical representations were language specific.

1.2.1.3 Repetition priming with short SOAs

Quite surprisingly, a minor variation on the priming technique revealed a contradictory pattern of priming effects, this time suggesting that perhaps the bilingual's two languages were represented in a common store. When researchers used a very short SOA along with no intervening lexical information appearing between prime and target (i.e. prime and target were presented in rapid succession), robust cross-language priming effects were observed (Chen & Ng, 1989; De Groot & Nas, 1991; Keatley, Spinks, & De Gelder, 1994; Kirsner et al., 1984; Schwanenflugel & Rey, 1986). This same pattern of dissociative effects had already been observed in priming studies with monolinguals when a short versus long SOA was used. Namely, it had already been observed that repetition priming effects (e.g. between *butter-butter*) survive over long durations as well as over intervening material, but that semantic priming effects (e.g. between *prince-boy*) dissipate very quickly and do not span across intervening material (Neely, 1991). Because of this, cross-language priming effects between translation equivalents were thought to resemble those observed in within language semantic priming experiments much better than they did the effects observed in within language repetition priming studies. In accordance with the notion that the locus of cross-language translation priming was semantic. Smith (1991) showed that cross-language repetition priming (word-fragment completion task) could be observed if the study conditions encouraged conceptual processing of the prime (i.e. by having the participants make an inference about the prime).

Particularly devastating evidence against the distinct meaning hypothesis came from a cross-language masked priming study, in which the participants were not aware of the prime at all. Using this masked priming paradigm, Williams (1994) revealed a dissociation between cross language associative and semantic priming by showing that L1-L2 priming effects were obtainable when the prime-target pairs were semantically related (fence-haie ["hedge" in French]) but not when the pairs were associatively related (e.g. shoe-pied ["foot" in French]). This study effectively won the debate between the distinct and common meaning hypotheses because it showed that the cross-language priming effect was in fact semantic. Due to the primes being masked, participants were not able to engage translation or predictive strategies. Also, because word pairs with strong within-language associations (but semantically dissimilar, like *needle-thread*) did not produce cross language priming, this study showed very effectively that the semantic information in common between cross language word pairs was both sufficient and necessary to produce cross-language priming effects.

In summary, it is clear that it would be very difficult to reconcile the predictions of the distinct meaning models with the data reviewed here. Although there is evidence from tasks such as word association and cross-language repetition priming (with long SOAs) that lexical representations are language specific, there is equally strong evidence to suggest that word forms from each language access a common semantic store. When very short SOAs are used in repetition priming tasks, so that the conditions of the task are identical to those in which within-language semantic priming effects occur, strong cross-language priming effects are observed between translation equivalent words. The simplest and most parsimonious way to make sense of this pattern of results is to assume a hierarchical organization, such as the one discussed above, where lexical forms are represented language specifically and where meaning specifications are represented in a common store. Evidence relevant to the hierarchical models that assume a common meaning store is discussed in the next section.

1.2.2 Common meaning models

Evidence bearing on the common meaning class of models comes from a variety of different sources. The debate within this literature, which by now is quite large, has not been over whether the hierarchical common meaning architecture is correct or not; rather, the focus of research has been on which model within this class is better able to account for the performance of bilingual lexical processing at all levels of proficiency. Findings germane to this debate come from a variety of different types of studies. In this section, I will discuss findings from acquisition studies, translation and picture naming studies, as well as priming studies.

1.2.2.1 Translation and picture naming studies

The first explicit test of two models within this class of models, the word association model and concept mediation model, that made use of online reaction time measurements was conducted by Potter, So, Von Eckardt, and Feldman (1984). Potter et al. (1984) pitted these two models against each other by comparing the performance of fluent bilinguals on picture naming and translation tasks. In order to name a picture, the object must be recognized, a conceptual representation that corresponds to the object needs to be activated, and then the lexical item for that concept can be retrieved and articulated. Picture naming, then, is essentially a two-step process. Potter et al. (1984) reasoned that if the word association model was correct, then L1-L2 translation should be much faster than naming pictures in L2 because translation could make use of the direct link between the L1 and L2 lexicons – a one-step process. Potter et al. similarly reasoned that if the predictions of the concept mediation model were better able to account for the performance of bilinguals, then participants should take approximately the same amount of time to perform both picture-naming in L2 as well as L1-L2 translation. This is because, according to the concept mediation model, translation and picture-naming alike are two-step processes involving first concept activation and then lexical retrieval. The results of their study confirmed the concept mediation hypothesis. Potter et al. (1984) found that participants took no longer to perform L2 picture-naming than they did to perform L1-L2 translation. This was true for two groups of participants: one was a highly proficient group of bilinguals, the other was a group of less proficient L2 learners.

Subsequently, studies by Kroll and her colleagues (e.g. Kroll, 1993; Kroll & Sholl, 1992; Kroll & Stewart, 1990, 1991, 1994) challenged the conclusion that bilinguals are 'concept mediators' at all levels of proficiency. In a series of studies, Kroll and Stewart (1990, 1991, 1992 & 1994) reported a translation asymmetry such that L1-L2 translation was slower than L2-L1 translation, a finding that the concept mediation model, as it was specified by Potter et al. (1984) could not account for. Furthermore, Kroll and Stewart reported that L1-L2 translation was sensitive to semantic interference while L2-L1 translation was not. Kroll and Stewart found that L1-L2 translation was slower when to-be-translated items were grouped according to semantic category than when they were translated in random order. Translation in the L2-L1 direction was insensitive to this manipulation. Again, this was a finding that the concept mediation model could not account for because it predicted equal sensitivity to semantic interference in each translation direction. Kroll and Stewart hypothesized that translation followed different routes depending on the direction: L1-L2 translation was proposed to be conceptually mediated, while L2-L1 translation was proposed to be lexically mediated. Kroll and Stewart (1992, 1994) put forth the revised hierarchical model (RHM) in an attempt to explain the performance of bilinguals in translation tasks.

The RHM combines both the word association and concept mediation model into a developmental model of L2 lexical processing. Early on in development, the RHM assumes that L2 forms access meaning via L1 equivalents – essentially the architecture proposed by the word association model. This results in strong asymmetric connections at both the lexical and conceptual levels. That is, L2 forms have strong L2→L1 lexical-level connections but L1 forms have weak (if any) L1→L2 lexical level connections. Also, L2 form-meaning connections are very weak (non-existent, actually) initially, while L1 form-meaning connections are strong. As a learner becomes more proficient, L2 forms become more capable of accessing meaning directly, which serves to diminish the degree of asymmetry over time, resulting in performance best characterized by the concept mediation model. Kroll and her colleagues have shown the RHM to be very successful in accounting for a wide range of bilingual lexical processing data, but it too has had its challenges.

The RHM assumes that because L2 forms relied on L1 equivalents to access meaning initially, there are strong residual L2-L1 form-level connections. Even in highly proficient bilinguals, L2 words still strongly activate their respective L1 translation equivalents, and only weakly activate meaning-based representations (Kroll, Michael & Sankaranarayanan, 1998, p. 371). L1 words, on the other hand, will weakly (if at all) activate their L2 translation equivalents through form-level connections, but will strongly activate meanings. This is how Kroll and Stewart (1994) explained the presence of semantic interference effects in the L1-L2 translation direction when items were grouped according to semantic category and the absence of any interference effects in L2-L1 translation despite the same blocking manipulation. Unfortunately for the RHM, the pattern of results that Kroll and Stewart (1994) reported has not been replicated. La Heij, Hooglander, Kerling, and van der Velden (1996) found semantic interference effects in both translation directions. Furthermore, contrary to the predictions of the RHM, the semantic interference effect in this study was greatest in magnitude in the L2-L1 translation direction. The RHM would allow that for highly proficient bilinguals a semantic interference effect could be observed in both translation directions, but because of the strong residual asymmetries, it would never predict a larger semantic effect in the L2-L1 direction.

In another study, de Groot and Poot (1997) found that nonfluent bilinguals at a very low level of proficiency were sensitive to conceptual factors of to-be-translated words in both translation directions. More recently, Vigliocco, Lauer, Damion and Levelt (2002), using essentially the same experimental design (items blocked according to category) and subject pool as Kroll and Stewart (1994), found semantic category effects in the L2-L1 translation direction, where Kroll and Stewart (1994) reported none.

Considering carefully the performance of bilinguals in translation tasks at different levels of language proficiency, the evidence in support of direct L2 form-meaning connections (even at low levels of proficiency) seems to outweigh the evidence against such direct connections. In an effort to resolve these apparently contradictory findings, Green (1986, 1993, 1998) has introduced the Interference Control hypothesis. Basically, Green assumes the architecture of the concept mediation model along with the notion that in order to perform a production task, like picture naming or translation, interference from the non-target language must be suppressed. According to the IC model, suppressing L1 interference is more difficult than suppressing L2 interference. Consequently, production in the L2 will be more effortful (and hence slower) than L1 production. This is how Green accounts for the speed asymmetry in translation. With increased proficiency, the bilingual's ability to modulate interference from the L1 system increases, allowing the bilingual to perform L2 production faster and with less effort. Green accounts for Kroll and Stewart's finding that semantic interference occurs in L1-L2 translation and not the reverse by saying that their manipulation of categorizing items served to increase the number of semantically related competitors and that bilinguals are less able to control such competition in their L2. Note, however, that this account is very different from Kroll and Stewart's in that it predicts semantic interference in both translation directions, only assuming that the interference is regulated differently in the two language production systems. According to Kroll and Stewart's account, semantic interference is not possible with beginning level bilinguals due to the lack of direct L2 form-meaning connections. This is an issue taken up in the next section.

1.2.2.2 Vocabulary learning studies

The relevance of vocabulary learning studies to models of bilingual lexical representation and processing comes in comparing the performance data of novice learners with the assumptions made by each of the models in terms of how newly established L2 lexical information comes to be represented and subsequently processed. Within this class of models, only the RHM is a developmental model. That is, only the RHM assumes a fundamentally different architecture, and, hence, makes different predictions, for novice L2 learners and advanced bilinguals. Unfortunately, this (developmental) aspect of the RHM has not been tested as thoroughly as other aspects of the model. In fact, I am only aware of one study that was designed explicitly to test the developmental aspects of the RHM.

Altarriba and Mathis (1997) conducted a set of experiments in which monolinguals learned Spanish-English translations. After the translations had been learned to a pre-determined criterion, they tested their participants with a variety of different tasks, including translation recognition (L2-L1) with orthographically related foils (Experiment 1a), semantically related foils (Experiment 1b), as well as a Stroop color-word task (Experiment 2). Altarriba and Mathis (1997), in accordance with the predictions of the RHM, hypothesized that their novice learners would exhibit interference effects when orthographically related foils were used compared to unrelated control words because of the assumption that lexical level connections are established first in second language acquisition. Similarly, they assumed that semantically related foils would not produce significant interference effects with this group of novice learners because, according to the predictions of the RHM, direct L2 form-meaning connections are not established during the earliest stages of bilingual language development. Altarriba and Mathis also included a group of fluent bilinguals in each experiment. In contrast to the predictions made for the novice learners, they predicted that the fluent bilinguals would exhibit interference for semantically and orthographically related foils because the RHM indicates strong L2 form-meaning connections as well as strong residual lexical level connections for proficient bilinguals.

In Experiments 1a and 1b, Altarriba and Mathis found that novice learners and advanced bilinguals alike exhibited interference effects for both orthographically related and semantically related foils. That is, all participants, regardless of proficiency levels, took longer to reject both orthographically and semantically related foils than they took to reject unrelated control words in each condition. Although this finding was rather convincing in showing that even the novice learners in their experiment were processing the semantic properties of L2 lexical forms, there still existed the possibility that the semantic interference effect arose *after* translation had taken place. In order to address this possibility, they used the Stroop task in Experiment 2. Altarriba and Mathis reasoned that if novice learners exhibited a Stroop effect both within as well as between languages, then it would show that their participants had formed direct L2 form-meaning connections even for just-learned L2 words. Altarriba and Mathis found that both groups of participants (novice and advanced alike) produced significant Stroop effects both within and between languages. They conclude their paper by suggesting that even novice learners represent newly acquired words at both a lexical and conceptual level of representation, which, they point out, contradicts the predictions made by the RHM for beginning-level bilinguals.

This topic is taken up again in Chapter 2, where I describe a study in which monolingual participants learned a set of new L2 vocabulary items under two different conditions: one where the items were learned in semantically grouped sets, and one where the same items were learned in no particular order. At test, participants were asked to translate their newly learned items both in the L1-L2 and L2-L1 directions. It is widely assumed in the SLA literature that learning is enhanced when targets are grouped according to semantic categories. The purpose of this experiment was to test this assumption by determining the effect of increasing activation within semantic categories during learning has on subsequent lexical performance. According to the predictions of the RHM, a manipulation of semantic-level properties should have minimal effects, if any at all, on L2 word learning. Contrary to the assumptions widely held in the field of SLA and the predictions of the RHM, the findings reported in Chapter 2 indicate that learners' vocabulary learning performance suffered when items were learned in semantic sets.

In summary, within the common meaning class of models, only the RHM makes clear predictions in terms of word learning. Unfortunately for the RHM, the few data that have been collected on this aspect of the model do not sit well with the predictions of the RHM. The next section will similarly reveal that the predictions made by the RHM in terms of cross-language priming effects do not sit well with the performance of bilinguals.

1.2.2.3 Cross language priming studies

Cross-language priming has been one of the most productive experimental paradigms in investigating the nature of bilingual lexical representation and processing (e.g., Chen & Ng, 1989; De Groot & Nas, 1991; Frenck & Pynte, 1987; Keatley, Spinks, & De Gelder, 1994; Kirsner, Smith, Lockhart, King, & Jain, 1984; Meyer & Ruddy, 1974; Williams, 1994; Schwanenflugel & Rey, 1986; Tzelgov & Eben-Ezra, 1992). Although these studies provide strong support for the most general form of the common meaning hierarchical architecture hypothesis, where distinctly represented L1 and L2 lexical forms both access a commonly represented semantic store, these studies have not provided clear evidence as to which of the common meaning models is better suited to explain cross-language priming.

For example, in cross-language priming experiments using visible primes and the lexical decision task, a well-known asymmetry exists. Essentially, L1 words have been found to be very effective as primes for L2 targets, producing a large and reliable priming effect, while L2 primes have been found to be much less effective as primes for L1 targets, producing a dramatically reduced priming effect (Schwanenflugel & Rey, 1986;

Frenck & Pynte, 1987; Chen & Ng, 1989; Jin, 1990; Altarriba, 1992; Keatley et al., 1994). Later experiments that made use of the masked priming paradigm found an even stronger priming asymmetry. These later studies found that although masked primes in the dominant language (L1) prime translation equivalent words in the non-dominant language (L2), there is little or no priming in the reverse direction, unless the prime and target words are cognates from same-script languages (e.g., rich-rico) (Sanchez-Casas, Davis & Garcia-Albea, 1992), despite the fact that the very same L2 primes are capable of priming L2 targets (Gollan, Forster & Frost, 1997; Jiang, 1999).

This pattern of findings is clearly very perplexing within the context of the common meaning models. The concept mediation model would first have to say that the locus of cross-language priming effects is conceptual, and second that priming effects, when observed, would have to be symmetrical between L1-L2 and L2-L1 directions. The data clearly do not support the predictions of the concept mediation model. The word association model, on the other hand, would have to say that the locus of cross-language priming effects is at the lexical level (since this is the only route that connects the two lexicons), and that interlingual priming effects must be symmetrical. Again, the data clearly do not support the prediction of the word association model. The RHM, on the other hand, has had some success in accounting for the priming asymmetry (c.f. Kroll & Tokowicz, 2001). According to Kroll and her colleagues, the RHM can explain the priming asymmetry by suggesting that (1) the locus of the cross-language priming effect is conceptual and (2) that relative to L1 representations, L2 lexical representations are only weakly connected to the conceptual system. That is, priming is effective in the L1-

L2 direction because the masked L1 prime serves to activate a shared conceptual node, which then preactivates the L2 translation-equivalent lexical form, thus facilitating recognition. However, priming is not effective in the L2-L1 direction because L2 primes do not automatically activate their conceptual representations, resulting in no preactivation of the L1 translation-equivalent form and thus no priming.

This account is appealing, but it suffers from a number of weaknesses. First, the nature of the connection between an L2 lexical form and its associated concept needs to be better explained because if the locus of the priming effect is at the level of concepts, and L2 primes are too weakly connected to conceptual representations to be effective as masked primes, then we should not observe within-language priming with masked L2 primes – yet L2-L2 masked priming effects are very robust (Gollan et al., 1997, Jiang, 1999). Second, the Revised Hierarchical Model (RHM) posits that in order to obtain L2-L1 priming, the necessary links are L2-C and C-L1, where C represents the relevant conceptual node. To explain why there is no priming, the model assumes that the L2-C link is too weak. However, the relevant links for L1-L2 priming are L1-C and C-L2, and since there is priming in this case, it must be assumed that the C-L2 link is strong enough to permit priming. Thus, the asymmetry in priming would have to be explained by an asymmetry in the links between L2 words and conceptual nodes: The C-L2 link must be stronger than the L2-C link. As currently specified, the RHM assumes a symmetrical (weak) connection between L2 forms and their associated concepts; but note how further specification of the L2 form-meaning connection would prima facie require the L2-C link to be stronger than the C-L2 link because comprehension precedes production in L2 language development and comprehension makes use of the L2-C link, while production makes use of the C-L2 link. It is not clear, then, how, even with further specification, the "connection strengths" hypothesis could account for the priming asymmetry as well as the normal development of L2 comprehension and production skills.

A further problem for an account expressed in terms of strengths of connections is the possibility that the priming asymmetry may be task specific. Jiang & Forster (2001) claim to have obtained L2-L1 priming for non-cognates in a speeded episodic recognition task. Also, although Grainger & Frenck-Mestre (1998) reported that "The responses of English-French bilinguals performing semantic categorization and lexical decision tasks were facilitated by [L2] prime stimuli that were non-cognate translation equivalents of the [L1] targets…" (p. 601), a careful reading of their paper reveals that significant L2-L1 priming effects were observed in the semantic categorization task, with only a trend towards priming in the lexical decision task. These results are intriguing because they suggest that the type of task employed may reveal differential sensitivities to the cross-language priming effect.

In summary, even though the strongest evidence in support of the common meaning hypothesis has come from cross-language priming experiments, none of the individual models within this class of models have been particularly successful in accounting for the cross-language priming asymmetry. As will be seen below, the distributed meaning models, particularly the sense model, have had a much easier time of accounting for this particular set of findings.

1.2.3 Distributed meaning models

Evidence bearing on the distributed meaning class of models is quite limited compared to the amount of evidence speaking to the common meaning models. The majority of work done specifically on this class of models has been conducted by De Groot and her colleagues, and she has primarily made use of translation studies to test the predictions of the distributed feature model. For this reason, there is relatively little evidence bearing specifically on the distributed meaning class of models, especially in the domain of vocabulary learning or cross-language priming studies. Nevertheless, both De Groot's distributed feature model and the sense model make specific predictions with regards to cross-language priming, bilingual lexical development and translation performance. Hence, in the following section I will discuss how both of the distributed meaning models reviewed here would account for the available data in each of these areas.

1.2.3.1 Translation and picture naming studies

The strongest evidence to date in support of De Groot's distributed feature model has come from studies looking at word type effects in translation. A fundamental tenet of De Groot's distributed feature model is that cross-language translation performance is sensitive to the type of word being translated. For example, translation equivalent concrete words are thought to share more conceptual features across languages than abstract words, and thus are predicted to be translated faster than abstract words. This is a prediction that sets De Groot's distributed feature model apart from the common meaning models. The common meaning models assume that the same meaning representation is shared between translation-equivalent L1 and L2 forms, and thus must assume a general connectivity between equivalent forms and their meaning representation that is consistent across word types. That is, when all other factors are held constant (e.g. word frequency and length), the common meaning models would not predict word type effects in translation. In a series of experiments, De Groot (1992, 1994, 1998) and her colleagues have shown that bilingual participants respond more quickly to concrete than abstract words on a variety of different lexical processing tasks including: (1) normal translation production from L1 to L2 and from L2 to L1, (2) cued translation production from L1 to L2, which involves a single L1 word being presented along with the first letter of the expected L2 response, (3) translation recognition, where possible translation equivalent word pairs are presented to the participant to decide whether they are correct translations of each other or not, and (4) lexical decision in both of the bilingual's languages (Van Hell& De Groot, 1998). De Groot has accounted for these findings within the parameters of her model by suggesting that there are more conceptual features in common between translation-equivalent forms for concrete items than there are for abstract items.

De Groot (1992) presents an intuitively plausible argument in terms of why concrete translation equivalents would share more conceptual features relative to abstract translation equivalents. She suggests that because concrete words refer to entities whose function is highly similar across languages, the behaviors that these entities elicit, as well as the contexts within which the words referring to these entities are used, are very similar across language settings. As a consequence, the meaning representations for these words will largely be the same across languages, which quite plausibly could lead to their representations being shared. De Groot argues that abstract words, on the other hand, have no external referents and thus no guarantee that the contexts within which they are used are highly similar across language settings. And because the meanings of abstract words must, for the most part, be inferred from the context in which they are used, there is a good possibility that the meanings of these words will differ to a large extent, if not completely, across languages. For this reason, De Groot proposes that concrete translation equivalents are mediated by a common (distributed) meaning representation (i.e. a compound architecture), while abstract items are represented with language specific meaning representations (i.e. a coordinate-like architecture).

Although De Groot's hypothesis that abstract words are represented coordinately for highly proficient bilinguals is an intuitively appealing one, it is not clear within the constraints of her model that an L2 learner in the beginning stages of L2 word learning would represent the meanings of the two words distinctly. According to the way that the distributed model is specified, L2 learners associate a newly learned L2 form with an L1 meaning, resulting in a compound architecture at the earliest stages of development. De Groot (1993) acknowledges this when she says, "...beginners confronted with a new language ... may assume pure meaning equivalence of the two terms in an associated pair (whether concrete or abstract). As a consequence, they may simply link all new words onto the existing conceptual representations for the corresponding L1 words..." (p. 41). This results in a very counterintuitive prediction; namely, that beginners should not exhibit a word-type effect in translation since both concrete and abstract words are represented identically (i.e. compoundly with equal amounts of featural overlap for all word types). Yet, De Groot and Poot (1997) report word-type effects in translation for beginning level bilinguals.

The reason that De Groot is forced into this counterintuitive (and apparently incorrect) prediction is because she assumes that meaning representations, though distributed, are not dissociable. That is, although meaning representations are comprised of several distinct features, those distinct features are "meaningless" in isolation. As a consequence, the L2 learner is assumed to form new L2 form-meaning connections by mapping a newly learned L2 form onto a complete set of features that comprise the L1 lexical concept. This results in a compound architecture early on in L2 lexical development, which, over time, becomes more coordinate-like as the learner notices that the specifications of the L1 lexical concept are not completely correct in the L2 context – which, according to De Groot, is more likely to happen for abstract items than concrete items.

The sense model introduces a different type of "distributivity" that is able to avoid the counterintuitive (and apparently incorrect) prediction made by De Groot's distributed feature model. According to the sense model, lexical concepts are comprised of several distinct senses, some of which are "core" senses and some of which are more peripheral to a particular word's meaning. As a consequence, this model assumes that learners, when forming new L2 form-meaning connections, will map new L2 forms on to only those senses of the L1 word that they think would exist in the L2. There, in fact, has been support for this assumption for quite some time. Kellerman (1978) showed that Dutch learners of English presumed that only the core senses of Dutch (L1) words would exist in English (their L2). For example, Kellerman showed that beginning-level learners of English thought that only the "break a pencil" sense of "break" in Dutch would exist in English, and that more peripheral usages of the word (e.g. "take a break") would not exist in English.

Although the sense model is able to avoid the strange prediction made by the distributed feature model that there would not be word type effects for beginning-level learners, it should be noted that the sense model views word type effects in translation as a by-product of a separate variable. According to the sense model, effects such as the concreteness effect in translation must be the result of a co-variable being confounded with word type. The sense model assumes that translation is constrained by the number of senses in common between translation-equivalent lexical forms. That is, translation is possible when two lexical forms (one L1 and one L2) share a similar sense representation. When the ratio of common senses across translation equivalents is low, translation is assumed to be negatively affected. As it turns out, translation-equivalent concrete words tend to have fewer senses in general and more senses in common across languages than do abstract words (support of this claim comes from translation norming procedures, which reveal that it is easier to find single translation equivalents for concrete words than it is for abstract words – c.f. Tokowicz et al. 2002). For this reason, translation, according to the sense model, should be faster for concrete words – not because they are concrete, but because the proportion of common senses across languages for these words is greater. Because the sense model assumes that the amount of "overlap" between equivalents is determined on a word-by-word basis (as opposed to word type), it predicts that the concreteness effect in translation would disappear when the number of senses in common between translation equivalents is held constant. That is, according to the sense model, there should be no difference between abstract and concrete words in translation (or cross-language priming) when both types of words are selected so as to ensure that each sense represented in one lexical concept is also represented in the lexical concept of its translation equivalent (i.e. the ratio of common senses between translation equivalents is 1:1).[1] A recent study by Tokowicz and Kroll (in press) provides early support for this particular prediction of the sense model. In their study they show that the concreteness effect is present only when words with multiple translation equivalents are used but disappears completely when words with single translation equivalents are used. In other words, when ambiguous words like "glass" are used (glass is both an object as well as a material, a distinction that many languages make lexically), a concreteness effect is observable, but when unambiguous words are used, the concreteness effect disappears. Presumably, the reason the concreteness effect disappeared with unambiguous words is because these words had fewer senses in general and more senses in common across languages. Tokowicz and Kroll (in press) did not control for the number of senses, nor for the number of possible translations in the

---

[1] Presently, because cross-linguistic sense norms are not available, the best way to test this hypothesis would be to select translation equivalents with only one sense in each language. This way, one can assure that the two equivalents do share a common sense and that unrelated senses peculiar to one language or the other are not affecting translation performance. An interesting prediction of the sense model is that "concreteness" or, more appropriately, "sense" effects should be observable within concrete word types. One-sense concrete words should be translated differently than many-sense concrete words (assuming that several senses of the many-sense words are language specific).

multiple translation condition, so it is not clear at present what gave rise to the concreteness effect when words with multiple translation equivalents were used. There clearly is much more work that needs to be done to determine whether the (speculative) sense model account presented here of word type effects in translation will hold up when the proper manipulations are used. As things stand currently, though, the findings reported by Tokowicz and Kroll (in press), if they are found to be replicable, already demand further specification of the bilingual lexico-semantic system than is currently offered by the distributed feature or common meaning models. As these models are currently specified, they cannot account for translation performance being sensitive to the number of meanings that a particular pair of translation-equivalent words may have.

Turning now from word type effects in translation to the translation asymmetries reported by Kroll and Stewart (1994), it is clear that the distributed meaning models are not very well equipped to account for the asymmetries. The first asymmetry described by Kroll and Stewart (1994) was their finding that semantic interference only occurs in the L1-L2 direction. They did not observe a semantic interference effect in the L2-L1 direction. According to both of the distributed meaning models being discussed here, semantic interference should occur in both translation directions at all levels of L2 proficiency. As mentioned above (see section 1.2.2.1), though, attempts to replicate the translation asymmetry (semantic interference only in the L1-L2 direction) reported by Kroll and Stewart (1994) have not been successful. Quite to the contrary, there are several studies showing that translation in both directions exhibits semantic interference effects, even when participants are at low levels of L2 proficiency (c.f. Chapter 2). As such, the amount of evidence in favor of translation being semantically mediated in both directions regardless of translation direction presently outweighs the evidence to the contrary, which sits rather well with the predictions of the distributed meaning models reviewed here.

The second asymmetry reported by Kroll and Stewart (1994) is the speed of translation asymmetry. K&S (1994) found that translation latencies were much shorter when L1 was the language of production compared to when L2 was the language of production. This pattern of translation latencies has proven to be consistent across several different studies by several different researchers. Furthermore, the difference in production latencies cannot be attributed to differences in skill alone between the L1 and the L2 production systems. For example, in Chapter 3 I report a study in which participants took 30ms longer to produce L2 labels than they did L1 labels in a word naming task, but then in a translation task took 85ms longer when L2 was the language of production than when L1 was the language of production. That is, translation imposes a differential cost on L2 production that exceeds the demands of simple articulation. Kroll and Stewart's Revised Hierarchical Model can account for this asymmetry quite easily. According to the RHM, L2-L1 translation is performed via direct lexical-level connections and L1-L2 translation is performed indirectly via meaning-level representations lying in common between the two translation-equivalent forms. Models that do not propose lexical-level connections (e.g. the concept mediation model and the distributed meaning models discussed here) have to explain this asymmetry in terms of something other than types or strengths of connections. As mentioned above, Green

(1993, 1998) has proposed the interference control (IC) hypothesis to account for this asymmetry, while maintaining semantically mediated translation in both directions. Again, the essence of the IC hypothesis is that bilingual language production involves suppressing cross-language interference, and that suppressing L1 interference during L2 production is much more difficult than suppressing L2 interference during L1 production. Hence, faster L1 productions times. This account for the translation asymmetry is quite appealing and has received some empirical support from language switching studies (c.f. Meuter & Allport, 1999; Finkbeiner, Nicol & Nakamura, in prep). In a language-switching paradigm, participants are asked to name pictures (or numbers) in alternating languages. The variable of interest is the "switch cost", which is determined by comparing naming times for a particular stimulus in a language congruent condition (preceding item was named in the same language) with naming times in a language switch condition (preceding item was named in the other language). The difference in naming times for the same item between the two conditions is the switch cost. Previous work (c.f. Meuter & Allport, 1999; Finkbeiner, Nicol & Nakamura, in prep) has shown that the L1 incurs the largest cost in a switching paradigm. This finding has been interpreted to mean that, in line with Green's IC hypothesis, the L1 is suppressed during L2 production and that reactivating the L1 (necessitated by the switch back into L1) takes longer relative to reactivating the L2 because a greater amount of suppression was necessary to suppress the L1 than was necessary to suppress the L2.

Although there is some empirical support for the IC model, it may be possible to account for the speed of translation asymmetry without resorting to a control mechanism outside the lexico-semantic system. According to the sense model, there are many more senses represented within L1 lexical entries than there are represented within L2 lexical entries. Given this representational asymmetry, L1 stimuli will naturally activate more semantic-level information than L2 stimuli, making lexical selection (for the purposes of production) much more difficult when translation is in the L1-L2 direction. When translation is in the L2-L1 direction, L2 stimuli activate less information overall in the semantic system, resulting in conditions where lexical selection is relatively easier.[2]

In summary, then, the distributed meaning models reviewed here can account for the speed asymmetry in translation without having to change anything architecturally in their models. In terms of the reported semantic interference asymmetry in translation, the available evidence suggests that, contrary to Kroll and Stewart (1994), semantic interference is symmetrical and, consequently, is something that the distributed meaning models of bilingual lexical memory are already able to provide an account for. With regards to word type effects in translation, the distributed feature model (De Groot, 1992) is best suited to account for these findings; but in light of recent evidence (Tokowicz & Kroll, in press), it appears that word type effects are not as stable as originally thought, only appearing when words with multiple translations are used. The way in which the

[2] Note how this account may appear at first to be at odds with how the sense model accounts for the cross-language priming asymmetry. Namely, in lexical decision, L1 primes facilitate L2 lexical decision times, but in translation production L1 stimuli serve to create "too much" activation, interfering with the selection of the appropriate L2 lexical entry. This seemingly contradictory set of accounts is perfectly in line with how category effects have been observed to affect performance in production versus comprehension tasks. Manipulations designed to raise the level of activation with a semantic category (e.g. grouping items according to semantic category) interfere with production tasks such as picture naming, but facilitate comprehension tasks such as lexical decision or categorization.

sense model accounts for the reported word type effects was presented as well, though this account is still speculative as it has not yet been tested.

1.2.3.2 Vocabulary learning studies

As mentioned above, there have been very few studies done to test the developmental aspects of bilingual lexical processing models in general; and no studies, that I am aware of, have been done to test the specific predictions of the distributed meaning models for beginning-level bilinguals. The study done by Altarriba and Mathis (1997) mentioned above, which was designed to test the predictions of the RHM, found that even novice learners exhibited interference effects from semantically related distractors in translation tasks. This finding was taken to mean that translation was mediated by the semantic specifications of newly-learned L2 words, a finding that stood in opposition to the claims made by the RHM for how bilinguals in the earliest stages of acquisition processed L2 forms. These findings do nothing to distinguish the distributed meaning models discussed here, though, since both models would predict that translation is semantically mediated even for newly learned L2 words. In fact, the differences between De Groot's distributed feature model and the sense model in terms of L2 vocabulary learning are very subtle and difficult to test. As mentioned above, the distributed feature model assumes that L2 word learning involves mapping a newly learned L2 form onto the complete set of meaning specifications (or features) for the L1 equivalent. According to the distributed feature model, as the bilingual becomes more proficient, the meaning representations (especially for abstract words) reconfigure themselves so as to become more L2-like over time. The sense model, on the other hand, assumes that learners map their newly learned L2 form onto only the sense (or, in some cases, senses) of the L1 equivalent that they understand the new L2 form to refer to as well. This typically results in a relatively impoverished L2 form-meaning mapping consisting only of form and "core" sense(s) of the L1 equivalent.

Because both of these models are very similar in terms of how they assume new L2 form-meaning mappings are established, it is very difficult to tease the two models apart in a vocabulary learning study. Chapter 3 presents a vocabulary learning study, which reveals that L2 word learning is enhanced when new L2 forms are learned in association with uniquely colored pictures relative to black and white line drawings. The advantage conferred by learning new L2 forms in association with uniquely colored pictures is observed in both picture naming and translation tasks. This suggests that the nature of the L2 form-meaning connections established in the "unique pictures" condition are fundamentally different from those established in the control condition. I will argue that the sense model is best equipped to account for this finding, but this study admittedly does not serve to eliminate a distributed feature account of the data. The cross-language priming data provides a stronger comparison of these two distributed meaning models, revealing that the sense model is best suited to account for the cross-language priming asymmetry with masked primes.

1.2.3.3 Cross language priming studies

As we saw above, one of the more perplexing findings in the bilingual lexical processing literature is the cross-language masked priming asymmetry. Remember, the asymmetry is such that L1 primes have been found to be effective in facilitating decision times on L2 targets, but L2 primes have not been found to have the same effect on L1 targets. In the discussion above, it became clear that the common meaning models have not had much success in accounting for this asymmetry. In the present section, I will discuss how the distributed meaning models account for this finding.

In terms of a distributed feature model account of cross-language priming effects in general, De Groot , in her work with Gerard Nas (De Groot & Nas, 1991), proposed that both cognate and noncognate translation equivalents were linked across languages lexically. Her reasoning for doing this was because she found that both cognate and noncognate translation equivalents primed each other in a lexical decision task (in the L1-L2 direction) under masked conditions. She ruled out the possibility that this pattern of cross-language priming could be attributed to semantic mediation because, while she observed cross-language associative priming for cognates (e.g. author/auteur – book/boek), she did not observe cross-language associative priming for noncagnates (e.g. boy/jongen – girl/meisje).[3] As a result, De Groot and Nas (1991) argued that cognates must share a common conceptual representation, but that the conceptual representations for noncognates must be language specific (see Figure 1-6). This account is at dramatic odds with the distributed feature model De Groot proposed in 1992 to account for the word type effects she observed in translation. In the '92 model, the amount of overlap at the conceptual level between translation equivalent lexical representations is constrained by word type (i.e. concrete words share more conceptual features across languages than

[3] Note how Williams (1994), discussed above, similarly did not find cross-language priming effects between associatively related word pairs (e.g. shoe – foot) but did find cross-language semantic priming between pairs that were semantically similar (e.g. fence-hedge).

do abstract words). In the '91 model, conceptual representations fall into two classes depending on cognate status: cognates share conceptual representations across languages and conceptual representations for noncognates are represented language specifically. Regardless of the remarkable differences between the '91 and '92 models, it is not clear how either of De Groot's models would account for the finding that only L1 primes are effective in driving cross-language masked priming effects. In the '91 model, De Groot proposes symmetrical connections between languages, which clearly do not predict a directional asymmetry in priming; and in the '92 model, she proposes a variable amount of conceptual overlap (no lexical links in the '92 model) between equivalents, with concrete words sharing more conceptual features in common than abstract words. Regardless of word type, though, this latter account assumes that for any given L1-L2

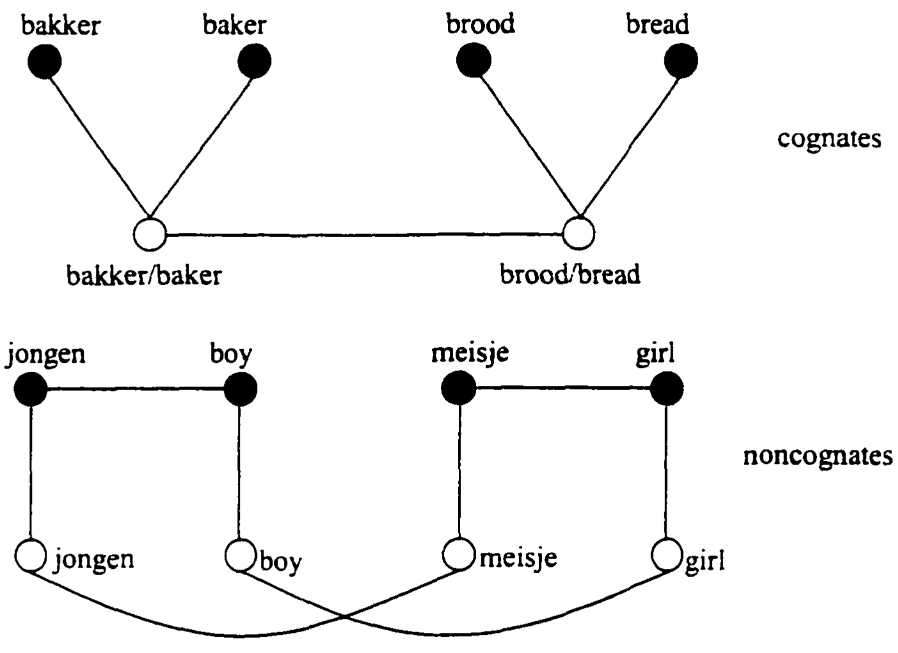

Figure 1-6: Representations and connections in bilingual memory for cognates and noncognates (adapted from De Groot & Nas, 1991)

word pair, the amount of conceptual information coming to bear on priming effects is equal between the two words and should not be sensitive to the direction of information flow. As a result, it is not clear how this latter '92 account or the prior '91 account would predict a directional asymmetry in priming.

With regards to the account provided by the sense model of this priming asymmetry, the first thing to mention is that it has been found that generating a decision in lexical decision is sensitive to the number of senses represented within a lexical entry. The nature of this "sensitivity" is that when factors known to affect lexical decision times are held constant, words with many senses are responded to faster than words with fewer senses. Given that generating a decision in a lexical decision task is sensitive to the number of senses present, it logically follows that preactivating these senses will produce priming and that the magnitude of priming will depend upon the proportion of senses that are preactivated by the prime. This is how the sense model accounts for the priming asymmetry. It is assumed that more senses are represented within L1 entries than are represented within L2 entries. Furthermore, because several of the senses represented within L2 entries are copied from their L1 equivalents, senses within L2 entries are also represented in the L1 equivalent. The opposite cannot be assumed. That is, there presumably are many senses represented within L1 entries that are not present within their L2 equivalents (see Figure 1-5b). A result of this representational asymmetry is that L1 primes will activate all (or close to all) of the senses of their L2 equivalents, resulting in priming. L2 primes, though, will only serve to activate a portion of the senses of their L1 targets, resulting in weak priming, if any priming is observed at all.

In Chapter 4. I discuss a series of experiments in which the priming asymmetry is replicated in lexical decision tasks with Japanese-English bilinguals. Interestingly, when the task was changed to semantic categorization, L2-L1 masked priming effects were observed. A complete account of this task effect will be discussed in Chapter 4.

1.3 Summary

This introduction has raised some important issues regarding bilingual lexical performance and the likelihood that a single psycholinguistic model could account for the several different findings reported in the literature. We saw how one of the first attempts to account for bilingual lexical performance, the coordinate model (Weinreich, 1953), failed when it was found that translation latencies were no different for compound and coordinate bilinguals – a finding that did not resonate well with how Weinreich had characterized the distinction between the two types of bilinguals. The word association model suffered a similar fate when Potter et al. (1984) found that the word association model could not account for the pattern of translation results that they reported (i.e. equal L2 picture naming and L1-L2 translation times). Similarly, the concept mediation model (Potter et al., 1984) was found not to be able to account for the cross-language priming asymmetries because it, as specified by Potter et al. (1984), would predict equal priming effects in both directions. An attempt by Kroll and Stewart (1994) to combine the word association and concept mediation models into their revised hierarchical model in order to account for both low- and high-level bilinguals has not been received well either. The RHM proposes that L2-L1 translation is lexically mediated in the beginning stages of bilingualism. Contrary to these predictions, several researchers have shown that translation is semantically mediated in both translation directions at even the lowest levels of proficiency (Altarriba & Mathis, 1997; De Groot & Poot, 1997; see also Chapter 2). Also problematic for the RHM is its account of the priming asymmetry, which has been called into question. The distributed feature model (De Groot, 1992; De Groot & Nas, 1991) has also been faced with several problems. The first of these concerns word type effects in translation with beginning-level bilinguals. According to De Groot (1993), L1 and L2 form-meaning connections are identical for beginning-level bilinguals, meaning that the amount of featural overlap between translation equivalents for abstract and concrete words should be the same. Hence, bilinguals in the earliest stages of bilingualism should not exhibit word type effects, yet De Groot and Poot (1997) reported exactly such effects with very low-level bilinguals. Another problematic issue concerns how the different distributed feature models that De Groot has proposed – one in 1991 and one in 1992 – would account for the translation priming asymmetry. The earlier model proposed lexical links between lexicons, but the later one did not. The earlier model constrained conceptual representations on the basis of cognate status, but the later model constrained conceptual representations on the basis of word type (e.g. concrete versus abstract). Regardless of the different assumptions made by the two models, it is not clear how either model of bilingual lexical memory could account for the masked priming asymmetry.

The sense model, proposed here as an alternative to previous models, is able to account, albeit quite speculatively at this point, for many of the phenomena that other models have found problematic. By proposing that lexical concepts are comprised of distinct sense representations, the lexical concepts of L1 and L2 translation equivalents can actually be quite different overall. This resonates well with many bilinguals who argue that there are no true translation equivalents. Also, by structuring L1 and L2 lexical concepts in this way, the sense model is able to account for asymmetrical performance across languages rather straightforwardly by proposing an asymmetrical amount of semantic information in the two languages. That is, L2 sense representations are thought to be completely (or almost completely) activated by their L1 equivalents, allowing for large L1-L2 semantic effects, in both translation and priming tasks. L1 representations, on the other hand, are thought to be only partially activated by their L2 equivalents as the result of a large number of L1 senses not found in L2 entries. The consequence of this is that semantically mediated L2-L1 effects are diminished in size, unless the task is such that only the senses in common across languages are relevant. Though the sense model is very appealing theoretically, there are still several aspects of the model that are underspecified and very speculative. The present thesis tests a few aspects of the sense model in its attempt to address the motivating question of this thesis: can a single, parsimoniously constructed psycholinguistic model of bilingual lexical representation and processing account for bilingual lexical performance at all levels of proficiency?

In Chapter 2, I present a vocabulary learning study designed to address the question of how vocabulary should be presented to learners: in semantic sets or not. This study reveals that grouping to-be-learned items according to semantic category serves to increase the activation of competitors, which results in a deleterious effect on learning. This study also confirmed that even novice learners exhibit an interference effect in translation when items are grouped according to semantic category at test. These findings will be discussed in terms of how the different models reviewed here would account for these interference effects.

In Chapter 3, I present a second vocabulary learning study. This study was designed to investigate the influence that already-established L1 information can have on L2 word learning. It was found that the faster L1 information was likely to "come to mind" the more L2 word learning suffered. A follow-up study was conducted in which half of the learning cues were designed to elicit L1 information normally and half were designed to inhibit L1 information from coming online. The results clearly reveal an advantage to learning L2 vocabulary in association with cues designed to inhibit L1 lexical information from being retrieved. Again, these results are discussed in terms of how the different models reviewed here could account for this pattern of findings.

In Chapter 4, I present a cross language masked priming study which reveals a clear task effect, such that L2-L1 masked priming is not observed in lexical decision but *is* observed when the task is changed to semantic categorization. I will argue that the sense model best accounts for this set of results.

Chapter 2 Category effects in word learning

The first study to be discussed addresses a longstanding question within the field of second language acquisition over how L2 vocabulary should be taught: in semantic groups or not? Several researchers and textbook authors work under the assumption that presenting new L2 vocabulary in semantically related sets to learners is beneficial. For example, Gairns and Redman (1986) argue in their well-known book *Working with Words* that grouping words by meaning provides "... greater precision in guiding students towards meaning, and in helping them to define the boundaries that separate lexical items" (p. 32). Seal (1991) argues similarly, saying that when words are learned in semantic sets, "the learning of one item can reinforce the learning of another..." as well as facilitate understanding because "...items that are similar in meaning can be differentiated" (p. 300). These authors, like many others, assume two benefits associated with learning words in semantic sets; the similarity between items (1) serves to facilitate the learning task, and (2) causes the learner to notice fine-grained distinctions between words, which leads to a better understanding of the words that are learned.

Looking at several of the current ESL textbooks, one notices that the practice of introducing new vocabulary in semantic sets is a popular one. For example, in one widely used textbook, *Vistas* (Brown, 1991), the 'target vocabulary' for chapter one consists of items necessary in a classroom (e.g. paper, pen, pencil, chalk, blackboard, eraser). In another popular textbook, *ExpressWays* (Molinsky & Bliss), family members (e.g. husband, wife, mother, father, brother, sister) and places in the community (e.g.

airport, bank, post office, mall, park, library, museum) serve as the target vocabulary in the first two lessons.

Despite many SLA theorists and practitioners endorsing the arguably plausible position that teaching new L2 vocabulary in semantically grouped sets is effective, there is actually very little empirical evidence to support this position. The literature often cited in support of presenting learners with words in semantic sets includes (monolingual) memory studies which involve the following tasks: first there is a study phase, in which subjects are given a series of words (all of which are well-known to the subjects) and told to memorize them; then there is a test phase, which requires subjects to either recall all the words from the study phase, or to recognize which words from a list of words had appeared in the study phase. Such studies have found that semantically grouping the study words facilitates later recall or recognition (Bousfield, 1953; Cofer, 1966; Cohen, 1963). In a closely related set of studies, it has been shown that the use of category labels during the study phase and/or during later recall of word lists has improved performance (Lewis, 1971; Tulving and Pearlstone, 1966; Tulving and Psotka, 1971).

Some SLA researchers have interpreted findings reported in the monolingual memory literature as providing support for the idea of using semantic sets in L2 vocabulary teaching. For instance, Schmitt (1997) argues that just as grouping works to enhance recall of words in list-recall tasks in native speakers, "there is no reason to believe it does not do the same for L2 learners" (p. 213). It is not entirely clear, however, how a task which requires subjects to recall already well-known words which appeared on a study list is similar to the task of learning brand new L2 labels for already-established concepts (concepts that have well established L1 labels). In a recall task, subjects could benefit from semantic grouping because it provides an organizational schema. If the subject notices that the study items all belong to a particular semantic category, at test the participant simply has to generate a list of appropriate exemplars for that category and check each one off against a record of items established during the study phase. If a generated item matches one of the subject's record of items, then she or he can "recall" it for the experimenter. This effectively turns the recall task into a recognition task, which is much easier to perform. Recall is much more difficult when participants are unable to make use of an organizational schema (Bower, Karlin & Dueck, 1975).

In contrast, the demands of an L2 vocabulary-learning task are very different from those of remembering a list of already-known vocabulary items. The learning of new L2 vocabulary involves establishing new L2 forms in memory and then forming associations between those forms and already-established information. It is unclear how being aware of what semantic category the new L2 form's referent belonged to could affect the establishment of these associations – unless, of course, one assumes that newly established L2 forms are associated directly with meaning representations. As the review in Chapter 1 revealed, the question of whether newly established L2 forms are associated with already-established L1 equivalent forms or meaning representations is still being debated. The present experiment was designed to address this question directly by manipulating the semantic properties of to-be-learned items during learning. If newly established L2 forms are associated with L1 forms, as the word association and revised hierarchical models suggest, then a manipulation of semantic properties during learning should have little to no effect. If, though, newly established L2 forms are associated with meaning representations, then a manipulation of semantic properties during learning should have a dramatic effect on learning performance. The purpose of this experiment was to determine what effect (if any) manipulating semantic level properties of to-be-learned items during vocabulary learning would have on subsequent lexical processing tasks. The results of this study will serve to answer two important questions: (one) is L2 vocabulary learning enhanced or interfered with when target items are grouped according to semantic categories? and (two) are the associations between newly established L2 forms and already-established information affected by a manipulation of semantic level properties? (If so, then we can assume direct L2 form-meaning connections for newly learned L2 forms.)

The question of whether L2 vocabulary learning is affected positively or not when to-be-learned items are grouped into semantic sets has been addressed before in learning to criterion studies. In these type of studies, it has been shown that participants take longer to learn new labels for sets of semantically related items than for sets of semantically dissimilar items (Higa, 1963; Kintsch & Kintsch, 1969; Nation, 2000; Tinkham, 1993, 1997; Underwood, Ekstrand & Keppel, 1965; Waring, 1997). Note, however, that although learning semantically related words appears to take *longer*, it is possible that words learned under these conditions are learned *better* for the purpose of actual language use (e.g. the retrieval of vocabulary during production and comprehension). That is, the very difficulty associated with learning the new labels may make them easier to process once they are learned (c.f. Craik & Lockhart. 1972; Craik & Tulving, 1975 for a discussion of the effect of "depth of processing" on retrieval). The present study addresses this possibility directly by using reaction times in a language production task as its dependent variable.

The second question, of whether new L2 forms are associated directly with equivalent L1 forms or meaning representations has also been addressed before. Altarriba and Mathis (1997; see introduction) found that novice learners exhibited semantic interference effects in a translation recognition task as well as in a Stroop task, suggesting that even beginning-level L2 learners establish direct L2 form-meaning connections. The present study addresses this same question but with a different manipulation and a different task. In the present study, semantic grouping was manipulated both during training and during test. The test phase included L1 to L2 translation and L2 to L1 translation.

2.1.1 Method

2.1.1.1 Participants

Forty-seven undergraduates (29 female, 18 male) at the University of Arizona participated for course credit. All of the participants were monolingual English speakers.

2.1.1.2 Materials

Thirty-two novel words were created and each was paired with a picture of a familiar concept (see Appendix A for a complete list). The new L2 words were created to conform to English phonotactic constraints in order to reduce memory load (Ellis & Beaton, 1993; Gathercole & Baddeley, 1989, 1990; Service, 1992). For variety, half of the words for each category were one syllable in length (e.g., *birk*, *plap*, *floop*), while the other half were two syllables in length (e.g., *walloon*, *dopal*, *fonteen*). The pictures used were adopted from Snodgrass and Vanderwart (1980). The eight most prototypical category exemplars available from the Snodgrass and Vanderwart (1980) set of drawings were selected from each of four categories: animals, kitchen utensils, furniture, body parts.

2.1.1.3 Procedure

Individuals participated on two separate days within a 5-day period. Each session spanned approximately 45 minutes and consisted of: (1) vocabulary training (participants were told that they were learning a new "alien" language), followed by (2) a recognition task, and then (3) two blocks of translation in each direction (L1 to L2 and L2 to L1) for a total of four blocks of translation. The recognition task and translation tasks at the end of the first session were included because pilot experiments revealed that participants benefited from being "tested" part way through training. Pilot testing also indicated that performance on these first session tests was generally poor, however, so only translation times from the second session were recorded for analysis.[4] Participants were seated individually in sound-resistant computer booths during the entirety of each session. At the beginning of the first session, participants were shown the pictures used in the

---

[4] Because of this design, learning performance was not assessed until the language production test. For this reason, all participants were tested, regardless of whether they had sufficiently learned the new labels or not.

experiment on flashcards and asked to name them in English. After they had named each of the pictures correctly, they were asked to write the names of the items down before beginning their training on the computer. This was done to ensure name agreement on all of the items because, although L1 labels were never present during training, the translation task did use L1 labels. During training, participants first heard a recording of the L2 word over headphones, then saw the L2 word and its corresponding picture for 500 ms. on the monitor, then heard a second recording of the L2 word. Participants were asked to repeat the L2 word twice into the microphone placed in front of them for recording purposes. The purpose of the repetitions was simply to facilitate learning. Post-hoc assessment of the recordings revealed that all of our participants repeated the L2 words at least twice during training.

Items were displayed on computer using the DMDX system developed by J.C. Forster at the University of Arizona. In the Related training condition, semantically related items were blocked into groups of eight. Each block of eight was presented four times during training, in pseudorandom order such that no block appeared twice in a row. In the Unrelated training condition, the 32 items were scrambled within a block; each block was presented four times (with the order of items in different random order each time). The vocabulary training was followed by a recognition task, which consisted of the presentation of a picture followed by one of the L2 labels. Of the 64 picture-label pairs, half were correct (the picture was paired with its new label) and half were incorrect (the picture was paired with the wrong label). Participants were instructed to press a "yes" button if the picture and L2 word matched and a "no" button if they did not.

Participants were given feedback for each item, including whether they were correct or not, as well as their reaction times. After the recognition task, participants were given the translation task. In this task, participants were shown a row of hash marks ("#######") for one second (for orienting purposes only) followed by the word to be translated. For example, in the L1 to L2 blocks, an English word appeared and participants were asked to speak the "L2" translation equivalent into the microphone as quickly as possible. Their vocal response triggered a voice key, stopping the computer's timer. Latencies were measured from the time the word to be translated was presented until the voice key was triggered. All responses were recorded on tape so that they could be checked for errors. As per standard procedure in timed language production studies (Damian, Vigliocco, & Levelt, 2001; Levelt, Roelofs, & Meyer, 1999; Schriefers, Meyer, & Levelt, 1990), all incorrect responses, as well as fluency errors like stutterings and "um's" were counted as errors because they would lead to an inaccurate measure of timing (triggering the voice key before the onset of the word was uttered). In order to control for any differences that may arise due to order of translation direction, translation direction was counterbalanced such that half of the participants in each design cell performed forward translation (L1 to L2) first, while the other half performed backward translation (L2 to L1) first.

2.1.2 Results

As is typical in learning experiments of this type (in which there is no concrete incentive for participants to perform well), many participants failed to reach the preset learning criterion (see also Altarriba & Mathis (1997) who also had a high number of

		Training		
		Related	Unrelated	*mean*
Testing	Related	1272 (181)	1148 (113)	*1210*
	Unrelated	1323 (213)	1062 (155)	*1193*
	mean	*1298*	*1105*	

Table 2-1: Translation means (in msec) and standard deviations by training and testing condition

subjects fail to meet the accuracy criterion). Data from 23 participants were excluded from analysis for not reaching the predetermined accuracy criterion of 80%.[5] Mean reaction times (RTs) for correct responses per experimental condition and participant were calculated for the remaining 24 participants (14 female, 10 male). The mean translation times for the four treatment conditions appear in Table 2-1. The average error rate for the experimental conditions was 9%, and did not differ across conditions.

Separate analyses of variance (ANOVA) with participants (F_1) and items (F_2) as random variables reveal significant main effects of translation direction ($F_1(1, 20) = 41.81, p < .0001$; $F_2(1, 124) = 16.87, p < .0001$) and training condition ($F_1(1, 20) = 11.55, p = .003$; $F_2(1, 124) = 126.56, p < .0001$), but no effect of testing condition ($F_1(1, 20) < 1$; $F_2(1, 124) < 1$). The interaction between translation direction and training condition was significant in analyses by items ($F_2(1, 124) = 8.01, p = .005$) and approached significance by participants ($F_1(1, 20) = 3.00, p = 0.09$). There were no other significant interactions.

[5] Ishould point out, however, that the overall pattern of reaction times is similar when all participants are included and when just those who met the accuracy criterion are included.

		Training		
		Related	Unrelated	*mean*
Testing	Related	1164 (127)	1100 (98)	*1132*
	Unrelated	1230 (182)	1002 (125)	*1116*
	mean	*1197*	*1051*	

Table 2-3: Backward translation (L2>L1) means (in msec) and standard deviations by training and testing condition

Now let us consider the results for each translation direction separately. The results for L1 to L2 translation can be seen in Table 2-2, and the results for L2 to L1 translation can be seen in Table 2-3. Translation in the forward direction was 238 ms longer for those participants who learned semantically grouped L2 words compared to those who learned a random grouping of the same L2 words. Planned comparisons showed that this difference was highly significant ($F_1(1, 20) = 12.60, p = 0.002$; $F_2(1, 63) = 71.60, p < .0001$). Translation in the backward direction was 146 ms slower for participants who learned semantically related words compared to participants who learned the same words in unrelated sets. Planned comparisons showed that this difference was significant in both the participants analysis ($F_1(1, 20) = 7.46, p = .013$) and the items analysis ($F_2(1, 63) = 57.78, p < .0001$). It is quite clear from these results that there was a semantic category effect in training, and that this effect was negative.

		Training		
		Related	Unrelated	*mean*
Testing	Related	1380 (180)	1196 (112)	*1288*
	Unrelated	1415 (215)	1121 (164)	*1268*
	mean	*1397*	*1159*	

Table 2-2: Forward translation (L1>L2) means (in msec) and standard deviations by training and testing condition

Participants translated L2 labels learned in semantic sets significantly more slowly than they did L2 labels learned in random order. This was the case in both translation directions.

Mean translation latencies as a function of training condition (related vs. unrelated items) and translation direction are shown in Figure 2-2. As can be seen, when translation latencies are grouped according to training condition, there is a clear difference between participants who learned L2 words in semantic groups and those who learned the same L2 words in random order.

Figure 2-4 shows mean translation latencies as a function of testing condition (related vs. unrelated items) and translation direction, revealing that participants who translated items that were grouped into semantic sets were not reliably slower than participants who translated random sets of those same items.

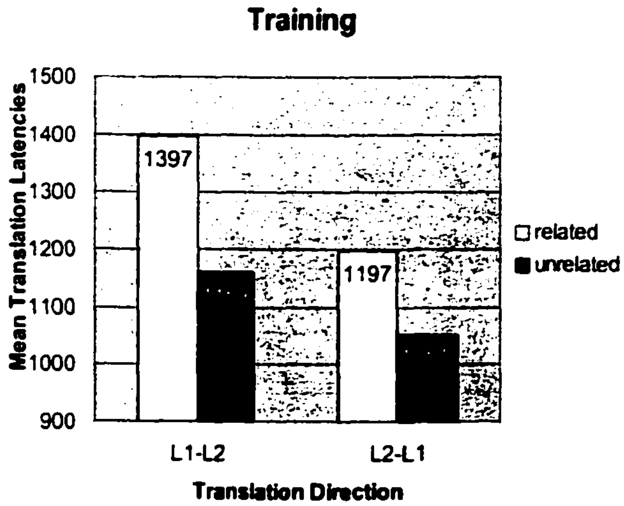

Figure 2-2: Response Latencies by training condition (related vs. unrelated) and translation direction

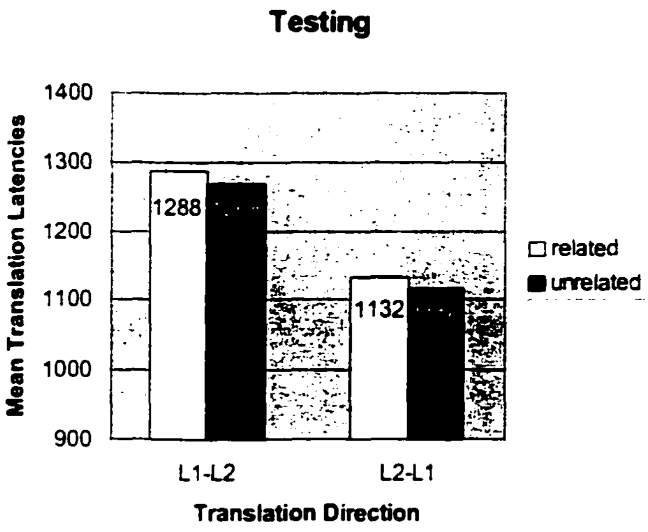

Figure 2-4: Response Latencies by testing condition (related vs. unrelated) and translation direction

2.1.3 Discussion

The purpose of this study was twofold. On the one hand, I wanted to test directly the question of whether grouping to-be-learned items in semantic sets facilitated vocabulary learning or not. On the other hand, I wanted to test the assumptions of various different models of bilingual lexical memory regarding the nature of L2 form-meaning associations early on in development. This is an extremely important issue in modeling the bilingual lexicon, yet, as the review in Chapter one revealed, there has been only one study (that I am aware of) to use a vocabulary learning experiment to test the developmental aspects of these different models.

The answers to the research questions are clear. Translation times were significantly slower for words learned in semantic sets relative to those learned in random order. Because translation requires both comprehension and production under speeded conditions, it is considered a good measure of an individual's ability to make fluent use of encoded lexical information. As such, the results reveal very clearly that learning performance suffers when to-be-learned items are grouped in semantic sets. Given this state of affairs, it is clear that the establishment of L2 forms in memory is affected by manipulations of semantic properties. Because it is unlikely that grouping to-be-learned items in semantic sets would have an effect on the state of L1 form representations, it is unlikely that learners were associating new L2 forms with their equivalent L1 forms when learning the new words. Hence, the straightforward conclusion is that subjects established direct associations between new L2 forms and meaning-level representations.

Between the findings reported here and those reported by Altarriba and Mathis (1997) and De Groot and Poot (1997), it is becoming quite clear that direct L2 form-meaning connections are established in even the earliest stages of L2 lexical development. Together, these findings call into serious question models of bilingual lexical memory that exclude semantic level connections between the two lexicons in favor of lexical level connections at any stage of development (i.e. the word association and revised hierarchical models).

Though these findings are rather convincing in showing that learners develop direct L2 form-meaning connections in the initial stages of word learning, it is still incumbent on those models that assume direct form-meaning connections to explain these particular findings. How could grouping to-be-learned items in semantic sets have this particular effect on L2 word learning? I suggest that simultaneous activation of semantically related lexical items is at the root of the effect. Several different theories of lexical representation (e.g. de Groot, 1992; Levelt, Roelofs, & Meyer, 1999; Schreuder & Flores D'Arcais, 1989) assume that when a concept is activated, there is spreading activation among representations that are similar in meaning. This "spreading activation" is responsible for the facilitory effects in categorization tasks (Glaser & Düengelhoff, 1984) and semantic priming tasks (Meyer & Schvanveldt, 1971; Neely, 1991); but in production tasks, activation of related information at the semantic level leads to inhibitory effects (Levelt et al., 1999).

One task that has been used extensively to explore interference effects in language production is the *picture-word interference task*. In this task, participants are asked to name a picture that appears more or less coincident with a written or spoken "distractor" word. A semantic relationship between a picture and a distractor word slows picture naming relative to an unrelated condition. Interference in this task is argued to arise as a result of co-activated lexical entries that are semantically related. The to-be-named picture of an object activates a cohort of related, mutually activating lexical concepts, which in turn activate their corresponding lexical representations, or "lemmas". Co-activated lemmas compete with each other and affect the speed with which a lexical form is retrieved and output (Damian, Vigliocco, & Levelt, 2001). When the distractor words are unrelated semantically, activation of competitors is not thought to be affected, but when a semantically related distractor word is presented, activation of competitors is increased to an even greater extent. The result of this increased activation amongst competitors is that the target lemma has a lower relative level of activation overall. With a lower relative level of activation, lemma selection is more difficult, which leads to longer production times in the related condition relative to the unrelated condition (Levelt, et al., 1999).

How do such results bear on the present set of findings? During the learning phase of the present study, participants saw a picture paired with L2 orthographic and phonological forms. Presumably, viewing the picture served to activate its corresponding concept, which, in turn, activated related concepts. In both training conditions, subjects established the new L2 form in memory and associated that directly with the appropriate meaning representation for the picture. In the Unrelated training condition, the appropriate meaning representation presumably had a suitably high relative level of activation because interference from semantically related representations, though present, is minimal. The resulting strength of the L2 form meaning connection established in this condition could be considered the "standard" or baseline strength. In the Related condition, however, interference from related meaning representations is heightened due to repetitive and residual activation of concepts within the same category. This results in a much lower relative level of activation for the target meaning representation. And, just as it is more difficult for monolinguals to select the appropriate lexical item in a picture naming task when target items have a lower relative level of activation overall, the L2 vocabulary learner finds it much more difficult to map newly established L2 forms onto meaning representations with lower relative levels of activation. Furthermore, not only is it more difficult to establish a form-meaning connection in the Related condition, but assuming that the quality of the form-meaning connection is affected by the degree to which a meaning representation is activated relative to its close competitors, then it logically follows that the strength of the form-meaning connections established in the Related condition would be weaker relative to those established in the Unrelated condition.

Though this study was not designed to test this question directly, a post-hoc assessment of translation interference effects when subjects are grouped according to training condition provides some support for this account. If, as argued above, subjects were able to establish "stronger" form-meaning connections in the Unrelated condition, then those subjects should show a stronger semantic interference effect in translation relative to the subjects in the Related condition. The data appear to be patterning so as to support this speculation (see Figure 2-5). There are what appear to be interference effects

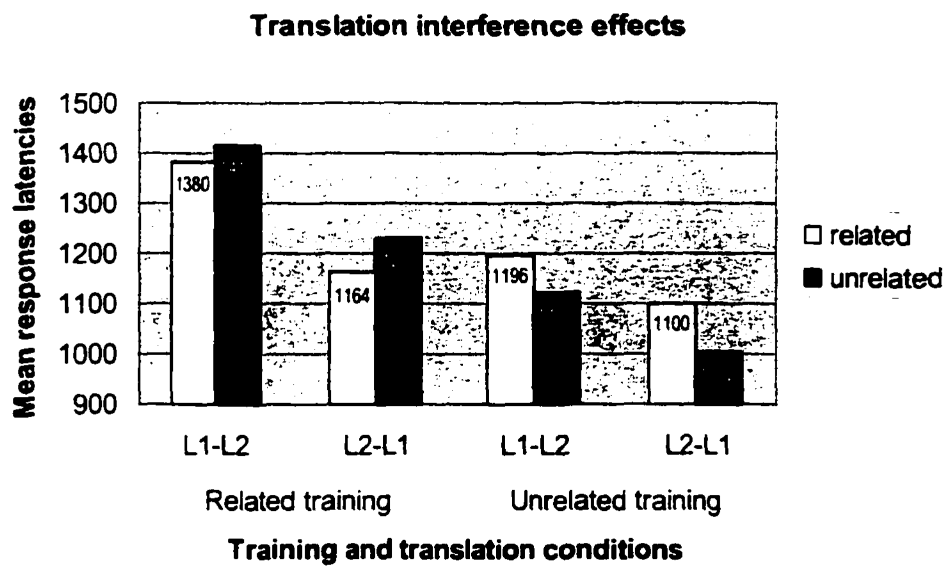

Figure 2-5: Translation interference effects (translation latencies of related items versus unrelated items) as a function of training condition.

in translation for items learned in the Unrelated condition, but not for items learned in the Related condition.[6]

Although I can only speculate at this point as to how the grouping manipulation had the effect that it did on learning, the results could not be any clearer with regards to the original research questions: (one) is L2 vocabulary learning enhanced or interfered with when target items are grouped according to semantic categories? and (two) are newly established L2 forms associated with equivalent L1 forms or with semantic representations directly? L2 vocabulary learning is clearly interfered with when to-be-learned items are grouped in semantic sets. Given this finding, it logically follows that learners establish direct L2 form-meaning connections during even the earliest stages of word learning.

The implications that this study has for vocabulary instruction and curriculum development are not trivial. As pointed out in the introduction, several authors in the teaching methodology literature have argued that vocabulary should be taught in semantic groups. The results of the present study converge with those of Tinkham (1993, 1997) and Waring (1997) to suggest that teaching words in semantic sets creates competition between items, which increases difficulty during learning as well as during memory retrieval in language production.

---

[6] Because this is a post-hoc assessment, it will be necessary to conduct more research that is designed to test this question directly before one could be confident that these are in fact interference effects that arise during testing and not "congruency" effects carrying over from the training condition.

Chapter 3 Minimizing L1 in L2 word learning

In Chapter 2 we saw that L2 word learning suffered when to-be-learned items were grouped into semantic sets. The source of this interference effect was attributed to increased activation among competitors, similar to the account given for interference effects observed with monolinguals in picture-word interference tasks as well as picture naming experiments where items are grouped semantically. The motivation for the present set of experiments was to see if the opposite effect could be elicited. That is, seeing as how a manipulation designed to increase activation among competing L1 lexical representations led to poorer L2 vocabulary learning, would a manipulation designed to inhibit activation of L1 lexical representations lead to enhanced L2 vocabulary learning?

Although the word learning study reported in Chapter 2 suggests that increased activation among L1 competitors during L2 word learning leads to poorer learning performance, it also seems reasonable to assume that it would be very difficult to establish L2 forms in memory unless an association between the new L2 form and already-established L1 information could be formed. That is, despite evidence that L1 information can interfere with L2 word learning (at least in "extreme" experimental conditions), it may, nevertheless, be the case that an L1 "influence" is facilitory, if not necessary, in the learning process. The first experiment in this study was designed to test this possibility by having participants learn novel L2 labels for concepts with single-word L1 equivalents (e.g. female sibling) and without L1 equivalents (e.g. female elephant). If associations between the newly learned L2 form and already-established information in the L1 facilitate learning, then learning a new label for "female sibling" should be easier; if, though, those associations are inhibitory, then learning a new L2 label for "female elephant" should be easier.

3.1 Experiment 1

Experiment 1 made use of a modified naming task to test how well participants had learned 28 new L2 labels. Participants learned these new L2 labels in association with two-word phrases having lexicalized and non-lexicalized referents. The purpose of this experiment was to investigate the role that associations with already-established L1 information have on L2 vocabulary learning: facilitory or inhibitory?

3.1.1.1 Method

3.1.1.1.1 Participants

Fifty-four undergraduates at the University of Arizona participated for course credit. All of the participants were native English speakers; all but 3 had had some second language instruction in high school.

3.1.1.1.2 Materials

An initial list of 40 two- or three-word phrases were generated and tested in a pilot experiment. This initial list was such that half of the phrases' referents were lexicalized in English (e.g. "young dog") and half were not (e.g. "young giraffe"). The phrases were constructed in pairs so that phrases with lexicalized referents (e.g. "melted ice") and phrases with non-lexicalized referents ("melted aluminum") were not different by more than one word for any given pair. The phrases were randomized and printed on a sheet of paper. Twenty participants were asked to write one-word equivalents for each phrase that they felt had an English equivalent and put an "X" next to the phrases without equivalents. Because we were interested in the influence that an association between an L2 form and already-established information has on L2 vocabulary learning, we only selected those phrases for which participants in the pilot experiment consistently produced the correct responses (i.e. equivalents when appropriate and an "X" when appropriate). If participants provided an equivalent for a phrase when an "X" was the appropriate answer (e.g. "monkey" for "young ape"), that item was scored as being incorrect. Conversely, if participants indicated an "X" for items with lexicalized referents (e.g. "X" for "group of birds"), that item was scored as being incorrect also. Our criterion was 90% accuracy. This procedure resulted in 28 phrases, 14 with lexicalized referents and 14 without.

Twenty-eight L2 labels were created, one for each phrase. These were designed to conform to English phonotactic constraints in order to reduce memory load (Ellis & Beaton, 1993; Gathercole & Baddeley, 1989, 1990; Service, 1992). For variety, half of the words in each learning condition (N=7) were one syllable in length (e.g., birk, plap, floop), while the other half were two syllables in length (e.g., walloon, dopal, fonteen).

3.1.1.1.3 Design and Procedure

The design of the experiment was completely within subjects. Each subject learned 28 new L2 labels, 14 in association with phrases having lexicalized referents and

14 in association with phrases without lexicalized referents. Two lists were constructed in order to counterbalance the pairing between L2 label and phrase. For example, on List A "glip" was paired with "dried grape" and "vid" was paired with "dried apple." On List B this pairing was reversed such that "vid" was paired with "dried grape" and "glip" was paired with "dried apple." The purpose of this counterbalancing was to ensure that any observed difference in performance between item types (lexicalized referents or not) was due to the experimental manipulation and not due to differences between the L2 labels.

The experiment consisted of two separate training sessions. These occurred within a 5-day period and took participants approximately 45 minutes each to complete. Each session consisted of (1) vocabulary training, followed by (2) a recognition task, and then (3) two blocks of L2 word naming. Generally speaking, performance is quite poor during the first session, so only naming times from the second session were used in the analysis. During each session, participants were seated individually in sound-resistant computer booths in front of a computer monitor. During training, participants first heard a recording of the L2 word over headphones, then saw the phrase (e.g. "female sibling") for 500 ms. on the monitor, followed immediately by a simultaneous auditory and visual presentation of the L2 word. Participants were asked to repeat the L2 word twice into the microphone placed in front of them for recording purposes. The purpose of the repetitions was twofold: one, to facilitate learning, and two, to provide participants with the opportunity to practice producing the target L2 labels. Because the test is a naming task, it is important that participants feel comfortable producing the novel targets.

Items were displayed on the computer using the DMDX system developed by J.C. Forster at the University of Arizona. During training, all 28 items were blocked and presented randomly within blocks four separate times. The vocabulary training was followed by a recognition task, which consisted of the presentation of a phrase followed by one of the L2 labels. In this task, half (N=28) of the phrase-L2 label pairs were correct (the phrase was paired with its new label) and half were incorrect (the phrase was paired with the wrong label). Participants were instructed to press a "yes" button if the phrase and L2 word matched and a "no" button if they did not. Participants were given feedback for each item, including whether they were correct or not, as well as their reaction times. After the recognition task, participants were given the naming task. In this task, participants were shown a row of hash marks ("#######") for one second (for orienting purposes only) followed by the phrase for which they were to produce the correct L2 label. All 28 items were blocked and presented randomly within blocks twice. Participants' vocal responses triggered a voice key, stopping the computer's timer. Latencies were measured from the time the phrase to be named was presented until the voice key was triggered. All responses were recorded so that they could be checked for errors. As per standard procedure in timed language production studies (Damian, Vigliocco, & Levelt, 2001; Schriefers, Meyer, & Levelt, 1990; Levelt, Roelofs, & Meyer, 1999), we counted as errors all incorrect responses, as well as fluency errors like stutterings and "um's" because they would lead to an inaccurate measure of timing.

3.1.1.2 Results

	Training	
	+ Referent	- Referent
Naming Times	1351 (17%)	1299 (13%)

Table 3-1 Mean L2 naming times (in msec) and error rates by training condition.

As is typical in learning experiments of this type, many of our participants failed to reach the pre-deterimined accuracy criterion of 80% (c.f. Altarriba & Mathis, 1997 and Chapter 2 for similar findings). Data from 25 participants were excluded from analysis for not reaching the predetermined accuracy criterion of 80%. Mean reaction times (RTs) for correct responses per experimental condition and participant were calculated for the remaining 29 participants. The mean naming times for the two training conditions are presented in Table 3-1. Separate analyses of variance (ANOVA) with participants (F1) and items (F2) as random variables reveal a significant main effect of training condition in the participants analysis ($F1(1, 28) = 4.58$, $p = 0.04$), but not in the items analysis ($F2 < 1$). The mean error rate was 15%; 17% in the +referent condition and 13% in the -referent condition. This difference in error rates was not significant ($F(1,56) = 2.03$, $p=0.16$).

3.1.1.3 Discussion

The question motivating this study was whether associations between newly established L2 forms and already-established L1 information facilitated or inhibited L2 vocabulary learning. At first glance, it appears that associating new L2 labels with +referent phrases is more difficult than associating new L2 labels with –referent phrases. But the observed 52ms interference effect did not reach significance in the items analysis. For a difference of 52ms to reach significance in the subjects analysis and not in the items analysis suggests a large degree of variance between the items; and, more specifically, that only some of the items may be responsible for the pattern of results. In fact, post-hoc assessment of the mean naming times for each pair of items revealed that the L2 labels in the –referent condition (e.g. the L2 label for "female elephant") were named faster than their +referent counterpart (e.g. the L2 label for "female sibling") for one half of the item pairs; for the other half of item pairs, this pattern of naming times was reversed. Clearly, then, the experimental manipulation of training condition was not very robust. When learning novel L2 labels, learners appear to be sensitive to something other than whether a new L2 label is learned in association with phrases having lexicalized referents or not. A post-hoc assessment of the data suggested that the frequency of the phrases' referents themselves might have led to the data patterning the way that it did. The slowest L2 naming times were for those phrases whose referents had the highest frequencies (according to Kucera & Francis, 1967). That is, it appeared that how quickly an L1 referent "came to mind" influenced L2 vocabulary learning more than the experimental distinction between +referent and –referent phrases; and furthermore, it appeared that the faster an L1 referent came to mind, the slower learners were to retrieve its translation-equivalent L2 label at test.

In order to confirm this possibility, we conducted a follow-up experiment with 13 participants separate from those in the vocabulary learning experiment. In the follow-up experiment, participants were presented with the same phrases that the learners had been presented with in the vocabulary learning experiment. If the phrase had a referent, participants were asked to produce that referent as quickly as possible; if they could not think of a referent for the phrase, they were asked to say "no word." Response times

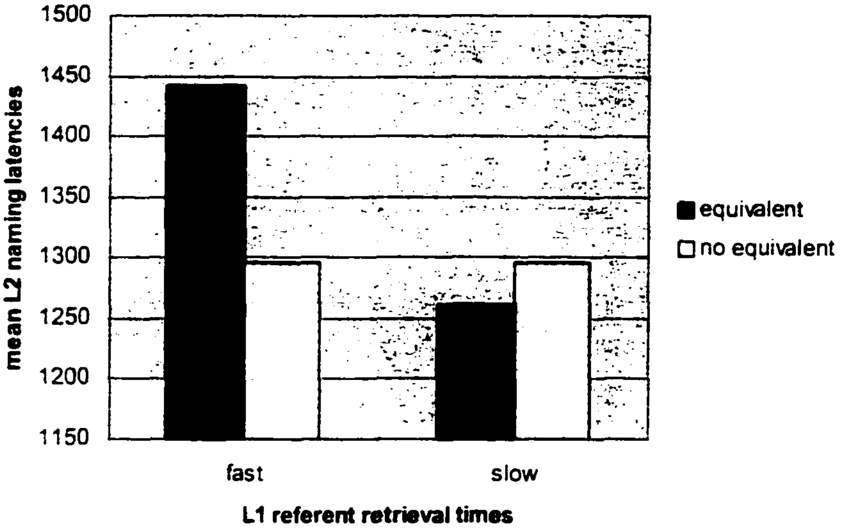

Figure 3-1: Mean L2 naming times for items grouped according to mean retrieval times for the L1 referents of the phrases. The baseline is provided by the mean L2 naming times for –referent phrases.

were recorded and analyzed in the same way described for the vocabulary learning experiment. Using a median split, the mean response times for the +referent items (N=14) were divided into two halves. Response times from the vocabulary learning experiment were then grouped according to this split and re-analyzed. In this re-analysis, mean L2 naming times for phrases with "fast" and "slow" L1 referent retrieval times were compared to the baseline provided by mean L2 naming times for minus-referent phrases. The results are presented in Figure 3-1. It is quite clear that participants are much slower to produce L2 labels for phrases with quickly-retrieved L1 referents than they are for phrases with slowly-retrieved L1 referents. An ANOVA revealed that this difference was significant ($F1(1, 28) = 19.38$, $p < 0.001$; $F2(1,12) = 5.05$, $p = 0.04$). Relative to the baseline, L2 labels for phrases with quickly-retrieved L1 referents are produced significantly slower ($F1(1, 28) = 18.59$, $p < 0.001$; $F2(1,40) = 5.41$, $p = 0.025$). Conversely, L2 labels for phrases with slowly-retrieved L1 referents are produced slightly faster relative to the baseline. but this difference is not significant (all $Fs < 1$).

Overall, it appears that at the earliest stages of L2 vocabulary learning L2 lexical processing is very sensitive to how quickly an L1 translation-equivalent "comes to mind" and that the nature of this effect is one of interference, not facilitation. When L1 equivalents come to mind quickly, L2 performance suffers greatly. When L1 translation equivalents are slow to come to mind, L2 lexical processing approximates that of when there is little chance of L1 interference (as is presumably the case in the -referent condition).

The possibility that L2 vocabulary learning can be enhanced by slowing down the time course of activation of L1 equivalents during learning is intriguing. Namely, if activation of L1 equivalents during L2 vocabulary learning causes interference, and if that interference effect can be reduced by using training stimuli designed to delay retrieval of L1 equivalents, then this could improve L2 vocabulary learning.

Experiment 2 was designed to test this possibility. Experiment 2a was designed to see what type of training stimuli (pictures) would best serve to slow participants' retrieval of their corresponding L1 labels. Once this was determined, Experiment 2b used these pictures (as well as control pictures) as stimuli in an L2 vocabulary learning task. At test, participants were asked to perform three different tasks: word naming, picture naming, and translation. Based on the results of Experiment 1, we predicted that L2 labels learned in association with the pictures in the experimental condition would be learned better (by virtue of slower L1 retrieval times) compared to L2 labels learned in association with the control pictures.

3.2 Experiment 2a – Picture-Word Matching

Experiment 2a was done in an effort to identify the best experimental items to use as stimuli in the subsequent vocabulary learning experiment. Participants performed a picture-word matching task with two different sets of stimuli: rotated black and white drawings, and uniquely colored pictures. The decision to test these two manipulations in particular was motivated by two previous studies. In one, Kroll and Tokowicz (2002) found that rotated pictures conferred an advantage in a vocabulary learning experiment when used as training stimuli. In another, Johnson (1995) found that uniquely colored pictures were named slower than their black and white counterparts. In the present study, we compared naming times between these two manipulations. The set of items that led to the slowest matching times relative to control pictures (standard black and white line drawings in upright position) were chosen to be used as training stimuli in Experiment 2b.

3.2.1.1 Method

3.2.1.1.1 Participants

Forty-three undergraduates at the University of Arizona participated for course credit. All of the participants were native speakers of English.

3.2.1.1.2 Materials

Ninety different pictures were adopted from the Snodgrass and Vanderwart (1980) set of items. Forty-five items with an inherent canonical orientation were selected and used as stimuli in the "rotated pictures" condition. For example, pictures of a chair and a giraffe were selected because they have an inherent canonical orientation. Conversely, a picture of a lemon could not be used because it does not have a canonical orientation and looks familiar at any degree of rotation. Five of the selected items were designated as practice items, with the remaining forty items used as experimental items. In the experimental condition, each item consisted of a rotated picture (at least 90 degrees) paired with a label that always appeared in the center of the screen, directly below the picture. Half of the pictures were paired with their correct labels (e.g. a picture of an airplane appeared with its correct label, "airplane") and half were not. In the control condition, each item consisted of a canonically orientated picture paired with a label that also always appeared directly below the picture.

The remaining 45 items were used in the "unique colors" condition. Five of the items were practice items. In the experimental condition, pictures were uniquely colored and paired with either a correct label (half of the time) or an incorrect label. In the control condition, the same pictures were used, but the pictures were not colored.

3.2.1.1.3 Design

The design of the experiment was entirely within-subjects. For each participant, half of the items were experimental and half were control. Two counterbalanced lists were constructed such that all items appeared equally across the experiment in both experimental and control conditions. For example, the items that appeared in a canonical orientation on List A appeared in a noncanonical orientation on List B, and the items that appeared in a noncanonical orientation on List A appeared in a canonical orientation on List B. Participants were randomly assigned to each list.

3.2.1.1.4 Procedure

The procedure was straightforward. Participants were instructed to press a "yes" button if the picture and word matched (e.g. when a picture of a chair appeared with the word "chair"), and a "no" button if they did not. Both the rotated pictures and the colored pictures were blocked and presented separately from each other (the order that the blocks appeared in was counterbalanced across subjects). The first 5 items of each block were practice items and reaction times were not recorded for them.

3.2.1.2 Results

All incorrect responses, as well as response times longer than 3000 ms or shorter than 200 ms, were counted as errors and excluded from analysis (5%). As seen in Table 3-2, colored pictures were matched with their labels 54 ms slower than the control pictures were. This difference was found to be significant in both participant (F1) and item (F2) analyses: ($F1(1, 42) = 23.99$, $p < 0.0001$; $F2(1, 19) = 5.89$, $p = 0.025$). Rotated pictures were matched with their labels 76 ms slower than the canonically oriented control pictures. Although this difference is larger numerically, it did not reach significance in the items analysis: ($F1(1, 42) = 21.99$, $p < 0.001$; $F2(1, 19) = 2.84$, $p > 0.05$). Statistically speaking, then, the difference between experimental and control items

Type of Test	Testing Condition: control condition	experimental condition
colored picture test	760.38	814.04
rotated picture test	847.36	923.71

Table 3-2: Mean Picture-Label Matching Times for Experiment 1

in the color test was reliable while the difference observed in the rotated-pictures test was not.

3.2.1.3 Discussion

The purpose of Experiment 2a was to find for which type of stimulus (colored pictures vs. rotated pictures) participants were reliably slower to match a picture with its label. In order to perform the picture-label matching task, one must first recognize the picture. Upon recognition, the lexical concept corresponding to the picture is activated, which serves to activate its lemma. Presumably, the presentation of the label serves to activate its lemma directly (Levelt et al., 1999; Roelofs, 2000). That said, the observed latency was likely due either to a delay in picture recognition or concept retrieval (not lemma activation as the label should be equally efficacious in activating the appropriate lemma in both conditions). In any event, it is clear that lexical concepts are activated ("come to mind") more slowly upon viewing an oddly colored picture (with the present experimental set, at least) than upon viewing a black and white line drawing. Based upon the results of the present experiment (in conjunction with those of Experiment 1), we predict that the colored pictures will delay activation of L1 equivalents, which should serve to limit the amount of L1 interference , leading to better L2 vocabulary learning. We tested this hypothesis in Experiment 2b.

3.3 Experiment 2b – Vocabulary Learning with Pictures

Just like Experiment 1, Experiment 2b consisted of participants coming into the lab two separate times. Each session consisted of a training and testing component. For the testing component, we used a picture-naming task, a translation task, and a word-naming task to test how well participants had learned the new vocabulary. All of the participants performed the translation task while half of them did the picture-naming task, and half did the word-naming task. The results for each task will be reported separately.

3.3.1.1 Method

3.3.1.1.1 Participants

Forty undergraduates at the University of Arizona participated for course credit. Twenty of the participants performed the picture-naming and translation tasks, while 20 performed the word-naming and translation tasks. All of the participants were native speakers of English.

3.3.1.1.2 Materials

Twenty novel words (taken from Experiment 1) were paired with a picture of a familiar concept. The pictures, adopted from Snodgrass and Vanderwart (1980), were selected from those used in Experiment 2a. Control items were the original Snodgrass and Vanderwart black and white line drawings. Experimental items were the same drawings with colors and designs added to them in an effort to make them unique.

3.3.1.1.3 Design

The design of the experiment was entirely within-subjects. Two training lists were constructed such that each participant learned 10 L2 words in association with control pictures and 10 L2 words in association with uniquely colored pictures. The items were counterbalanced so that all pictures appeared as both control and colored pictures across the experiment. For each participant pictures were presented consistently as either control or colored pictures during the entirety of the experiment.

3.3.1.1.4 Procedure

Individuals participated on two separate days within a 5-day span. The first day was a training session. No data were recorded for analysis from the training session. The second time that participants came to the lab was the testing session. Both the training and testing sessions were identical and consisted of: (1) vocabulary training (participants were told that they were learning a new "alien" language), followed by (2) a recognition task, then (3) either a picture naming task or a word-naming task, and finally (4) a translation task. At the beginning of the first session, participants were shown the pictures used in the experiment on flashcards and asked to name them in English. After they had named each of the pictures correctly, they were asked to write the names of the items down before beginning their training. This was done to ensure name agreement on all of the items because, although L1 labels were never present during training, the translation task did use L1 labels. Participants were then taken into a sound-resistant computer booth where they did their training and testing. All stimuli were displayed on Windows-based computers using the DMDX system developed by J. C. Forster at the University of Arizona.

During vocabulary training, participants first heard a recording of the L2 word over headphones, then saw the L2 word and its corresponding picture for 500ms on the computer monitor, and then heard a second recording of the L2 word. Participants were asked to repeat the L2 word twice into the microphone placed in front of them for recording purposes (see Appendix B for complete instructions). All 20 items were blocked and presented randomly within blocks four separate times for a total of 80 training trials.

The vocabulary training was followed by a recognition task, which consisted of the presentation of a picture followed by one of the L2 labels. The 20 picture-label pairs were presented in random order four separate times: two trials were correct (the picture was paired with its new label) and two trials were incorrect (the picture was paired with the wrong label). Participants were instructed to press a "yes" button if the picture and L2 word matched and a "no" button if they did not. Participants were given feedback for each item, including whether they were correct or not, as well as their reaction times. After the recognition task, participants were given either the picture-naming task or the word-naming task.

3.3.2 Picture Naming

Twenty of the participants performed the picture-naming task in addition to the translation task. In this task, pictures were presented on the computer monitor to participants. Each trial began with a fixation cross at the center of the monitor screen for one second, followed immediately by the picture to be named for 500 ms. Participants were instructed to speak the L2 name of the picture into the microphone as quickly as possible. Their vocal response triggered a voice key, which stopped the computer's timer. Latencies were measured from the time the picture appeared on the monitor until the voice key was triggered. All responses were recorded on tape so that they could be checked for errors. All 20 pictures were blocked and presented randomly within blocks four separate times for a total of 80 trials.

3.3.2.1 Picture Naming Results

Data from 6 participants of this group of 20 were excluded from analysis for failing to reach the predetermined accuracy criterion of 80%. For the remaining 14 participants, all incorrect responses, including fluency errors like stutterings and "um's" were excluded from analysis (6.6%). Additionally, response times longer than 3000 ms or shorter than 200 ms, were excluded from analysis (less than 1%). The average error rate was 7.1% in the control picture condition and 6.0% in the unique-pictures condition. This difference was not significant.

Figure 3-3 displays a clear naming advantage for words learned in association with uniquely colored pictures. Separate analyses of variance (ANOVA) with participants (F1) and items (F2) as random variables reveal that salient pictures (M = 889.94) were named significantly faster than control pictures (M = 944.84): ($F1(1, 13)$ =

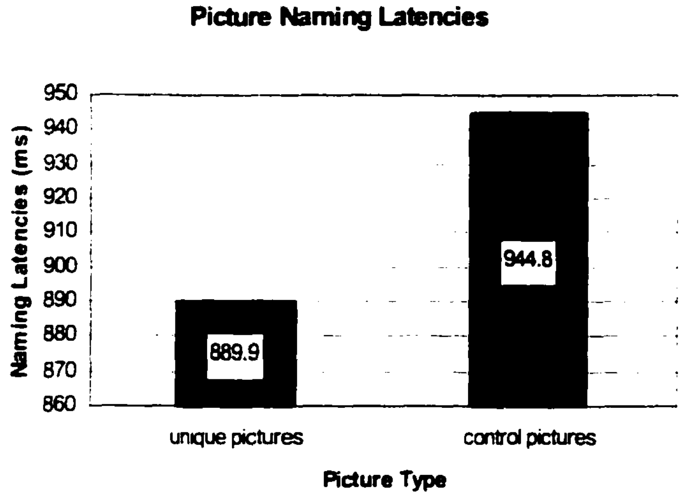

Figure 3-3: Mean L2 naming latencies by picture type at training

5.93, $p = 0.03$; $F2(1, 79) = 6.84$, $p = 0.01$). Thus, the experimental manipulation of colored pictures exerted a facilitation effect on picture naming times.

3.3.2.2 Discussion

The purpose of Experiment 2b was to see if L2 vocabulary learning is facilitated when uniquely colored pictures (which were determined to slow retrieval of their corresponding L1 labels in Experiment 2a) are used as training stimuli. In the present task, we used a picture naming task at test. A clear facilitation effect was observed for those items learned in association with uniquely colored pictures. Participants were able to retrieve from memory L2 labels learned in the unique condition faster than they were able to retrieve L2 labels learned in the control condition. Because the "unique" pictures were determined to slow L1 label retrieval (Experiment 2a), the present findings provide support for the hypothesis that participants' learning of L2 vocabulary can be enhanced when training stimuli designed to delay activation of L1 equivalents are used during L2 vocabulary learning tasks.

The picture naming results indicate a clear facilitation effect for unique pictures on L2 vocabulary learning, but the locus of the facilitation effect is still in question. In order to name a picture, one must first recognize the picture, retrieve the appropriate conceptual representation for that picture, select the appropriate lexical entry for that concept, and then articulate the word. The locus of facilitation observed in the present picture-naming experiment could, theoretically, have been at any one of these stages. It may have been that uniquely colored pictures led to faster concept retrieval times, lexical retrieval times, or word-form encoding times. However, given the results of Experiment 2a, it is hard to imagine how the experimental items could have led to faster concept retrieval times. In fact, the purpose of Experiment 2a was to find those experimental items that best inhibited concept retrieval. But it may have been the case that learning new words in association with uniquely colored items versus control items led to faster word-form encoding times in the production task. This seems unlikely, but we tested this possibility with the word-naming task.

3.3.3 Word Naming

Twenty of the participants performed the word-naming task in addition to the translation task. This task was essentially the same as the picture-naming task, the only difference being that participants named words instead of pictures. Participants named each of the 20 items four separate times: twice in English and twice in 'Alien.' All 20 items were blocked and presented randomly within blocks; and the order of target languages was counterbalanced across participants. Each trial began with a fixation cross at the center of the monitor screen for one second, followed immediately by the word to be named for 500 ms. Participants were instructed to name the word as quickly as possible. Again, latencies were measured from the time the word appeared on the monitor until the voice key was triggered. All responses were recorded on tape so that they could be checked for errors.

3.3.3.1 Word Naming Results

Data from 4 participants from this group of 20 were excluded from analysis for failing to reach the predetermined accuracy criterion of 80% (in the translation task). For the remaining 16 participants, all incorrect responses, including fluency errors like stutterings and "um's", as well as response times longer than 2000 ms or shorter than 200 ms, were counted as errors and excluded from analysis (less than 1%). Figure 3-4

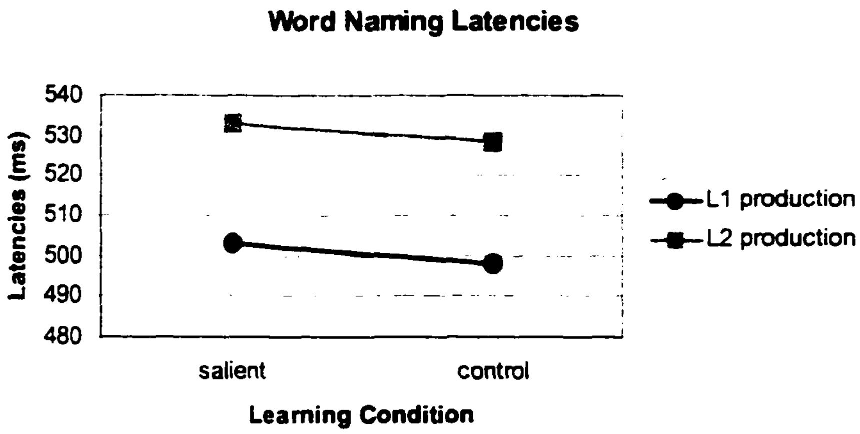

Figure 3-4: Word naming latencies by learning condition

displays a clear time advantage for words named in participants' L1 versus their "L2," as would be expected. Separate analyses of variance (ANOVA) with participants (F1) and items (F2) as random variables reveal a main effect for language of production, with English words (M = 500.59) being named significantly faster than "L2" words (M = 530.63): ($F1(1, 15) = 6.41$, $p = 0.023$; $F2(1, 78) = 23.05$, $p < 0.001$). Crucially, though, there was no interaction with learning condition (all Fs < 1), suggesting that learning words in association with salient versus control pictures has no bearing on how those words are recognized or articulated. This suggests that the locus of the saliency effect observed in the picture-naming experiment was not at the level where word-form encoding occurs, but rather at a more abstract level of representation.

3.3.3.2 Discussion

Participants named English words faster than newly learned 'L2' words, which was not surprising. Of more interest in this experiment was whether or not the saliency effect observed in the picture-naming experiment could be elicited with a word-naming task. Because word naming arguably takes place without activating the conceptual or semantic properties of the to-be-named word (Damian, Vigliocco & Levelt, 2001), any indication of the facilitation effect with this task would serve to isolate the form level of representation as a contributor in this effect. No effect was observed. Alien words were named equally fast, regardless of what type of picture they had been associated with during learning. With no interaction between item type (salient vs. control) and language of production, it is clear that the training manipulation did not affect how form-level properties of L2 words were encoded in memory during learning or how they were subsequently retrieved in lexical processing tasks. Consequently, the locus of the facilitation effect in the picture naming task could not have been due to a better encoding of the form-level properties of L2 labels into memory.

If the locus of the facilitory effect observed in the picture naming task was not attributable to differences at the form level, then we can narrow down the locus of the effect to differences at more abstract levels of representation (i.e. the conceptual level). Before discussing further how uniquely colored pictures may have led to representational differences in the minds of the learners, it is first necessary to address a possible confound present in the design of the picture naming experiment. In the picture naming experiment, the same pictures were used both during training and testing. This design leaves open the possibility that the associations formed between the unique pictures themselves (not the concepts they refer to) and their corresponding labels were stronger than those formed between the control pictures and their corresponding labels. In other words, the facilitation effect observed in picture naming could have been due to the unique pictures being better memory cues than the control pictures. In terms of lexical (mental) representations, it could be that there is nothing different between the L2 lexical representations established in association with unique pictures and those established in association with the control pictures. In order to test this possibility, we used a translation task. Because the stimuli used at test in the translation task are identical for all participants across learning conditions, any observed differences in translation performance for L2 items learned in association with "unique" pictures versus those learned in association with control pictures would have to be due to the differences in the training conditions. A significant interaction between translation performance and learning condition would indicate that the training manipulation significantly affected how L2 lexical representations came to be represented and subsequently processed.

3.3.4 L1-L2 and L2-L1 Translation

Forty participants performed the translation task. In this task, participants were shown a row of hash marks ("#######") for one second (for orienting purposes only) followed by the word to be translated. For example, in the L1 to L2 blocks, an English word appeared and participants were asked to speak the "L2" translation equivalent into the microphone as quickly as possible. Their vocal response triggered a voice key, stopping the computer's timer. Latencies were measured from the time the word to be translated was presented until the voice key was triggered. All responses were recorded on tape so that they could be checked for errors. All incorrect responses as well as fluency errors like stutterings and "um's" were counted as errors. In order to control for any differences that may arise due to order of translation direction, translation direction was counterbalanced such that half of the participants performed forward translation (L1 to L2) first, while the other half performed backward translation (L2 to L1) first.

3.3.4.1 Translation Results

Data from 10 participants were excluded from analysis for failing to reach the predetermined accuracy criterion of 80% (6 from the picture-naming group, and 4 from the word-naming group). For the remaining 30 participants, all incorrect responses, including fluency errors like stutterings and "um's", as well as response times longer than

3000 ms or shorter than 200 ms, were counted as errors and excluded from analysis (8.1%). The average error rate was 7.8% for items learned in association with control pictures and 8.9% for items learned in association with colored pictures. Separate analyses of variance (ANOVA) with participants (F1) and items (F2) as random variables reveal no main effects of learning condition ($F1(1, 29) = 2.08$, $p > 0.05$; $F2(1, 78) = 2.79$, $p > 0.05$) or translation direction ($F1(1, 29) = 3.45$, $p > 0.05$; $F2(1, 78) = 1.70$, $p > 0.05$). Crucially, though, the interaction between learning condition and translation direction was significant ($F1(1, 29) = 7.29$, $p = 0.01$; $F2(1, 78) = 7.83$, $p = 0.006$). Figure 3-5 presents the nature of this significant interaction. L2 labels learned in association with unique pictures are translated in a symmetrical fashion, with approximately equal L1 and L2 production times. L2 labels learned in association with control pictures, on the other

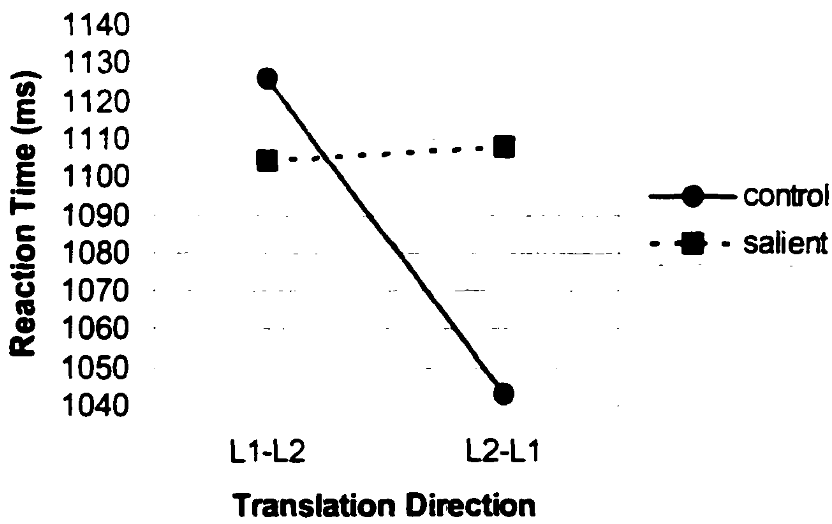

Figure 3-5: Translation latencies (msec) by training condition and translation direction.

hand, are translated asymmetrically, such that production times are much faster when L1 is the language of production than when L2 is the language of production. Planned comparisons show that the translation latencies for items learned in association with control pictures were significantly slower when L2 was the language of production ($F1(1, 29) = 8.69$, $p = 0.006$; $F2(1, 78) = 6.36$, $p = 0.014$) and no difference between L1 and L2 production for items learned in association with unique pictures (all $Fs < 1$).

3.3.4.2 Discussion

Participants translated 20 recently learned L2 labels in both translation directions (L1-L2 and L2-L1). L2 labels learned in association with unique pictures were translated in both directions symmetrically. L2 labels learned in association with control pictures (black and white line drawings) exhibited a strong asymmetry in translation performance, such that translation times were significantly faster when L1 was the language of production. A significant interaction between training condition and translation direction indicated that the training manipulation significantly affected how the items in the unique pictures condition were translated. A possible explanation for how the training manipulation could affect how the new L2 lexical entries were represented and processed is provided in what follows.

3.4 General Discussion

The present set of experiments has demonstrated that associating target vocabulary with unique cues during L2 lexical acquisition can have a rather dramatic effect on the way in which the newly learned items are processed. The findings of

Experiment 1 suggested that the faster L1 equivalents were likely to come to mind during the learning of a new L2 label, the slower participants were to produce that L2 label in a subsequent naming task. In terms of L2 vocabulary learning, these findings are important. The possibility that delaying the activation of L1 equivalents during L2 vocabulary learning can enhance L2 vocabulary learning holds very important implications for the field of second language vocabulary learning. Experiment 2 was explicitly designed to test this possibility by using training stimuli found to delay participants' retrieval of their L1 labels relative to control stimuli (Experiment 2a). L2 labels learned in association with these unique pictures were not named faster than L2 labels learned in association with control pictures in a word naming experiment, but they were named faster in a picture naming experiment. Furthermore, a significant interaction in the translation task between training condition and translation direction revealed that the training manipulation had significantly affected how the "unique" L2 labels were translated relative to the "control" L2 labels. The nature of the interaction was such that the unique L2 labels were translated symmetrically, while the control L2 labels exhibited a strong asymmetrical translation pattern, with translation times being significantly faster when L1 was the language of production.

An explanation for these findings would need to consider the following: (1) unique pictures slow retrieval of their associated L1 labels; (2) learning L2 labels in association with these unique pictures leads to faster picture naming times, but not word naming times; (3) learning L2 labels in association with these unique pictures leads to a symmetrical translation pattern, which is significantly different from the translation pattern exhibited by L2 labels learned in association with control pictures.

How is it that the unique stimuli used in Experiment 2 could facilitate L2 vocabulary learning the way that it did? In order to answer this question, we first need to review what is presumably involved in L2 vocabulary learning. Learning a new L2 word involves establishing a new L2 form (orthographic or phonological) in memory and then associating that form with appropriate meaning representations. In Chapters 1 and 2, several different models were reviewed that propose a variety of possible "routes" between new L2 form and meaning representations. One possibility is that the new L2 form is associated directly with its L1 equivalent form. When this is the case, meaning is accessed via the L1 translation-equivalent form. The word-association model (Potter et al., 1984) and the revised hierarchical model (Kroll & Stewart, 1994) capture this possible architecture best. As we saw in Chapter 2, though, as well as in the study reported by Altarriba and Mathis (1997), learners establish direct L2 form-meaning connection in even the earliest stages of word learning. For this reason, the predictions of the word association models (i.e. the word association and revised hierarchical models) cannot be correct for L2 word learners. The results of the present study lend further to this conclusion. The word-association models assume that new L2 forms are associated directly to L1 equivalent forms. This works well to account for how L2 words were learned in the control condition. The black and white line drawings used in the control condition did nothing to inhibit activation of the appropriate L1 lexical entry, allowing the new L2 label to be associated with its L1 equivalent. It is not clear, though, how the word-association models would account for the effect that the unique-pictures training condition had. If the critical difference between the unique pictures and the control pictures was that the unique pictures served to delay retrieval times of L1 labels, then the word association models would have to assume that the association between new L2 label and its L1 equivalent was delayed in the critical condition and not in the control condition. It is not clear, though, how delaying the formation of the mapping would lead to a difference in terms of how the two types of L2 entries were represented. Within the parameters of the word association models, it is not clear how the training manipulation could have had the effect that it did. For this reason, I will not consider it as a possible account of the present findings.

Another possibility is that the new L2 form is associated directly with an already-established meaning representation (e.g. concept), with no form-level associations between L1 and L2 equivalents. This possible architecture is best represented by the "concept-mediation model" (Potter et al., 1984), the distributed feature model (De Groot, 1992) and the sense model (c.f. Chapter 1).

In the strictest terms, the concept mediation model is not able to account for the effects of the training manipulation either. The concept mediation model assumes that new L2 forms are mapped onto existing conceptual representations. Again, it is not clear how delaying activation of the L1 equivalent would change the nature of the L2 form-meaning mapping within the parameters of this model. If the unique pictures in the critical condition served to slow recognition or concept retrieval, then according to the concept-mediation model, the formation of the new L2 form-meaning mapping would be delayed in the critical condition and not in the control condition. Again, though, the concept mediation model would have to predict that the nature of the new L2 form-meaning mapping was identical between training conditions – only the time at which the mapping was formed would be different. One could argue, though, that in the critical training condition of the present study, the unique pictures, by virtue of having significantly slowed L1 label retrieval times in Experiment 2a, may have been unique enough for learners to classify them differently from the way that they did the control pictures. That is, the critical-pictures were unique enough to lead participants to develop new concepts for them and then form direct L2 form-meaning connections between new L2 label and new concept. This is not completely implausible given the context in which the unique pictures were presented. Participants were told that they were going to learn several words from an "Alien" language and that the pictures from the "alien culture" were different in color and texture from what they were used to (see Appendix B for instructions). This difference in picture classification during vocabulary learning could very plausibly lead to a unique L2 form-meaning mapping. In the control condition the new L2 form-meaning mapping may have been something like <"floop" – "THUMB"> (where "THUMB" represents an L1 lexical concept). This is because the control picture served to readily activate the L1 lexical concept "THUMB." In the critical condition, on the other hand, the new L2 form-meaning mapping would be something like <"floop" – "ALIEN THUMB">, where "ALIEN THUMB" represents the newly established L2 lexical concept (see Figure 3-8).

It should be pointed out that each of the "meaning-mediation" models (i.e. the concept mediation, distributed feature and sense models) could adopt this type of account, even though all of them assume that under normal word learning conditions L2 labels are associated with already-established meaning representations. It is virtually impossible for either of the word-association models to make this adjustment because it is unlikely that the status of the L1 equivalent form would be affected during learning by the color of the pictures.

Even though the meaning-mediation models could allow for the possibility that the unique pictures training condition led to the establishment of unique L2 form-meaning connections, the question still remains whether their architectures would predict the patterns of picture naming and translation data observed in Experiment 2b for both the critical and control items.

In the picture naming experiment, the critical items were named significantly faster than were the control items. Because the meaning-mediation models cannot appeal to more connections being established in one condition than another, they must appeal to a separate mechanism. One such mechanism, Green's (1998) mechanism of interference control, was introduced in Chapter 1. According to Green's interference control hypothesis, suppression of the L1 must occur in order for L2 production to proceed. This account is very workable when applied to translation asymmetries because it seems quite plausible that L1 suppression is more difficult than L2 suppression, resulting in faster L2-L1 translation latencies than L1-L2 translation times. But in order for the interference control hypothesis to account for differences in L2 picture naming times, it must assume that some stimuli introduce more L1 interference than others. Specifically, it must assume that L2 labels learned in association with control pictures invoke more L1 interference (because they are named slower) than L2 labels learned in association with unique pictures. This falls in line with what was suggested above. That is, if participants formed L2 form-meaning connections between new L2 labels and already-established meaning representations in the control condition, and, conversely, the form-meaning connections established in the unique pictures condition were between new L2 label and newly established meaning representation, then it follows straightforwardly that the form-meaning connections established in the control condition would invoke more L1 interference. Form-meaning connections established in the unique pictures condition, on the other hand, would not invoke much L1 interference because the newly established meaning representation would have very weak interconnections with already established semantic representations (see Figure 3-8). This account, by appealing to an interference control mechanism, appears capable of explaining the picture naming data. The question still remains, though, whether this explanation will work to account for the translation data.

In the translation experiment, L2 labels learned in association with uniquely colored pictures were translated symmetrically while L2 labels learned in association with control items exhibited a strong asymmetry. The nature of this asymmetry was such that translation times were significantly faster when L1was the language of production. Again, assuming that form-meaning connections established in the control condition are between a new L2 form and an already-established sense representation copied over from the L1 lexical entry, it logically follows that invoking these form-meaning connections will introduce interference from semantically related information (because the sense representations integrated into these form-meaning connections are strongly interconnected with other related meaning-level representations). The impact that this interference has on the L1 production system should be fairly negligible seeing how the L1 system is well-practiced in resolving interference during production. The impact that interference at the level of lexical retrieval has on the L2 production system is much greater. This is to be expected seeing as how the L2 production system (of the subjects in the present experiment) was not well-practiced in resolving interference at this level. The interference-control explanation serves well, then, to account for the translation asymmetry exhibited by items learned in the control condition. The remaining question

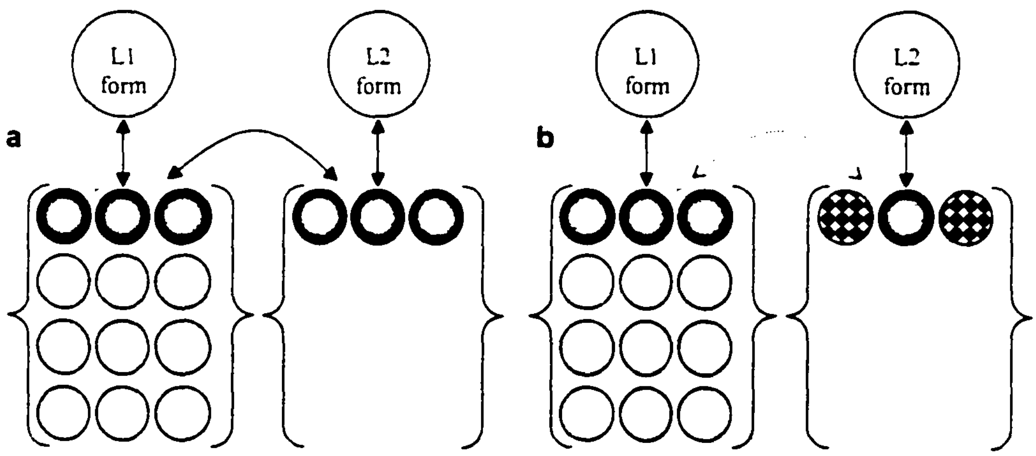

Figure 3-8: Sense model account of the form-meaning connections established in Experiment 2b. Figure 3-8a depicts the "representational consequence" of word learning in the control condition – L2 form-meaning connections between new L2 forms and an already-established sense representation copied over from the L1 equivalent lexical entry. Figure 3-8b depicts the representational consequence of learning new L2 forms in the unique pictures condition – L2 form-meaning connections between new L2 form and newly established meaning representation. Note how a limited amount of overlap between L1 and L2 semantic representations is assumed in Figure 3-8b.

concerns whether the interference-control account can explain the symmetrical translation performance exhibited for items learned in association with unique pictures.

According to the constraints of the meaning-mediation models, translation must be semantically mediated. For this reason, positing that the form-meaning connections established in the critical condition are comprised of newly created meaning representations that are only weakly interconnected with related representations in the semantic system predicts that translation will be rather difficult. If subjects found the uniquely colored (Alien) pictures to be sufficiently different to warrant developing new semantic representations for them, then it is clear that there will be little in common between these newly-established semantic representations and the already-established ones represented in their L1 "equivalents". Hence, translation should be an effortful task. The translation latencies are in accord with this. Translation latencies are slower overall for the labels learned in the unique pictures condition than in the control condition (though this difference is not significant).

In terms of the symmetrical pattern of translation latencies exhibited for critical items, the sense model assumes that this is a result of the limited amount of interference between systems during L2 production. That is, when L2 was the language of production in translation, the to-be-selected L2 lexical entry had so little overlap (at the semantic level) with "related" items that lexical retrieval was relatively un-interfered with. This allowed L2 production to proceed much like L1 production, where interference, along with the ability to control it, is much greater.

3.5 Conclusion

Overall, it appears that L2 vocabulary items learned in association with the unique pictures used in the present vocabulary learning experiment came to be represented in a fundamentally different way than did L2 vocabulary items learned in association with standard black and white line drawings. This difference is captured remarkably well by the sense model (though the other meaning-mediation models could account for this difference as well). When a new L2 label is learned in association with a unique variant of a familiar cue, it appears that learners classify the stimulus differently enough to warrant creating a new meaning-level representation. The resulting meaning-level representation is quite distinct within the semantic system, and for this reason shares very little overlap with "related" representations. This lack of overlap with semantically related representations benefits the learner. The nature of this benefit is a reduced amount of interference between competitors within the L2 production system when these unique representations need to be retrieved and produced. The present study provided support for this account with superior picture naming performance for labels learned in the "low-interference" condition as well as with symmetrical translation performance for these same items.

These findings have very important implications (both practical and theoretical) for the field of second language vocabulary learning. Practically speaking, this study shows that using training stimuli designed to slow activation of L1 equivalents can enhance L2 vocabulary learning. Theoretically speaking, this study serves to shed light on the nature of this "enhancement." When the learning context is sufficiently unique, learners appear to classify stimuli differently than they would in a typical learning context. As the present findings indicate, classifying a training stimulus differently can lead to "coordinate" L2 form-meaning connections, independent from those already established in the L1. I argue that this "coordinate" architecture was best suited to account for the enhanced L2 lexical processing performance observed in the present study.

Chapter 4 Masked Translation Priming

Turning now to masked translation priming, my attention shifts to how different models of bilingual lexical representation and processing can account for the lexical performance exhibited by fluent bilinguals as opposed to novice learners. In Chapters 2 and 3, I argued that models which propose lexical level associations between the lexicons (to the exclusion of direct L2 form-meaning connections) could not account for the semantic effects observed in word learning and subsequent picture naming and translation tasks. In the present chapter, I will discuss how these word-association models, as well as the meaning-mediation models, fare in terms of explaining the priming asymmetry seen in masked translation priming with bilinguals.

The nature of this well-known priming asymmetry is such that although masked primes in the dominant language (L1) prime translation equivalent words in the non-dominant language (L2), there is little or no priming in the reverse direction, unless the prime and target words are cognates from same-script languages (e.g., rich-rico) (Sanchez-Casas, Davis & Garcia-Albea, 1992). This is true despite the fact that the very same L2 primes are capable of priming L2 targets (Gollan, Forster & Frost, 1997; Jiang, 1999). Given this asymmetry, we can immediately eliminate all those models that posit symmetrical connections between L1 and L2 lexicons. Furthermore, it seems reasonable to exclude the interference control hypothesis as a viable account for this particular asymmetry. Remember, Green (1998) proposed the interference control hypothesis in order to account for the translation asymmetry while retaining the basic tenets of the concept-mediation model (i.e. direct L1 and L2 form-meaning connections). According to Green (1998), suppressing L1 is more difficult than suppressing L2, leading to slower L2 production times relative to L1 production. Green has also argued that bilinguals are less able to control competition (between semantically related competitors) within the L2 production system, leading to more semantic interference effects in L2 production than what is typically observed in L1 production. Though the interference control hypothesis is very attractive as an account of bilingual performance in translation production tasks, it is not clear how it could come to bear on masked translation priming findings. Namely, if bilinguals are less able to control competition from semantically related distractors within their L2 system, it is not clear how the appropriate L2 lexical entry could be selected in time under the speeded conditions that are used in masked priming. Yet both L1-L2 and L2-L2 masked priming effects are very robust.

It appears, then, that the only way a model of bilingual lexical representation and processing could account for this asymmetry would be to build into the model either asymmetrical form-meaning connections or semantic representations between the lexicons. There are only two such models that I am aware of in the bilingual lexical processing literature: the revised hierarchical model (Kroll & Stewart, 1994) and the sense model.

The Revised Hierarchical Model proposed by Kroll & Stewart (1994) has attempted to explain the priming asymmetry by suggesting that (1) the locus of the cross-language priming effect is conceptual and (2) that relative to L1 representations, L2 lexical representations are only weakly connected to the conceptual system. That is, priming is effective in the L1-L2 direction because the masked L1 prime activates a shared conceptual node, which then preactivates the L2 translation-equivalent lexical form, thus facilitating recognition. However, priming is not effective in the L2-L1 direction because L2 primes do not automatically activate their conceptual representations, resulting in no preactivation of the L1 translation-equivalent form and thus no priming.

This account is appealing, but it suffers from a number of weaknesses. First, the nature of the connection between an L2 lexical form and its associated concept needs to be better explained because if the locus of the priming effect is at the level of concepts, and L2 primes are too weakly connected to conceptual representations to be effective as masked primes, then we should not observe within-language priming with masked L2 primes – yet L2-L2 masked priming effects are very robust (Gollan et al., 1997, Jiang, 1999). Second, the Revised Hierarchical Model (RHM) posits that in order to obtain L2-L1 priming, the necessary links are L2-C and C-L1, where C represents the relevant conceptual node. To explain why there is no priming, the model assumes that the L2-C link is too weak. However, the relevant links for L1-L2 priming are L1-C and C-L2, and since there is priming in this case, it must be assumed that the C-L2 link is strong enough to permit priming. Thus, the asymmetry in priming would have to be explained by an asymmetry in the links between L2 words and conceptual nodes: The C-L2 link must be stronger than the L2-C link. As currently specified, the RHM assumes a symmetrical (weak) connection between L2 forms and their associated concepts. Note, though, how further specification of this connection would lead to a very strange asymmetry in connection strengths. Namely, if the C-L2 link is stronger than the L2-C link, then language production should precede comprehension in bilingual lexical development; but this is clearly not right.

A further problem for an account expressed in terms of strengths of connections is the possibility that the priming asymmetry may be task specific. Jiang & Forster (2001) claim to have obtained L2-L1 priming for non-cognates in a speeded episodic recognition task. Also, although Grainger & Frenck-Mestre (1998) reported that "The responses of English-French bilinguals performing semantic categorization and lexical decision tasks were facilitated by [L2] prime stimuli that were non-cognate translation equivalents of the [L1] targets..." (p. 601), a careful reading of their paper reveals that significant L2-L1 priming effects were observed in the semantic categorization task, with only a trend towards priming in the lexical decision task. These results are intriguing because they suggest that the type of task differentially sensitive to the cross-language priming effect. That is, if it is the case that masked L2-L1 priming can be observed with some tasks (e.g. semantic categorization) and not others (e.g. lexical decision), then we could rule out the possibility that the L2-C link is too weak to drive masked translation priming.

Presently, it would be premature to put forth such a hypothesis about task differences without further investigation because the Grainger and Frenck-Mestre finding contradicted an earlier result. Sanchez-Casas, Davis and Garcia-Albea (1992) used a semantic categorization task in order to ensure the access of lexical knowledge, but found no priming in the L2-L1 direction for non-cognate translation pairs. The most striking difference between the Sanchez-Casas et al. (1992) study and the Grainger and Frenck-

Mestre (1998) study is in how the categories were presented to the participants. In the Grainger and Frenck-Mestre study, a blocked presentation of categories was used. Participants were presented with a category, followed by 24 exemplars and non-exemplars. In the Sanchez-Casas study, categories changed for each successive item. The disadvantage of this procedure is that each experimental item may effectively constitute a separate task, and one consequence of this "task-switching paradigm" may be diminished priming effects. As an example in support of such a possibility, Catchpole (1987) found that blocking of categories restored the frequency effect for non-exemplars, an effect that had previously gone undetected in experiments using different categories on each trial (Balota & Chumbley, 1984). Blocking of semantic categories may also affect cross-language priming.

However, just as there is good reason to question the null result reported by Sanchez-Casas et al. (1992), there is also good reason to question some aspects of the design of the Grainger and Frenck-Mestre study. Their design was completely within-subjects, where the same target was presented to the same participant in both the translation condition and in the control condition, and also as an exemplar and a non-exemplar. Furthermore, the design involved 4 different prime exposures. This resulted in each participant (N=12) responding to the same target word 16 different times during the course of the experiment. Obviously, this repetition would create a strong episodic record of the targets, and it is possible that this is critical for L2-L1 priming, as claimed by Jiang and Forster (2001), who reported L2-L1 priming in a speeded episodic recognition task, but not in a lexical decision task.

The key question addressed in this chapter is whether it is possible to obtain L2-L1 priming in a semantic categorization task without repeating any target items, and without changing categories on each trial. If this is possible, then the failure to observe such an effect in lexical decision could not be explained by weak L2 form-meaning connections. If anything, the switch to a semantic categorization task (where consideration of meaning is mandatory) would make the weakness of the connection between L2 forms and their concepts even more obvious. Observing L2-L1 priming in semantic categorization but not lexical decision would demand an alternative account of the cross-language priming asymmetry.

4.1 Experiment 1

This experiment was designed to test whether L2-L1 masked translation priming could be obtained in a semantic categorization task, as reported by Grainger and Frenck-Mestre (1998). In an effort to avoid the possible pitfalls of earlier experiments, items were blocked according to categories, and no item (target or prime) was repeated for any given participant.

4.1.1 Method

4.1.1.1 Participants.

Twenty Japanese-English bilinguals were recruited from the University of Arizona campus community. All participants were native speakers of Japanese and were employed as graduate students at the University of Arizona. Participants had received a minimum of 6 years of English instruction while in Japan, and at the time of testing all had been living in the United States for at least 2 years. All participants were paid for their participation.

4.1.1.2 Materials

In order to ensure translation equivalency for each English-Japanese prime-target word pair, 5 Japanese-English bilinguals (from the same population as the participants in the experiment) were asked to translate a list of 170 items from English into Japanese (L2-L1); and another group of 5 was asked to translate the same items in the opposite direction (L1-L2). Only those word pairs that were translated identically in each direction by all participants were selected as critical items. Fifty-two word pairs met this criterion. The critical items belonged to 11 different semantic categories, with a minimum of 4 per category. In order to make use of a blocked design, an additional 58 items were selected to serve as "practice" items on trials preceding the critical items in each category. This resulted in a total of 110 "Yes" items (10 per category), with a minimum of 4 per category being critical. Additionally, 110 non-exemplar targets (NO items) were chosen. These were chosen so as to ensure that they could not be construed as belonging to any of the 11 categories. An additional 220 English words were selected to serve as primes on control and practice exemplar trials, as well as on non-exemplar trials. These were unrelated to their targets, but were matched with the critical primes for frequency, concreteness, imageability and word-length.

Care was taken to ensure that none of the Japanese targets shared cognate status with their English primes. All targets were presented in Kanji characters except for those appearing in the INSECT category. Both exemplars and non-exemplars in this category were presented in the Katakana script.

4.1.1.3 Design and Procedure

Items were blocked according to semantic category. For each category, there were 10 exemplar and 10 non-exemplar trials. Each trial consisted of the following sequence (adapted from Grainger & Frenck-Mestre): first, the participant was presented with a forward mask (##########) for 500ms, followed by an English prime (translation or control) in lowercase letters for 50ms, followed by a backwards mask for 150ms, and then the Japanese target word for 500ms. Half of the critical targets per list were preceded by their translation equivalents and half were preceded by a control prime. Two counterbalanced lists were constructed such that if a target was preceded by its translation prime on List A, it was preceded by its control prime on List B and vice versa. No target word or prime word was repeated within lists. Participants were asked to indicate whether the target belonged to the category by pressing either a YES button or a NO button. Stimuli were presented randomly within categories, using the DMDX package developed at the University of Arizona by J.C. Forster, with the only constraint that the "practice" exemplars (and an equal number of non-exemplars) preceded the critical items. Response times (RTs) were recorded to the nearest millisecond.

4.1.2 Results

Data from trials on which an error occurred were discarded and outliers were replaced with values equal to cutoffs established 2 *SD* units above and below the mean for each participant. Mean response times for experimental exemplar targets were 475 ms for Japanese targets preceded by their masked English translation equivalents, and 494 ms for targets preceded by masked English control primes. An ANOVA showed that the difference of 19 ms was highly significant ($F1(1,19) = 12.84$, $P = 0.002$; $F2(1,51) = 24.38$, $P < 0.001$). The mean error rate was 1.2% and was not different for targets preceded by translation primes versus control primes.

4.1.3 Discussion

The existence of a reliable 19 ms L2-L1 priming effect confirms the findings of Grainger and Frenck-Mestre (1998), and indicates that L2-L1 priming is possible if a semantic categorization task is used rather than lexical decision. This finding also suggests that the most effective procedure is to block items by category, rather than changing category from trial to trial, as in the case of the Sanchez-Casas et al.(1992) study. Given the previous failures to find masked L2-L1 priming effects for non-cognate translation equivalents in lexical decision (Gollan et al., 1997; Jiang, 1999), the present findings, in conjunction with those of Grainger and Frenck-Mestre, are important in providing converging evidence that the conditions presented by the semantic categorization task produce an L2-L1 priming effect with masked L2 primes.

There is one alternative possibility that should be eliminated before considering how the semantic categorization task achieves this effect. It could be that the degree of proficiency in L2 is critical, and the participants tested in Experiment 1, unlike those in previous experiments, were able to exhibit L2-L1 priming regardless of the task. To rule out these possibilities, it is necessary to demonstrate that the same materials and participants would not have produced L2-L1 priming in a lexical decision task. In order to determine that this was the case, the same set of participants were tested six months later using the same materials in a lexical decision task.

4.2 Experiment 2

4.2.1 Method

4.2.1.1 Participants

Eighteen Japanese-English bilinguals from the same population were recruited for the lexical decision experiment. Fifteen of these 18 participants had participated in the semantic categorization experiment 6 months prior.

4.2.1.2 Materials

The materials were identical to those used in the semantic categorization except for the 110 Japanese nonwords that had to be created for the present experiment. These were created by dividing a column of two-character Japanese words into two columns and randomly sorting one of those columns and then rejoining the two columns. Any resulting combinations that were thought by a native speaker of Japanese to be remotely word-like were discarded.

4.2.1.3 Design and Procedure

The design and procedure was largely identical to that of Experiment 1 with the obvious difference that in this task participants were asked to indicate whether the target was a legal Japanese word or not instead of indicating whether the target belonged to a specific category. The sequence of forward mask, prime, backwards mask and target was identical to that of Experiment 1, as was the duration of each event in the sequence. Once again, two lists were constructed in order to counterbalance across the prime factor. Targets that were preceded by their masked L2 translation equivalent on List A were preceded by a masked control prime on List B, and vice versa. Items were presented randomly to each participant and RTs were recorded to the nearest millisecond.

4.2.2 Results

The same trimming procedure used in Experiment 1 was employed again here. Mean response times were 529 ms for Japanese targets preceded by their English (L2) translation equivalents, and 525 ms for targets preceded by English control primes. This slight inhibitory effect of 4 ms was not found to be significant (all Fs < 1). The mean error rate was 5.6% on nonword trials and 1.3% on experimental trials. Error rates on experimental trials did not differ between conditions. An ANOVA revealed a significant interaction between these RT results and those of Experiment 1 ($F1(1,36) = 4.62$, $P = 0.03$; $F2(1,51) = 6.70$, $P = 0.012$).

4.2.3 Discussion

This result confirms the general finding that masked L2 primes do not facilitate responses to translation-equivalent L1 targets in a lexical decision task. Hence the significant priming observed in Experiment 1 cannot be attributed to special properties of the materials or the participants. Clearly, the nature of the task is critical for L2-L1 priming, though not for priming in the reverse direction, since this effect is readily observed with lexical decision (De Groot & Nas, 1991; Goilan et al., 1997; Jiang, 1999).

The present set of findings demonstrates clearly that the semantic categorization task elicits L2-L1 priming and the lexical decision task does not. Current models of bilingual lexical processing do not account for this task effect. According to the RHM, the only model that has attempted to account for this asymmetry previously, L2 lexical forms do not prime L1 targets because of weak L2 form-meaning connections. Following from this claim, a task such as semantic categorization, which requires participants to focus on the meaning that a particular form refers to when generating decisions, certainly should *not* reveal L2-L1 priming. Yet it does.

In what follows, I discuss two possible explanations for the present set of findings. The first concerns the possibility that the effect represents a decision conflict, rather than a priming effect. The second is the explanation provided by the sense model. According to the sense model, the semantic categorization task serves to restrict the amount and type of semantic information recruited to generate a decision, which enhances the effectiveness of the L2 prime.

The decision conflict interpretation was pointed out by Davis, Kim, and Sanchez-Casas (in press) in their critique of the original finding reported by Grainger and Frenck-Mestre (1998). They argued that the control condition used in the Grainger and Frenck-Mestre (1998) experiment (same as in the present experiment) may have been inappropriate. For the critical items, the unrelated prime was a non-exemplar, while the target was an exemplar. This may have generated a decision conflict for unrelated items, but not for related items (since both L2 prime and L1 target were exemplars), which would produce a congruence effect in the place of a true priming effect. This explanation presupposes that the L2 prime is capable of activating semantic information, and that this information enters into the decision process. If such an effect occurs, it should also occur in a within-language experiment. However, such effects have not been detected in within-language experiments. In a recent experiment using the category ANIMAL (Forster, Mohan, and Hector, in press), the magnitude of repetition priming effects (e.g. *robin-ROBIN*) was found not to be affected by the type of control condition used. That is, the same priming effects were observed regardless of whether the control condition included category congruent items (e.g. *shark-ROBIN*) or incongruent items (*badge-ROBIN*). Furthermore, there was no difference in categorization times between congruent items and incongruent items. Likewise, a recent study by Bueno and Frenck-Mestre (2002) reported similar priming effects for *dolphin-WHALE* when the control condition used non-exemplars as primes (e.g. *helmet-WHALE*) and when the control condition used congruent, but semantically dissimilar, exemplars as primes (e.g. *sparrow-WHALE*). Since neither of these experiments report evidence that using a non-exemplar versus an exemplar as a control prime interacts with priming effects, and because both of these experiments involved L1-L1 priming, where one may expect a stronger semantic effect than in an L2-L1 situation, a "decision conflict" account for the priming effect seems very unlikely.

The sense model provides an alternative explanation of the task-specific nature of L2-L1 priming. As I pointed out in the Introduction, an issue that has not received adequate attention before in the bilingual lexical processing literature is that most words are polysemous and that the range of senses that a word can have tend to be language specific. In the Introduction I described how the English word "black" and the Japanese word "kuroi" are considered to be translation equivalents, despite having little besides the core sense (COLOR) in common. In English "black" can be used to refer to a type of humor, as well as a calamitous day on Wall Street (*black Friday*). In Japanese, "black" can be used to refer to those who are evil-minded (*black belly*), as well as those who are well tanned or guilty of a crime.

One of the central assumptions of the sense model is that learners, when forming new L2 form-meaning connections, map new L2 forms onto just the sense(s) associated with the L1 equivalent that they assume will exist in the L2 (c.f. Kellerman, 1978). Because learners (correctly) assume that most senses of a word are language specific, they only "copy over" the most central senses of the L1 equivalent to subserve the newly established L2 lexical form. This results in a representational asymmetry where all of the senses associated with an L2 form are also associated with the L1 equivalent form, but where only very few of the senses associated with an L1 form are similarly associated with the L2 equivalent form.

The relevance of this for priming is that the amount of translation priming may depend on the number of semantic senses in common between the prime and the target. That is, L2 primes may be ineffective because they are only able to activate a small proportion of the many senses associated with the L1 target. This effectively results in no observable priming in the L2-L1 direction when lexical decision is used. When the direction of priming is reversed, the proportion of L2 senses activated by L1 translation equivalent primes is presumed to be very high, if not 1:1. In so far as lexical decision times for a particular target can be facilitated by preactivation of its relevant semantic senses, priming should certainly be observed in the L1-L2 direction – and it is.

The question remains as to why semantic categorization was effective in eliciting L2-L1 priming effects. If it is correct to assume that L2-L1 priming does not occur because an insufficient number of L1 senses are preactivated by the L2 prime, then there must be something about the semantic categorization task that serves to restrict the number of senses recruited when a decision is generated. What may happen in a semantic categorization task is that only those semantic senses that are relevant to the category are taken into account (e.g., in order to decide that "black" is a color, one does

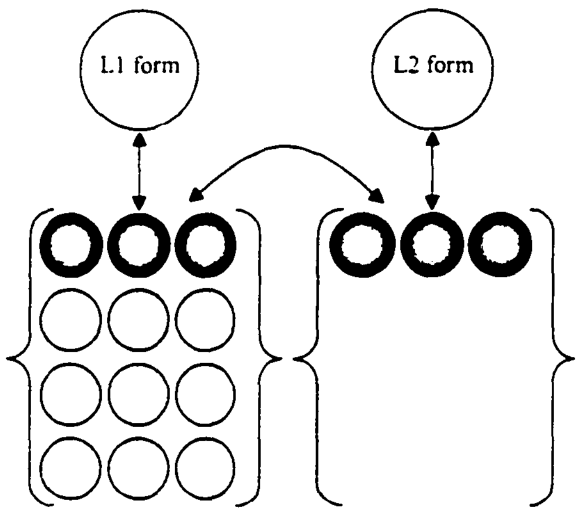

Figure 4-2: Sense model depicting the representational asymmetry between translation equivalents. All of the senses found in association with an L2 form are also represented in the L1 equivalent entry, but many of the L1 senses are not similarly associated with the L2 equivalent form.

not need to take into consideration the multiple senses that "black" may have). This suggests that an L2 prime will be more effective, because the only sense that is relevant for the decision will be the core meaning, and this will always be activated by the prime (because this is the very sense of the word that learners copy over into their L2).

Essentially, I am arguing that the semantic categorization task turns many sense targets into 1-sense targets. If this is accurate, then it should be possible to observe L2-L1 masked translation priming in lexical decision when L1 targets are restricted to just those that are thought to have only one semantic sense. Because sense counts are only provided for English (that I am aware of), testing this aspect of the sense model would require recruiting English-dominant bilinguals who are all roughly equally proficient in an L2 that has a different script. There is a more practical alternative: it is possible to effectively recreate the representational asymmetry assumed by the sense model by carefully selecting English word pairs such that each pair shares one sense in common and consists of a many-sense word and a 1-sense word. If English word pairs meeting these criteria effectively recreate the representational asymmetry thought to exist between L1 and L2 translation equivalents, then, according to the predictions of the sense model, priming should occur in the many-one direction (equivalent to the L1-L2 direction), but should not occur in the one-many direction (the equivalent of the L2-L1 direction).

4.3 Experiments 3a and 3b

This study was carefully designed in an effort to recreate the conditions of the bilingual representational asymmetry assumed by the sense model. Though difficult, it is possible to do this by choosing polysemous words (e.g. *air*) that can be paired with a one-sense synonym, or near synonym (e.g. *oxygen*). As long as all of the words in the experiment are familiar to the subjects, then we can be relatively sure that the participants will know several (if not all) of the senses associated with the many-sense words; and, similarly, we can be reasonably certain that they will know the meaning of the single-sense words (which, presumably, could involve no more than knowing one sense). For those words used in this experiment, then, our monolingual English speakers were much like bilinguals: they had a much richer set of semantic representations for one word than they did for the other word in each word pair. Furthermore, because of the way that the many-sense words were paired with the one-sense words, the word pairs were, at least according to the assumptions of the sense model, akin to L1 (many-sense) words being paired with L2 (1-sense) translation equivalents.

4.3.1 Method

4.3.1.1 Participants

Forty four undergraduates at the University of Arizona participated for course credit. Twenty two of the participants were tested in the many-one direction (Experiment 3a) and twenty two were tested in the one-many direction (Experiment 3b).

4.3.1.2 Materials

With respect to using all-English words to recreate the bilingual representational asymmetry between translation equivalent words, it was imperative to select word pairs that were thought to share a common sense. The "test" that I used to determine whether two words might share a sense or not was to think of whether or not the single sense word in each word pair could have been learned as a "translation" of the many-sense word. Taking *air-oxygen* as an example, it seems reasonable that when people first learn the word "oxygen" they might understand it in terms of what they already know about "air". If so, then this would be an ideal word pair to test the predictions of the sense model. Another way to think about this is to consider whether the word pairs could share a single translation equivalent form. If it is the case that two related words in one language are not distinguished lexically in another, then they must be very similar semantically – i.e., they must share a common sense. Typically, associated word pairs are not going to meet either criterion. Taking, for example, *doctor-nurse*, a very famous associative word pair, it is hard to imagine anyone extracting information from their meaning representation of "doctor" in order to form a new form-meaning connection for "nurse". Likewise, it is hard to imagine these two words sharing a single translation equivalent in any language. That is, every language is going to make a lexical distinction between "doctor" and "nurse". Contrast that with a word pair like *air-oxygen*. I am not sure if there are any languages that do not bother to distinguish these lexically, but it is rather easy to imagine a language not making a lexical distinction between the two (outside of scientific usage), suggesting that they may have a good chance of sharing a common semantic sense.

In addition to the word pairs needing to meet the common sense criterion, it was necessary that one of the words in each word pair had only one sense and that the other had several. Sense counts were taken from WordNet (Miller, 1996; Fellbaum, 1998). One-sense words were paired with words having eleven senses on average (see Appendix

C for a complete list of word pairs and sense counts). Forty word pairs in total were created.

4.3.1.3 Design and Procedure

Both Experiment 3a and 3b were identical in design and procedure. Participants were asked to indicate whether the target word was a legal word in English or not. Items consisted of the following sequence of events, each one following immediately after the other: forward mask (########) for 500ms, prime in lowercase letters for 42ms, target in uppercase letters for 500ms. Two lists were constructed for both Experiments 3a and 3b in order to counterbalance across the prime factor. Targets that were preceded by their related prime on List A were preceded by a control prime on List B and vice versa. Control primes were matched with the experimental primes in terms of frequency and word-length. An equal number of nonwords was generated by the ARC Nonword Database (Rastle, K., Harrington, J., & Coltheart, M., 2002) and interleaved with the experimental items. Items were presented randomly to each participant and RTs were recorded to the nearest millisecond.

4.3.2 Results

The same trimming procedure used in Experiments 1 and 2 was similarly employed here. In Experiment 3a, the many-one (or "L1-L2") condition, mean response times were 550 ms for targets (e.g. *OXYGEN*) preceded by a related prime (e.g. *air*) and 573 ms for targets preceded by a control prime. An ANOVA showed that this facilitation effect of 23 ms was very reliable ($F1(1,20) = 12.41$, $P = 0.002$; $F2(1,38) = 14.26$, $P <$

0.001). The mean error rate was 5.9% and was not different between experimental and control conditions (all *F*s < 1).

In Experiment 3b, the one-many condition (or "L2-L1" direction), the pattern of results was markedly different despite the use of the same word pairs. Mean response times were 538 ms for targets (e.g. *AIR*) preceded by a related prime (e.g. *oxygen*) and 528 ms for targets preceded by a control prime. An ANOVA showed that this inhibition effect of 10 ms was not reliable ($F1(1,20) = 1.71$, $P = 0.21$; $F2(1,38) = 3.24$, $P < 0.079$). The mean error rate was 5.7% and was not different between conditions (all *F*s < 1).

Not surprisingly, given how the priming effects for each direction of priming (many-1 versus 1-many) went in opposite directions, there was a significant interaction between the results of Experiments 3a and 3b ($F1(1,40) = 10.49$, $P = 0.002$; $F2(1,72) = 15.70$, $P < 0.001$).

4.3.3 Discussion

Taken together, this pattern of results suggests that the attempt to use all-English words to recreate the representational asymmetry thought to exist between translation equivalents across languages was successful. In the many-one condition, designed to reproduce the L1-L2 condition, robust priming effects were observed. This mimics the masked translation priming effects found with bilinguals in the L1-L2 direction (De Groot & Nas, 1991; Gollan et al., 1997; Jiang, 1999). Using the same word pairs, but now in the one-many direction (i.e. the L2-L1 direction), no priming was observed. This too mimics the pattern of effects seen with bilinguals – that is, no masked translation priming in the L2-L1 direction (Gollan et al., 1997; Jiang, 1999, Sanchez-Casas et al., 1992; Grainger & Frenck-Mestre, 1998; and Experiment 2 above).

This study provides very strong support for the assumptions of the sense model. According to the sense model, the amount of translation priming depends on the number of semantic senses in common between the prime and the target, and, because of the representational asymmetry, the number of common senses is different depending on the direction of priming. Because, according to the sense model, all of the senses associated with an L2 form are also associated with the L1 equivalent form, and only a few of the senses associated with an L1 form are similarly associated with the L2 equivalent form, L2 (few-sense) primes are less effective than L1 (many-sense) primes. L2 primes are only able to activate a small proportion of the many senses associated with the L1 target, which effectively results in no observable priming in the L2-L1 (few to many sense) direction when lexical decision is used. When the direction of priming is reversed, the proportion of L2 senses activated by L1 translation equivalent primes is presumed to be very high, if not complete. Being able to reproduce the same asymmetrical pattern of priming results with all-English word pairs simply by creating the representational asymmetry between words in each word pair suggests very nicely that the account given by the sense model for the cross-language priming asymmetry must be correct.

Experiment 4 extends these findings by looking further at the possible similarities in terms of representational structure and lexical performance for few- versus many-sense L1 words and for L2 versus L1 words.

4.4 Experiment 4

The priming asymmetry observed in Experiments 3a and 3b confirms the representational asymmetry assumed by the sense model to exist between (many-sense) L1 and (few-sense) L2 translation equivalents is essential in producing the priming asymmetry that has perplexed researchers for quite some time. The success of this manipulation suggests that it may be productive to look more closely at the possible similarities between few-sense L1 words and L2 words.

Traditionally, researchers working under the assumption that L1 and L2 forms are mediated by a common set of meaning representations have assumed that learners map L2 forms onto already-existing semantic representations such that both L1 and L2 forms come to have identical meanings. Under this (traditional) assumption, it would make no sense to look at the similarities between few-sense L1 words and L2 words since few-sense L1 words and their most closely related many-sense counterparts mean clearly different things and L2 words mean the same thing that their L1 counterparts do. But, as we saw in the Introduction, distributed meaning models, such as De Groot's distributed feature model and the sense model that I am proposing here, allow for L1 and L2 "equivalents" to have distinct, but overlapping, meanings. Under this view, the relation between L2 words and their L1 equivalents may, in fact, be very similar to the relation between few-sense words and closely related many-sense words.

According to the sense model, learners, when forming new L2 form-meaning connections, map new L2 forms on to only those senses of the L1 word that they think would exist in the L2. That is, L2 words are thought to instantiate only a circumscribed portion of the full range of meanings associated with their L1 equivalents. Quite interestingly, few-sense words with closely related many-sense words frequently serve the same purpose within a language. One of the general characteristics that differentiates few-sense words from their closely related many-sense words is that they are more specific in terms of their referents. It is as if the semantic sense representations of one-sense words are formed by extracting specific semantic properties from other many-sense words in order to create a finer-grained lexical distinction. Take for example the one-sense word "lawn". The meaning of "lawn" could be formed by taking from GRASS and YARD just those semantic properties that, when combined, refer to a specific type of grass that is grown in a specific area (namely, adjacent to people's houses).

If both few-sense L1 words and L2 words are acquired by extracting a circumscribed amount of semantic information from already-established semantic entries, then we may expect few-sense L1 words and L2 words to pattern similarly in lexical processing tasks. The present experiment tested this possibility by comparing lexical decision times for many-sense and few-sense L1 words. If few-sense L1 words and L2 words are represented (and processed) in similar ways, then we would expect slower reaction times and higher error rates for few-sense words.

4.4.1 Method

The present experiment was designed as a repetition priming experiment. The reason for using repetition priming instead of just lexical decision is that, in addition to comparing lexical decision times, I also wanted to compare few- and many-sense words in terms of repetition priming. An interaction would suggest that one word type (few-sense or many-sense) produced larger priming effects than the other. The design of a repetition priming experiment allowed me to make both reaction time and priming effect comparisons simultaneously.

4.4.1.1 Participants

Sixteen undergraduates at the University of Arizona participated for course credit.

4.4.1.2 Materials

When selecting items for the purpose of comparing reaction times, it is imperative that items are selected randomly to avoid inadvertent experimenter bias (Forster, 2000). With this in mind, the materials for the present experiment were selected in the following way. The frequencies (Kucera & Francis, 1967) of all of the one- and two-sense words (few-sense words) found in WordNet with a word length ranging from 4-6 were determined. This formed a list of possible few-sense words. Subsequently, the frequencies of all of the words with 15 or more senses (many-sense words) in WordNet with word lengths ranging from 4-6 were determined. This formed a separate list of possible many-sense items. The two lists were then sorted according to frequency. Eighty items appearing in a frequency band of 20-80 per million were then randomly selected from each list of possible words for a total of 160 experimental items.

An equal number of nonwords was generated by the ARC Nonword Database (Rastle, K., Harrington, J., & Coltheart, M., 2002). Control primes were generated by the MRC Psycholinguistic Database (Coltheart, 1981) and were matched with experimental items for frequency and word length.

4.4.1.3 Design and Procedure

The design of the present Experiment was identical to the previous priming experiments. Participants were asked to indicate whether the target word was a legal word in English or not. Items consisted of the following sequence of events. each one following immediately after the other: forward mask (########) for 500ms, prime in lowercase letters for 42ms, target in uppercase letters for 500ms. Two lists were constructed in order to counterbalance across the prime factor. Targets that were preceded by their identity prime on List A were preceded by a control prime on List B and vice versa. Counterbalancing across the prime factor in this way allows one to make accurate reaction time comparisons between word types because any influence of the prime is effectively removed when averaging across both prime conditions.

4.4.2 Results

The same trimming procedure used in Experiments 1-3 was similarly employed here. No participants were rejected for failing to meet the accuracy criterion of 80%. Looking first at the priming effects, there was an extremely robust repetition priming effect. Mean response times for targets preceded by their identity prime were 532 ms and 565 ms for targets preceded by a control prime. An ANOVA revealed that this main effect of priming was highly significant ($F1(1,14) = 73.53$, $P < 0.001$; $F2(1,78) = 37.63$, $P < 0.001$). Concerning the question of whether one word type exhibited greater priming than another, it was found that few-sense words exhibited a 40 ms repetition priming effect, and many-sense words exhibited a 28 ms effect. This slightly larger numerical effect for few-sense words was not found to be significantly different from the effect exhibited by many-sense words ($F1(1,14) = 1.13$, $P = 0.3$; $F2(1,78) = 2.7$, $P = 0.1$). This pattern of results suggests that many- and few-sense words are equally capable of producing repetition priming effects. This does not differ from what is found with bilinguals. Within-language repetition priming effects are found to be just as robust in L2 as they are in L1.

Turning now to the reaction time data, we see that many-sense words are responded to much faster than few-sense words. The mean response times for many-sense words were 533 ms and 564 ms for few-sense words, a difference that was found to be very reliable ($F1(1,14) = 14.44$, $P = 0.001$; $F2(1,78) = 36.54$, $P < 0.001$). This is quite a surprising result. Because these two sets of words were equated on frequency and word length, the best explanation for this difference has to do with the experimental manipulation. The more senses that are associated with a particular word form, the faster people are to say that it is a legal word. This pattern of results parallels what is frequently seen across languages. Response times for L1 targets are consistently faster than they are for L2 equivalents.

Looking at the error data, the parallel between the performance exhibited by participants with few-sense L1 words and L2 words becomes even more remarkable. In cross-language experiments, error rates associated with L2 stimuli are consistently higher than those associated with L1 stimuli. In the present experiment, the mean error rate for few-sense words was 13.8% and only 3.1% for many-sense words. This difference, just like the reaction time difference, was found to be very reliable ($F1(1,14) = 74.02$, $P <$

0.001; $F2(1,78) = 29.56$, $P < 0.001$). This particular difference is even more remarkable than the reaction time difference. Both sets of words were not only equated on frequency, but both sets were selected from a band of very high-frequency words (20-80 per million). Nevertheless, participants found it very difficult to respond to few-sense words. Participants took longer to determine that few-sense words were legal letter strings in English, and were more prone to be wrong once they finally did make a decision, despite the fact that these words appear in their environment just as often as the many-sense words.

4.4.3 Discussion

The present experiment was motivated by the notion that the semantic representations of few-sense L1 words and L2 words may be established in memory and, hence, processed in largely the same ways. Though it would be terribly difficult to provide conclusive evidence for this, the similarities between few-sense L1 words and L2 words are remarkable in terms of how they are processed relative to many-sense L1 words. The results of the present experiment reveal that few-sense words are responded to much more slowly in a lexical decision task relative to frequency- and length-matched many-sense control words. Furthermore, the error rates were significantly higher for few-sense words than they were for many-sense controls. This pattern of results is remarkably similar to the pattern of results typically observed for L2 words and L1 words in cross-language experiments.

4.5 General Discussion

The present set of experiments serves to lay the ground work for the development of a plausible explanation for a very important and hitherto perplexing issue in bilingual lexical processing: the masked translation priming asymmetry. The nature of this priming asymmetry is that masked primes in the dominant language (L1) prime translation equivalent words in the non-dominant language (L2), but the same does not occur in the reverse direction, unless the prime and target words are cognates from same-script languages (e.g., rich-rico) (Sanchez-Casas, Davis & Garcia-Albea, 1992).

In developing an account for this asymmetry, several things must be taken into consideration. First, within-language masked repetition priming effects are very robust in both L1 and L2. Second, the priming asymmetry is task specific; that is, L2-L1 masked priming can be observed when semantic categorization is used but not lexical decision. Third, the priming asymmetry can be recreated with within-language word pairs as long as those word pairs share a semantic sense and as long as there is a representational asymmetry between the two such that one is a many-sense word and one is a one-sense word. And finally, response times to many-sense words are faster and more accurate.

The fact that within-language masked repetition priming is robust in both L1 and L2 suggests that trying to account for the asymmetry by positing that L2 form-meaning connection-strengths are too weak to drive priming cannot be correct. Because this constitutes the explanation given by the RHM for the translation priming asymmetry, it is not clear how this model could provide a coherent account of both within- and cross-language priming effects. (The other models reviewed in this thesis are effectively excluded because they predict symmetrical priming effects across languages.) The sense model proposed here may provide a workable account of these findings.

The sense model does not assume differences in connection strengths, and so it is able to account easily for robust within-language priming effects. The difference between L1 and L2 words lies in the number of semantic senses associated with each. The sense model assumes that L1 words, on average, have many semantic senses associated with them relative to L2 words, which, on average, have dramatically fewer associated senses. This representational asymmetry is the consequence of how L2 words are thought to be learned and established in memory. It is assumed that learners establish new L2 form-meaning connections by mapping their newly learned L2 form onto a circumscribed amount of semantic information extracted from the already-established semantic entry of the L1 equivalent. This results in a representational asymmetry where all of the senses associated with an L2 form are also associated with the L1 equivalent form, but where only very few of the senses associated with an L1 form are similarly associated with the L2 equivalent form.

The priming asymmetry "falls out" of this representational asymmetry quite naturally. Insofar as priming can be attributed to semantic overlap, the amount of priming that occurs is assumed to depend on the proportion of primed to unprimed senses. In the L1-L2 direction, the proportion of L2 senses primed by the L1 prime is going to be very high, if not complete. This is because most, if not all, of the senses associated with the L2 form are also associated with the L1 equivalent form. Priming in this direction, then, should be very reliable, and it is. In the L2-L1 direction, the proportion of L1 senses primed by the L2 prime is going to be very low (e.g. 1:15). This is because there are many senses associated with the L1 form that are not similarly associated with the L2 prime. Accordingly, priming in this direction should be very weak, and it is.

At this point it is necessary to consider in greater detail how the sense model could account for asymmetrical priming effects both within and across languages. So far, I have described priming as being attributed to the proportion of overlapping senses between prime and target. One way to characterize this is to think of form entries as being associated with a distributed set of semantic features and that a word's different senses are attributed to different patterns of activation across these features. According to this characterization, many-sense (L1) words would be associated with many different patterns and one-sense (L2) words would be associated with just a few. Characterizing the representational asymmetry in this way appears promising at first glance because, for one, the response time advantage for many-sense (L1) words as well as the proportional priming effects follow from this type of model rather easily. For the faster response times associated with many-sense words, all that one would have to do is say that the orthographic level of representation receives feedback from the semantic level and that the feedback received from many-sense words is much greater, leading to faster lexical decision times (c.f. Rodd, Gaskell & Marslen-Wilson, 2002, for a similar argument). As for the priming asymmetry, one would simply say that an L2 prime can only preactivate a small proportion of the semantic features associated with a many-sense (L1) word, which is insufficient to noticeably facilitate decision times. In the many-one (L1-L2) direction, on the other hand, priming is observed because the proportion of the target's features preactivated by the prime is 100% – due to all of these features also being associated with the L1 form.

Though it may be possible to account for the present set of findings with an activation model like the one described here, there are at least two problems that this type of account faces. The first comes in trying to account for why there is no interaction in repetition priming for many-sense versus few-sense words. If decisions are generated on the basis of feedback from the semantic level to the orthographic level, and there is more feedback from the semantic level for many-sense words relative to few-sense words, then many-sense words should exhibit larger repetition priming effects. Activation accrues much faster when there are multiple inputs (as is the case with many sense words) than when there are few inputs. Following from this, preactivating, say, 100 features for a many-sense word versus 10 features for a few-sense word, should lead to a disproportionate amount of priming for the many-sense word. But it does not. One way to compensate for this would be to vary the threshold at the orthographic level for each word, saying that many-sense words need more feedback from the semantic level to generate a decision than do few-sense words. This would account for the lack of an interaction, but then it would be hard to explain why many-sense words are responded to faster.

A second problem that an activation model of this type would face has to do with the features themselves. If each semantic feature is allowed to be an input, whether that is in the form of feedback to the orthographic level or to a decision node, one should be able to detect congruence effects in semantic categorization. But as we saw above, two separate experiments using ANIMAL as the category found no difference in categorization times between congruent items, such as *shark-ROBIN* and incongruent items, such as *badge-ROBIN* and no difference in priming effects between *dolphin-WHALE* when the control condition used non-exemplars as primes (e.g. *helmet-WHALE*) or congruent, but semantically dissimilar, exemplars as primes (e.g. *sparrow-WHALE*). Presumably, if each feature is allowed to be an input, the set of features representing the properties of "animalness" would have had an effect. In sum, casting the sense model as an activation model is very promising in many respects, but it is not clear that characterizing it in this way is the most "trouble free".

Another way to characterize the sense model is to say that each form representation is associated with a set of semantic features or arguments, and that a semantic sense is, essentially, a function that operates on a specific set of the semantic arguments. According to this characterization, an individual is able to make a decision on a stimulus as soon as one of the functions (senses) has been assembled. This metaphor is taken from computer programming. According to this metaphor, more than one function may operate on any particular argument. This is important because it allows some arguments (semantic features) to be present in many, if not all, of a word's senses, constituting the "core" meaning of the word's many senses without predicting congruence effects. Words with one sense will only have one function (i.e. that word's semantic sense) associated with it, and it will operate on just a few select arguments.

Words with many senses, on the other hand, will have many functions, several of which will operate on the same arguments.

According to this characterization, priming is achieved when the prime pre-assembles the target's sense(s). This, of course, could only happen if the prime and target share senses (i.e. functions). In the many-one direction, this is effective because the target has only sense associated with it that comes to bear on generating a decision, and it is shared with the many-sense (L1) word. In those cases when this sense has been pre-assembled by the prime, a decision can be generated straight away (sense assembly does not need to occur) and priming will be observed. In the one-many direction, the prime similarly serves to pre-assemble the target's sense that is common to both the prime and target. Again, according to this particular characterization, a decision can be generated for a target as soon as one of its senses has been assembled. This is problematic, of course, given how we do not observe priming in this direction, and so characterizing the priming asymmetry in this way requires one to assume two things: first, senses must be assembled in parallel, and two, there must be a rather large random component in terms of how senses get assembled. A consequence of this random component is that there will always be a "first" and a "last" sense to be assembled, the order of which would be inconsistent from instance to instance. These two assumptions predict faster (unprimed) reaction times for many sense words relative to few-sense words, because the greater the number of senses to be assembled, the faster the fastest one to be assembled will be. It is not clear, though, that the first sense of the target to be assembled could be assembled before assembly of the sense associated with the prime could be completed.

Of course, the parameters of the model could be set so as to ensure that, on average, the fastest target sense to be assembled would "beat out" the assembly time of the primed sense. All one would have to do is increase the magnitude of the random component in sense assembly; but it would come at a cost. Although increasing the random component in sense assembly procedures would prevent one-many priming, and may even generate a reaction time effect of the magnitude seen in Experiment 4 for many-sense words (approximately 30 ms), it would concomitantly predict higher error rates for many-sense words. Yet the empirical data point out rather clearly that subjects produce very few errors with many sense words (3.1%) relative to few-sense words (13.8%), a difference that is found to be very reliable statistically.

In sum, the sense-assembly characterization of the sense model is promising in accounting for the pattern of findings observed in this study, but it, at this point, is still far from being able to provide a complete account. The largest hurdle facing this particular way of constructing the sense model is the error data. According to the sense-assembly hypothesis, a semantic sense is a function that operates on a specific set of the semantic arguments. In the case of many-sense words, several different functions are all operating on essentially the same semantic arguments. This, technically, should lead to slower processing times with higher error rates being associated with many-sense words because the potential for ambiguity is greater with these words. Yet, as Experiment 4 made clear, subjects are faster and more accurate when making decisions on many-sense words. It is not clear at this point if the sense-assembly hypothesis would be able to account for how the error data pattern.

A third possibility, and the last one to be discussed here, presents a slight variation of the sense-assembly hypothesis discussed above. Given that the potential for ambiguity (and, consequently, "throwing an error") is high for many-sense words, it may be that the system is compelled to restructure itself in these cases. That is, when a particular "neighborhood" (i.e. semantic entry) of arguments and functions becomes too dense, and the likelihood of producing errors becomes too high, a restructuring may occur so as to reduce that particular neighborhood's density. One way in which this restructuring could occur is to "encapsulate" the functions with their arguments, which would reduce the need for online computation of each sense. Essentially, the result of this restructuring would be such that each semantic sense would be stored as a distinct "value". This, of course, is the last option for the system to entertain because storing semantic senses as representations takes up much more in terms of resources than does computing those same senses by assembling a set of arguments. The motivation for *not* restructuring a semantic entry is only overcome, then, when the degree of ambiguity within a semantic entry, and the likelihood of generating errors during sense computation, becomes too great. In other words, semantic senses for few-sense words would be computed online because the neighborhoods of few-sense words are rather sparse.

Assuming that semantic senses of many- and few-sense words are represented in such fundamentally different ways imposes, at the very outset, a great theoretical cost because it is clearly not the most parsimonious way to go about specifying the mechanisms of the sense model; but it may be these assumptions are sufficiently well motivated (similar "restructuring arguments" have been made for densely populated orthographic neighborhoods). Furthermore, it may be that beginning with these assumptions in hand allows us to account for the pattern of findings in question.

First, by assuming that the senses of many-sense words are stored representations instead of functions that need to assemble a set of arguments (as is the case for few-sense words), the unprimed reaction time and error data are easily accounted for. If a decision can be generated as soon as a sense for that word is either assembled or activated, then decision times are clearly going to be faster when all that is involved is activating a particular representation. Likewise, error rates should be lower when activation of a particular representation is all that is involved compared to when a set of computational procedures needs to be carried out. Reaction times for few-sense words will be expected to be slower as a result of having to compute its semantic sense(s), and error rates should be higher because the procedure of sense-assembly is more susceptible to error.

With respect to the priming asymmetry in lexical decision, this account is largely the same as that of the sense-assembly account. In the many-one direction, there is only one sense associated with the target that can come to bear on lexical decision times, and when it is primed decisions can be generated faster. In the one-many direction, priming is very unlikely because, if we assume a random component in the parallel activation of the target's many senses, the likelihood that the first sense to be activated is going to be the sense that the prime preassembled is very low. In this latter case, because activation times are so much faster than assembly times, it is very likely that the sense of the target to be activated first is able to trigger a decision before assembly of the primed sense is completed.

Although the third possibility discussed here appears to be best equipped to account for the complete pattern of findings reported in this study, it may be the least desirable theoretically because it is clearly the least parsimonious. It remains to be seen whether this latter set of assumptions can be better motivated theoretically, or if one of the other mechanisms proposed above can work out a solution to account for the complete pattern of findings.

In conclusion, the present set of experiments has served to form the basis of a plausible explanation for particularly troubling findings in the bilingual lexical processing literature (i.e. those constituting the masked translation priming asymmetry). First, the priming asymmetry was argued to be the logical consequence of the representational asymmetry assumed by the sense model. In the L1-L2 case, priming is expected because all of the senses associated with the L2 form are also associated with the L1 equivalent form. The proportion of L2 senses primed by the L1 prime, then, is very high, if not complete, when priming is in this direction. In the L2-L1 case, though, the sense model assumes that there are, on average, many senses associated with L1 forms, many of which are not similarly associated with the L2 equivalent form. Following from this, priming in this direction should be very weak because the proportion of L1 senses primed by the L2 prime is very low. It is clear, then, that the sense model straightforwardly predicts the priming asymmetry in lexical decision.

The present study revealed, though, that the masked translation priming asymmetry is task specific. Masked L2-L1 priming does not occur in lexical decision, but reliable effects are found when semantic categorization is used. Though this particular task-specificity was not predicted, it was shown that the sense model is able to account for it very straightforwardly. Assuming that L2-L1 priming does not occur because an insufficient number of L1 senses are preactivated by the L2 prime, then it must be the case that the semantic categorization task serves to restrict the number of senses that can come to bear on generating a decision. It was argued that in semantic categorization only those semantic senses that are relevant to the category are taken into account. This suggests that an L2 prime will be more effective, because the only sense that is relevant for the decision will be activated by the prime (because, presumably, this is the very sense of the word that is shared between L1 and L2).

The assumption that the semantic categorization task restricts the number of senses recruited when the subject generates a decision (allowing the L2 prime to be effective) was supported in a within-language experiment. In this all-English experiment, semantic priming was observed when word pairs were carefully selected such that each pair shared one sense in common and consisted of a many-sense word and a 1-sense word. When priming was in the many-one direction, robust priming effects were observed. When priming was in the one-many direction, no priming was observed. This experiment convincingly shows that, though semantic similarity is necessary to produce priming, it is not sufficient. In order for masked translation (or masked within-language semantic) priming to occur, the targets must have a reduced number of senses. Following from this, I argued that what the semantic categorization task did was turn many-sense targets into one-sense targets, thereby allowing the prime to be more effective. Taken together, the set of experiments presented in this study provide very strong support for the basic tenets of the sense model. As we saw above, though, preliminary efforts to develop a highly specified computational model capable of producing the necessary pattern of effects have not yet met with much success.

Chapter 5 General Discussion

How do bilinguals access meaning? How do L2 lexical semantic representations differ from their L1 counterparts? Or do they? Is it possible to develop a general theory of adult vocabulary learning that is capable of accounting for both L1 and L2 adult word learning? These are the questions that have motivated the present set of studies.

Because it is well beyond the scope of this work to pursue all aspects of word learning, I have focused on just one level of this process: namely, how meaning comes to be represented for words learned in adulthood and how those meaning representations come to be associated with an appropriate form-level representation. A separate line of work could have focused, for example, on how form-level information comes to be established in memory and subsequently utilized.

Previous work on bilingual lexical processing has been interpreted as providing support for one of three different classes of models: distinct meaning, common meaning and distributed meaning models. Early on, there was some initial support for the "distinct" hypothesis. According to this hypothesis, learners represented L1 and L2 lexical items distinctly, at both the level of form and meaning. This conclusion soon had to be modified when it was found that it was possible to observe cross-language priming with short SOAs (Chen & Ng, 1989; De Groot & Nas, 1991; Keatley, Spinks, & De Gelder, 1994; Kirsner et al., 1984; Meyer & ruddy, 1974; Schwanenflugel & Rey, 1986). A later study by Williams (1994) effectively showed that the locus of this cross-language priming effect was semantic.

This pattern of findings, along with the work of Potter et al. (1984), who looked at picture-naming and translation performance, led to the development of the common-meaning class of models. According to this class of models, meaning-level representations were thought to be represented in a common store. These commonly represented semantic representations were considered to be capable of subserving both L1 and L2 equivalent form representations, which were represented in distinct lexicons. This class of models has been very successful in accounting for much of the bilingual lexical processing data.

With respect to adult L2 word learning, the common-meaning class of models suggests that learners map newly established L2 forms onto already-existing meaning-level representations. In other words, the semantic representations of L1 and L2 words are one in the same according to how the common-meaning models are specified. This simplicity, though appealing in terms of theory construction, has led to several difficulties for models within this class. First and foremost is the very real problem that bilinguals are quick to point out: there are very few true translation equivalents. A clear example of this was provided by looking at the meanings of the word "black" in English and Japanese. This example showed very nicely how translation equivalency could not be assumed even for very common, relatively concrete and completely unambiguous words like "black". For example, "black" can be used to refer to a type of humor, as well as a calamitous day on Wall Street (*black Friday*), while "kuroi", the Japanese translation of "black", can be used to refer to those who are evil-minded (*black belly*), as well as those who are well tanned or guilty of a crime. Given how these two words can be used language-specifically to mean such different things, it does not logically follow to think that "black" and "kuroi" map onto a single semantic representation. This is a very real problem for the common-meaning models.

Common-meaning models have also had difficulty in accounting for the empirical data found in some of the cross-language lexical processing studies. For example, the concept-mediation model would have to predict symmetrical priming effects in masked translation priming, yet there is a clear asymmetry. Proponents of the RHM, another common-meaning model, have attempted to show how the RHM could account for the priming asymmetry, but this attempt has been largely unsuccessful (see the introduction of Chapter 4).

A third class of models, the distributed-meaning class of models, has met with more success in terms of allowing for translation equivalent words to have language-specific meanings. According to this class of models, meaning is distributed across a series of "meaning elements", some of which are shared across languages and some of which are not. The distributed-feature model (De Groot, 1992) proposes that these "elements" are semantic features, and the sense model proposes that these "elements" are semantic senses. According to the distributed feature model, a particular form is associated with a specific set of semantic features. The degree to which the sets of semantic features associated with L1 and L2 translation-equivalent forms overlap is constrained by the word type. De Groot argues that concrete words share more semantic features across languages than do abstract words.

One aspect of the distributed feature model that is very appealing is its ability to account for language specific meanings. That is, translation equivalent words can mean almost the same exact thing (characterized by a high degree of featural overlap), or very different things (characterized by a very low degree of featural overlap). But this model fails to capture the full range of language-specific meanings that can exist in addition to a common meaning between translation equivalents. For example, it is not clear how this model could allow for translation equivalents like "black – *kuroi*" having language-specific meanings in addition to the one that they share. Again, according to the distributed feature model, "black" would be associated with a particular set of features and "kuroi" would be associated with a different set of semantic features. Because these two words are fairly concrete, several semantic features would presumably be shared between the two sets of features. In other words, "black" and "kuroi" would mean almost the same things according to the distributed feature model; but there is nothing in the model to allow for "black" to mean a particular color in both English and Japanese *as well as* a type of humor in English but not Japanese.

A separate problem for the distributed feature model is its inability to account for the masked translation priming asymmetry. Although this model predicts varying degrees of semantic overlap between translation equivalents depending upon word type, the degree of overlap for any given translation pair is symmetrical. Accordingly, this model would predict symmetrical priming effects across languages, which is clearly not the case.

The sense model, by assuming that a particular word's meaning is comprised of distinct senses (one or more of which may be shared across languages), is able to allow for the words in a translation equivalent pair to have language-specific meanings in addition to sharing a particular sense across languages. According to this model, learners map newly established L2 forms onto an appropriate semantic sense extracted from the semantic entry of the translation equivalent L1 word. For example, an English speaker learning Japanese would establish an L2 form-meaning connection between the form "kuroi" and the COLOR sense extracted from the semantic entry of "black". There is no reason to think that this new L2 form-meaning connection would include the HUMOR sense that exists in the learner's L1 (c.f. Kellerman's (1978) study with Dutch learners of English assuming that only the core sense of "break" existed in English).

Because L2 learners are going to be more familiar, on average, with the range of meanings that their L1 words encode compared to their L2 words, a representational asymmetry is assumed. There are two aspects of this representational asymmetry. The first is a simple numerical asymmetry. That is, more senses on average will be associated with L1 forms than L2 forms. The second has to do with the nature of the representations themselves. Typically, each sense associated with the newly learned L2 form will also be associated with the equivalent L1 form. The same is not true when considering the senses associated with the L1 form. Many of the senses associated with any given L1 form will *not* be similarly associated with the equivalent L2 form. This second aspect of the representational asymmetry suggests that, typically speaking, L2 words are highly specified L1 words (in terms of their semantics).

This latter point constitutes a very exciting feature of the sense model. The sense model assumes that learning a new L2 word is essentially the same as learning a new term for a specific meaning (sense) of an already-known word. In this respect, adult L2 word learning could be very similar to adult L1 word learning.

Typically, adult L1 word learning involves learning new terms that allow for a greater degree of specificity in communication. Take for example the word "ground". This is a frequently used word that has a relatively low age of acquisition rating (suggesting that it is learned early on in life). Because this form has several different senses associated with it, using it may give rise to an unintentional ambiguity as in "He stood on good ground." This may mean that the position he took on a particular issue was well justified or supported; or it may mean that the soil he stood on was of a good quality. In order to avoid such ambiguities, it behooves the language user to learn more vocabulary so as to allow for a greater deal of specificity. For example, if the latter meaning of the above sentence was intended, then it would be less ambiguous to say, "He stood on good soil." Not surprisingly, words like "soil", which serve to refer to a circumscribed meaning (sense) of other potentially ambiguous words, are learned later on in life and tend to encode fewer senses themselves. It is not surprising, then, to see that the age of acquisition rating for "soil" suggests that it is learned much later on in life than "ground", and that, according to WordNet, has many fewer senses (4) associated with it than does "ground".

I am not suggesting that *all* of L2 word learning can be compared to adult L1 word learning. First, the difficulty associated with learning new L2 forms is going to be much greater than what is involved with learning new L1 forms. And even at the level of semantic representations, there are differences as well. There are many instances where learners will establish new L2 form-meaning connections for words that are translations of one-sense L1 words. In this case, the one sense of the L1 word would be copied into the L2 lexical entry, constituting, perhaps, true translation equivalency – as long as other L2 language-specific senses were not encoded as well.

Even though it is clear that not all of L2 word learning can be said to parallel adult L1 word learning, it is intriguing to note how well this comparison works for the "typical case" both logically and empirically. As we saw in Chapter 4, when all-English word pairs were constructed so as to reproduce the representational asymmetry assumed by the sense model, performance resembled that of bilinguals. Words with many senses (e.g. "air") were responded to faster and with fewer errors relative to frequency-matched words with few senses (e.g. "oxygen"). This pattern of performance is remarkably similar to that of bilinguals with L1 words versus L2 words. Bilinguals typically respond faster and with fewer errors to L1 words than they do to L2 words.

It was also found that these all-English word pairs produced the priming asymmetry found between L1 and L2 words. That is, many-sense words (thought to be the representational equivalent of L1 words) primed their few-sense counterparts (the representational equivalent of L2 words) in masked priming with lexical decision, but when the direction of priming was reversed, the same few-sense words did not prime their many-sense counterparts. This again resembles the pattern of performance exhibited by bilinguals. In masked translation priming, L1 words prime L2 equivalents, but those same L2 words do not prime their L1 equivalents.

The sense model offers a new way to think of bilingual lexical representation and processing. By assuming that learners establish direct L2 form-meaning connections, the model is able to account easily for the semantic interference effects exhibited by novice learners in the earliest stages of L2 development, which is something that has proven difficult for other models. By assuming a representational asymmetry, the model is able to account for asymmetrical performance in translation and masked priming in a straightforward way. And perhaps the most intriguing aspect of this model lies in the parallels that it draws between L2 vocabulary learning and adult L1 vocabulary learning, suggesting that it may be possible to provide a general account of adult vocabulary learning. As we saw in Chapter 4, though, there is still much to be done before the assumptions of the sense model can be implemented into a highly specified and workable computational model.

Appendix A Experimental items from Experiment 2.1

Below, the stimuli used in Experiment 2.1 are presented along with the novel "L2" forms and their rank by frequency of the concept in each category (according to Battig & Montague, 1969).

	Items	Ranking	L2 word
animals	cat	2	birk
	cow	4	gorp
	dog	1	floop
	elephant	7	glip
	horse	3	larpell
	lion	5	treffim
	pig	8	ploozette
	tiger	6	walloon
Kitchen Utensil	bowl	9	fonteen
	cup	11	blikeet
	fork	3	dopal
	frying pan	15	tilkoon
	knife	1	dax
	pot	5	plap
	spoon	2	fremm
	stove	8	cald
Furniture	bed	3	carm
	chair	1	lig
	couch	7	vid
	desk	5	zek
	dresser	8	fozzeel
	lamp	6	soudleen
	table	2	beleem
	television	9	chabouk
Body Part	ear	8	detalle
	eye	4	ecrus
	foot	5	fossay
	hair	13	jacoll
	hand	9	milg
	leg	1	pelf
	nose	6	roop
	toe	10	yume

Appendix B Instructions for Experiment 3.3

During the training session, you will see a picture and then you will see the "Alien" word for that picture. You will also hear the alien word pronounced. After you hear the word, please pronounce it yourself two times. Your voice will be recorded so that the experimenter can check all of your vocal responses later. Remember, it is very important that you memorize the alien words.

All of the pictures that you see will be familiar to you. You will not need to guess what the picture is of. Some of the pictures are from our culture, some are from the alien culture. The alien pictures are different in color and texture from what you are used to.

Appendix C Experimental items from Experiment 4.3

one sense	many sense	# of senses for many sense words	LSA values	1-many association strengths	many-1 association strengths
oxygen	AIR	20	0.31	0.45	0.14
moist	WET	8	0.53	0.54	0
movie	FILM	7	0.65	0.19	0.54
sofa	CHAIR	7	0.73	0.13	0.07
surgeon	DOCTOR	7	0.64	0.47	0.04
oar	PADDLE	10	0.62	0.08	0.1
tomb	GRAVE	8	0.37	0.24	0.01
violin	BASS	9	0.43	0.01	0.02
clergy	MINISTER	6	0.39	0	0
elderly	AGED	10	0.67	0	0
wealthy	RICH	12	0.61	0	0
coyote	FOX	9	0.48	0.02	0
lantern	LIGHT	30	0.52	0	0
huge	LARGE	12	0.56	0.2	0.02
lawn	GRASS	10	0.47	0.39	0.09
wheat	GRAIN	11	0.72	0.03	0.36
lung	HEART	10	0.33	0.13	0.04
cottage	HOUSE	13	0.42	0.38	0
dairy	MILK	7	0.68	0.49	0.01
poverty	POOR	9	0.71	0.54	0.02
saucepan	POT	11	0.53	0	0
fist	HAND	15	0.47	0.23	0.01
pistol	GUN	8	0.62	0.77	0.06
prince	KING	6	0.61	0.13	0.01
beer	DRINK	11	0.84	0.17	0.01
fabric	MATERIAL	12	0.36	0.16	0.08
lawyer	JUDGE	7	0.74	0.01	0.01
tiny	SMALL	17	0.54	0.65	0.08
pebble	STONE	10	0.36	0.3	0.03
marina	DOCK	11	0		0
wrist	ARM	8	0.73	0.15	0
ankle	FOOT	12	0.54	0.36	0
lengthy	LONG	15	0.22		0
puppy	DOG	7	0.76	0.75	0.12
kitten	CAT	9	0.61	0.78	0.16
gale	WIND	16	0.76		0
suitcase	BAG	15	0.51	0.13	0
creek	STREAM	10	0.43	0.05	0.04
mosquito	FLY	20	0.34	0.05	0.04
rapidly	FAST	15	0.37		0
	means	**11.25**	**0.529**	**0.289**	**0.084**

REFERENCES

Altarriba, J. & Mathis, K. (1997). Conceptual and lexical development in second language acquisition. Journal of Memory and Language, 36, 550-568.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 340-357.

Basden, B. H., Bonilla-Meeks, J. L., & Basden, D. R. (1994). Cross-language priming in word-fragment completion. Journal of Memory and Language, 33, 69-82.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80, 1-46.

Bousfield, W.A. (1953). The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 49, 229-240.

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition, 3*(2), 216-220.

Bradley, D. C., & Forster, K. I. (1987). A reader's view of listening. Cognition Special Issue: Spoken word recognition, 25(1-2), 103-134.

Brown, H.D. (1992). Vistas. Englewood Cliffs, New Jersey: Prentice-Hall.

Catchpole, J. (1987). The locus of the word frequency effect. Unpublished honors thesis, Department of Psychology, Monash University.

Chen, H. -C., & Ng, M. -L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. Memory and Cognition, 17, 454-462.

Chen, H. -C.,H.-C., & Ho, C. (1986). Development of stroop interference in Chinese-English bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 397-401.

Chen, H.-c., & Leung, Y.-s.H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a nonnative language. Journal of Experimental Psychology: Learning, Memory, & Cognition, 15(2), 316-325.

Cofer, C.N. (1966). Some evidence for coding processes derived from clustering in free recall. Journal of Verbal Learning & Verbal Behavior, 5, 188-192.

Cohen, B.H. (1963). Recall of categorized word lists. Journal of Experimental Psychology, 66, 227-234.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. Psychology Review, 82, 407-428.

Damian, M.F., Vigliocco, G. & Levelt, W.J.M. (2001). Effects of semantic context in the naming of pictures and words. Cognition, 81, 77-86.

Davis, C. W., Kim, J., & Sanchez-Casas, R. M. (in press). Masked priming across languages: An insight into bilingual lexical processing. To appear in S. Kinoshita & S. Lupker (Eds.), Masked Priming: State of the Art.

De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. Journal of Memory and Language, 30, 90-123.

De Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. Language Learning, 47(2), 215-264.

De Groot, A. M. B., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. Journal of Memory & Language, 33(5), 600-629.

De Groot, A.M.B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), Orthography, phonology, morphology, and meaning. (pp. 389-412). Amsterdam: North-Holland.

De Groot, A.M.B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), The bilingual lexicon (pp 27-51). Amsterdam: John Benjamins.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Durgunoglu, A. Y., & Roediger, H.L., (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language, 26*, 377-391.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. Journal of Experimental Psychology: Human Experimental Psychology, 46, 533-558.

Ervin, S. and Osgood, C.E. (1965). Second language learning and bilingualism. In C.E. Osgood and T.A. Sebeok (Eds.), Psycholinguistics: A survey of theory and research problems. Bloomington, IN: Indiana University Press.

Finkbeiner, M., & Nicol, J. (2000). Early L2 lexical representation: An episodic memory explanation. Paper presented at the Second Language Research Forum, Madison, WI.

Forster, K. I., Mohan, K., & Hector, J. (in press). The mechanics of masked priming. To appear in S. Kinoshita & S. Lupker (Eds.), Masked Priming: State of the Art.

Fox, .E. (1996). Cross-language priming from ignored words: Evidence for a common representational system in bilinguals. *Journal of Memory and Language, 35*, 353-370.

Frenck, C., & Pynte, J. (1987). Semantic representation and surface forms: A look a across-language priming in bilinguals. Journal of Psycholinguistic Research, 16, 383-396.

Friedrich, J. F., Henik, A., & Tzelgov, J. (1991). Automatic processes in lexical access and spreading activation. Journal of Experimental Psychology: Human Perception and Performance, 17, 292-306.

Gairns, R., & Redman, S. (1986). Working with words: A guide to teaching and learning vocabulary. New York: Cambridge University Press.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 133-177). New York: Academic Press.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), Language: Psychological and biological aspects (pp. 69-96 Series Linguistics The Cambridge survey, Vol 63). New York, NY: Cambridge University Press.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Journal of Memory and Language, 28, 200-213.

Gekoski, W. L., (1980). Language acquisition context and language organization in bilinguals. Journal of Psycholinguistic Research, 28, 429-449.

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 305-315.

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

Gollan, T., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. Journal of Experimental Psychology: Learning, Memory and Cognition, 23, 1122-1139.

Graf, P., Mandler, G., & Haden, P. (1982). Simulating amnesic symptoms in normal subjects. Science, 218, 1243-1244.

Grainger, J., & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. Language and Cognitive Processes, 13, 601-623.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. Brain & Language, 27(2), 210-223.

Green, D. W. (1993). Towards a model of L2 comprehension and production. In R. Schreuder & B. Weltens (Eds.), The bilingual lexicon (pp. 249-277). Amsterdam. John Benjamins Publishing Company.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. Bilingualism: Language & Cognition, 1(2), 67-81.

Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. Journal of Verbal Learning and Verbal Behavior, 2, 170-175.

Jackendoff, R. (1997). The Architecture of the Language Faculty. Cambridge, MA: MIT Press.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. Bilingualism: Language and Cognition, 2, 59-75.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*, 47-77.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. Journal of Memory and Language, 44, 32-51.

Johnson, C. J. (1995). Effects of color on children's naming of pictures. *Perceptual & Motor Skills, 80*, 1091-1101.

Keatly, C. W., Spinks, J. A., & De Gelder, B. (1994). Asymmetrical cross-language priming effects. Memory and cognition, 22, 70-84.

Kellerman,E.(1978).Giving learners a break: Native language intuitions as a source of predictions about transferability.Working Papers on Bilingualism,15,59-92.

Kintsch, W., & Kintsch, E. (1969). Interlingual interference and memory processes. Journal of Verbal Learning and Verbal Behavior, 8, 16-19.

Kirsner, K., & Smith, M. C. (1974). Modality effects in word identification. Memory & Cognition, 2, 637-640.

Kirsner, K., Smith, M.C., Lockhart, R.S., King, M.L., & Jain, M. (1984). The bilingual lexicon: language-specific units in an integrated network. Journal of Verbal Learning and Verbal Behavior, 23, 519-539.

Kolers, P. (1963). Interlingual word associations. Journal of Verbal Learning and Verbal Behavior, 2, 291-300.

Kolers, P. (1966). Interlingual facilitation of short-term memory. Journal of Verbal Learning and Verbal Behavior, 5, 314-319.

Kroll, J. F. (1993). Accessing conceptual representation for words in a second language. In R. Schreuder & B. Weltens (Eds.), The bilingual lexicon (pp 53-81). Amsterdam: John Benjamins.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second languages words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), Practical aspects of memory (Vol. 2). London: Wiley.

Kroll, J. F., & Sholl, A. (1992). Lexical and conceptual memory in fluent and nonfluent bilinguals. In R. J. Harris (Ed.), Cognitive processing in bilinguals (pp. 191-204; Amsterdam: Elsevier.

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

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J.L. Nicol (Ed.), One mind, two languages: Bilingual language processing (pp. 49-71). Malden, MA: Blackwell Publishers.

Kroll, J. F., Michael, E., & Sankaranarayanan, A. (1998). A model of bilingual representation and its implications for second language acquisition. In A. F. Healy, & L. E. Bourne, Jr. (Eds.), Foreign language learning: Psycholinguistic studies on training and retention (pp. 365-395). Mahwah, NJ: Lawrence Erlbaum Associates.

La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. Journal of Memory and Language, 35, 648-665.

Lambert, W.E., Havelka, J. & Crosby, C. (1958). The influence of language acquisition context on bilingualism. Journal of Abnormal & Social Psychology, 56, 239-244.

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

Levelt, W.J.M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22, 1-75.

Lewis, M.Q. (1971). Categorized lists and cued recall. Journal of Experimental Psychology, 87, 129-131.

Maegiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 304-315.

Meuter, R.F.I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. Journal of Memory and Language, 40, 25-40.

Molinsky, S.J., & Bliss, B. (1996). Expressways. Englewood Cliffs, New Jersey: Prentice-Hall.

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. TESOL Journal, 9, 6-10.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In Besner D., & Humphreys, G. (Eds.), Basic processes in reading: Visual word recognition (pp. 264-336). Hillsdale, NJ: Erlbaum.

Neuman, E., McCloskey, M., & Felio, A. C. (1999). Cross-language positive priming disappears, negative priming does not: Evidence for two sources of selective inhibition. Memory and Cognition, 27, 1051-1063.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), Information processing and cognition: The Loyola symposium (pp. 55-85). Hillsdale, NJ: Erlbaum.

Potter, So, Von Eckardt, & Feldman (1984). Lexical and conceptual representation in beginning and proficient bilinguals. Journal of Verbal Learning and Verbal Behavior, 23, 23-38.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. Journal of Memory and Langauge, 46, 245-266.

Roelofs, A. (2000). Attention to action: Securing task-relevant control in spoken word production. In L. R. Gleitman & A. K. Joshi (Eds.), Proceedings of the 22nd Annual Conference of the Cognitive Science Society, pp. 411-416. Mahwah, NJ: Erlbaum.

Sanchez-Casas, R. M., Davis, C. W., & Garcia-Albea J. E. (1992). Bilingual lexical processing: Exploring the cognate/noncognate distinction. European Journal of Cognitive Psychology, 4, 293-310.

Scarborough, D.L., Gerard, L., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. Journal of Verbal Learning and Verbal Behavior, 23, 84-99.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), Vocabulary: Description, Acquisition and Pedagogy. (pp.199-227). Cambridge: Cambridge University Press.

Schreuder, R., & Flores d'Arcais, G.B. (1989). In W. Marslen-Wilson (Ed.), Lexical representation and process. (pp. 409-436). Cambridge, MA: MIT Press.

Schriefers, H., Meyer, A.S., & Levelt, W.J.M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory and Language, 29, 86-102.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. Journal of Memory and Language, 25, 605-618.

Seal, B.D. (1991). Vocabulary Learning and Teaching. In M. Celce-Murcia (Ed.), Teaching English as a second or foreign language (2nd ed.). (pp. 296-311). Boston: Heinle & Heinle.

Service, E. (1992). Phonology, working memory, and foreign-language learning. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 45, 21-50.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Smith, M. C. (1991). On the recruitment of semantic information ofr word fragment completion: Evidence from bilingual priming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 234-244.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Experimental Psychology, 6, 174-215.

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

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. System, 21, 371-380.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. Second Language Research, 13, 138-63.

Tokowicz, N., & Kroll, J. F. (in press). Accessing Meaning for Words in Two Languages: The Effects of Concreteness and Multiple Translations in Bilingual Production.

Tokowicz, N., Kroll, J. F., de Groot, A. M. B., & van Hell, J. G. (2002). Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production. Behavior Research Methods, Instruments & Computers, 34(3), 435-451.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. Journal of Verbal Learning & Verbal Behavior, 5, 381-391.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. Journal of Experimental Psychology, 87, 1-8.

Tulving, E., Schacter, K. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 336-342.

Tzelgov, J., & Eben-Ezra, S. (1992). Components of the between-language semantic priming effect. European Journal of Cognitive Psychology, 4, 253-272.

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 760-771.

Underwood, B.J., Ekstrand, B.R., & Keppel, G. (1965). An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. Journal of Verbal Learning and Verbal Behavior, 4, 447-462.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. Bilingualism: Language & Cognition, 1(3), 193-211.

Vigliocco, G., Lauer, M., Damian, M. F., & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. Journal of Experimental Psychology: Learning, Memory, & Cognition, 28(1), 46-58.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. System, 25, 261-274.

Watkins, M. J., & Peynircioglu, Z. F. (1983). On the nature of word recall: Evidence for linguistic specificity. Journal of Verbal Learning and Verbal Behavior, 22, 385-394.

Weinreich, U. (1953). *Languages in contact: Findings and problems.* New York: Linguistic Circle of New York.

Williams, J. N. (1994). The relationship between word meanings in the first and second language: Evidence for a common, but restricted, semantic code. European Journal of Cognitive Psychology, 6(2), 195-220.