An investigation of L2 English learners' knowledge of polysemous word senses

Mark Maby

Centre for Language and Communication Research
School of English, Communication and Philosophy

Cardiff University

Abstract

Polysemy is a challenge for L2 learners because it confounds the mapping of form to meaning. We can therefore consider learners' capacity to manage polysemous words as an indication of their L2 lexical and conceptual knowledge. To investigate what factors affect L2 learners' knowledge of polysemous meanings, a test was created in which Arabic learners of English judged whether various meanings of polysemous words were used acceptably in sentence-length contexts. Analysis of the results revealed that two key factors determined learner responses. First, learners were more likely to respond that a polysemous sense was acceptable if it was more frequently used in English. Second, learners were more likely to judge a polysemous sense as acceptable if was semantically closer to the core sense, such as when *head* is used in the test item, "I went to sleep early to have a clear head for the exam," in contrast to this less closely related use, "The president sat at the head of the table." Semantic similarity was further addressed through distractor items that were unacceptable to native English speakers but logically related to the core sense, such as this use of *head*, "I thought she was upset because she had a sad head," in contrast to the illogical use, "Come through into the dining head." Again, L2 learners generally judged the distractor items as more acceptable if the usage of the polysemous word was semantically related to the core sense. Further analysis revealed that learners with high scores on a receptive vocabulary size test were more likely to correctly reject distractor items; however, there was little indication that L1 form-meaning mappings affected perceptions of L2 polysemy. The implications of these findings for theories of lexical processing, and for the teaching of polysemous words in the classroom, are considered.

Acknowledgement

Table of contents

List of tables and figures

XIII

Chapter 1: The case for investigating L2 knowledge of polysemous senses

It is typically claimed that second language learners are more likely to know words that occur more frequently in the language than words that occur less frequently. There is a clear rationale for this view, namely that learners can recall words better if they encounter words repeatedly across different contexts (Nation, 2001). However, the assumption risks being too broad and inadequately nuanced unless one addresses the question of why a particular word form is frequent. If a word form is frequent because it has many different meanings, what exactly is it that the learner knows about that word? For example, the word *air* does not have just a single meaning as the following sentences illustrate:

(1)   Let's go outside and get some fresh air.
(2)   He threw the ball through the air.
(3)   Air travel was growing rapidly.
(4)   Trudy is always putting on airs.

In these sentences, we can count four instances of the word form *air* or *airs.* However, just as a word can be counted for the frequency of its form, individual meanings associated with that form can also be counted. The meaning of *air* in (1) is GAS[1]. According to the Longman Dictionary of Contemporary English (LDOCE), this is the most frequent meaning, and as such, one would expect that if the learner were to know any of the meanings, she[2] would likely know this one. According to the expectation of intraword meaning-frequency, the other less-frequent senses would be relatively less well known. Besides meaning-frequency, however, the use of *air* in each sentence is notable for how similar or different its meaning is to its use in the other sentences. If the learner only knew the GAS meaning of *air*, we might expect that she could guess the SPACE meaning in (2) because a single real-world experience

---

1. I will follow the practice of Fillmore (1982) and Croft and Cruse (2004) who use lower-case italics for word forms and capitals for concepts.

2. I will use 'she' and 'her' for examples of an L2 learner throughout the thesis. I will use 'he' and 'his' for examples with individuals from other populations.

can account for both concepts. The learner might also guess the meaning of PLANES in (3), which is less related to GAS than SPACE, but still somewhat similar. However, it is unlikely that the learner would be able to guess the meaning of BEHAVIOUR in (4) which doesn't seem to share much semantic similarity to GAS.

The question of how learners develop their knowledge of polysemous senses is interesting because it relates to the larger question of how adult L2 learners develop their L2 semantic knowledge. One could argue that developing semantic knowledge of known word forms is comparable in importance to developing knowledge of new word forms. In terms of vocabulary, a native speaker (or native-like speaker) has two advantages over the adult L2 learner. Generally speaking, the native speaker knows a greater number of words to express himself, and he is also able to use each individual word to represent a greater range of semantic meaning (Nation & Waring, 1997). For example, the native speaker knows if the L1 word can be used in an uncommon context and he may also know how to use the word creatively. If the L2 learner doesn't have the range of word forms to express herself, then she may rely more greatly on the semantic range of the words she does know. However, the L2 learner might be more conservative in her use of the word. She may be reluctant to the use the word differently from the way in which it was taught, let alone to use it metaphorically (Kellerman, 1986). In this way, the L2 learner is limited not only by the fewer number of words she knows, but also by her semantic knowledge of those words she does know.

Polysemous words have specific properties that offer the researcher a useful way to track developing semantic knowledge of L2 words. Unlike homonyms, the senses of a polysemous word are related. While it might be the case that a learner would have to be taught that a homonym like *bank* has both a MONEY sense and a RIVER sense, it is possible for the L2 learner to learn through inference that a polyseme like *air* has both a SPACE sense (2) and a PLANES sense (3). It is also relatively straightforward to test the learner's knowledge of this semantic content, because the researcher can either ask the learner to translate the L2 word in a given sentence or ask whether the word is used correctly or not. The study of polysemy also offers more than this

because some relationships may facilitate learning better than others (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007). For example, the word *hand* has a BODY sense, a WORKER sense and a HELPING sense. All three senses are related, but the relationships between the senses are not equal. Each has its own individual characteristics, which might facilitate or even inhibit inference on the part of the learner.

The aim of this thesis is to investigate how adult L2 learners learn the different senses of L2 polysemous words. The investigation will take the form of a set of empirical studies, and the findings from these will be used to inform the larger question of how these learners develop their L2 semantic knowledge.

There are, however, a number of challenges associated with working with polysemous words. These broadly relate to selecting words for investigation and to defining and labelling the relationships between senses that will enable findings to be extrapolated in a reliable way to polysemy. Labelling relationships between the senses of a polyseme can be quite complex. The word *air* is considered to be polysemous because many of its meanings, or senses, are semantically related to one another. Polysemy is usually defined as the association of a number of distinct but related senses with one linguistic form (Croft & Cruse, 2004, p. 109; Taylor, 2012, p. 219). While this definition refers to different 'senses', very often the term 'meaning' is also used. Some researchers reserve 'meaning' for distinct uses of homonymous words and 'sense' for polysemous words (Degani & Tokowicz, 2010b). However, this distinction is not consistently used in the literature and even where it is, the difference between homonyms and polysemes can blur, as some polysemous senses are less related than others (Croft & Cruse, 2004). A clear example of a homonym would be *air* [GAS] and *airs* [BEHAVIOUR], while a clear example of a polyseme would be *air* [GAS] and *air* [SPACE]. The distinction between homonyms and polysemes can be seen to blur in the example of *horn* [ANIMAL] and *horn* [MUSICAL INSTRUMENT]. Some speakers might see a connection between these two senses, while others may not. In this thesis, 'meaning' will be used as the general term for any distinct

semantic unit and 'sense' will be used for the specific case when relatedness between semantic units is proposed.

Turning now to the selection of appropriate items for investigation, a necessary question to ask is whether an individual word is a representative example of polysemous words in general. For example, is *air* representative of most polysemous words? A*ir* is a noun in the above examples (1-4, page 1), but are other parts of speech also representative of polysemous words? In the examples typically used to describe polysemy, the words are often frequent in a corpus and represented by a single orthographic unit according to the traditions of lexicography. These examples have been not only nouns (*air, line*) (Caramazza & Grober, 1976), but also verbs (*hold, keep*) (Csábi, 2004) and spatial prepositions (*over*) (Brugman & Lakoff, 1988; Lakoff, 1987). The reason behind this choice of words is not arbitrary. The basis for identifying a polysemous word is that the similarity between the word's senses can be traced through conceptually simple similarities. As a corollary to this, most polysemous words will have a core meaning that expresses easily perceptible characteristics. By contrast, the semantic content of function words, such as *of* or *to*, are more bound to their linguistic environment, which makes it difficult to identify distinct senses for these words.

The case has been made that multi-word phrases, such as *of course* or *at all*, could also be considered single lexical-units despite the conventions of orthography presenting these phrases as separate 'words' (Wray, 2015). While polysemy is not defined to exclude multi-word phrases per se, in the treatment of polysemy in this thesis, the form of a polysemous word will be restricted to a string of letters with a gap at either side. To identify polysemous words according to the conventions of orthography has the benefit that the word forms can be counted for their frequency in large corpora. Not surprisingly, it has been found that polysemous words are among some of the most frequent word-forms in the English language (Zipf, 1945). While counting the word forms is relatively straightforward, it is a challenge to count the different instances where a particular polysemous sense is used. For example, how many times is *air* used to mean GAS as opposed to SPACE? This is a

problem which needs to be addressed when assessing the intra-word frequency of polysemous senses.

The most practical way of identifying different polysemous senses is through reference to dictionaries. The benefit of using a dictionary is that the senses are clearly listed, and in some cases with some reference to their intra-word frequency. Intra-word frequency refers to how frequently a polysemous sense occurs in a corpus. In dictionaries this is presented as a relative measurement, whereby a given sense is more frequent than those listed below it and more frequent than those listed above it. Modern dictionaries take advantage of large, comprehensive corpora to aid in their identification of different senses. Writing in the field of lexicography, Hanks (2013) has used frequency in a corpus to identify different senses of a word. In his analysis, a distinct sense is established as a norm through a frequently recurrent lexical pattern. Specifically, Hanks says that, "A norm is a pattern of ordinary usage in everyday language with which a particular meaning or implicature is associated. A pattern consists of a valency structure[...], together with sets of preferred collocations," (Hanks, 2013, p. 92). For example, different senses of *file* (as a verb) can be distinguished by whether it is followed by a noun phrase or an adverbial phrase: [NP] *file* [NP] contrasts with [NP] *file* [AdvP] as in 'Deacon filed his first patent' vs. 'The mourners filed into the church' (Hanks, 2013, p. 130). According to Hanks' method of corpus analysis, these recurrent lexical patterns are referred to as norms, and the most frequent patterns will be the most central norms.

While Hanks is writing about lexicography, it is reasonable to expect that the centrality of norms in the corpus has some equivalence in a native speaker's mental lexicon. The corpus is an approximation of the speaker's language experience and so the tendencies of a word in the corpus should reflect the usual linguistic context in which the speaker encounters the word. To a degree this is somewhat applicable to L2 speakers as well. Many researchers have written about how it is important for learners to recognise and understand the most frequent 2000 - 3000 words in English for comprehension of a text (Laufer, 1992; Laufer, 1997; Nation, 1990; Nation, 2001; van Zeeland & Schmitt, 2013). While encountering a word frequently will facilitate

5

learning the form of the word, we need to question how easy it will be for a learner to learn the semantic content of the word if it is polysemous. Laufer has noted that polysemous words are difficult for L2 learners because they fall into the category of "words you think you know," but don't (Laufer, 1997, p. 26). Even if a learner knows enough words for adequate text coverage, she may still misinterpret the meaning of a polysemous word if she doesn't realise that the word expresses a different sense from the one she knows. The learner may be familiar with a word's most frequent norms of use, but she may not recognise that certain differences in context indicate a change in meaning.

One way an L2 learner might correctly interpret an unknown sense of a word is if it bears similarity to other senses of the word that the learner already knows. As mentioned above on page 2, a learner might be able to infer the meaning of *air* in the sense of PLANES (3) if she already knew *air* in the sense SPACE (2). In a qualitative study, MacArthur and Littlemore (2008) presented L2 learners with a variety of different uses of words as listed in a corpus. They found that the learners would use a 'core' or 'enabling' sense to guess the figurative meanings. Furthermore, some researchers have investigated whether noticing the similarity between senses of a polysemous word can aid an L2 learner in learning new senses (Csábi, 2004; Huang, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). Their research, reviewed in Chapter 2, investigates the learning benefit of explicitly teaching the similarity between senses. There have been some positive results from these studies for teaching. However, there has been less research about whether L2 learners would use the semantic relatedness expressed by polysemous words to learn new senses outside explicit teaching methodologies. Evidence that semantic relatedness facilitates implicit L2 learning of polysemous words would support further research into more explicit teaching methodologies. In the empirical studies of this thesis, different constructs of semantic relatedness will be used to determine how L2 learners of English develop their knowledge of the senses of polysemous words.

The research in this thesis has been informed by three different fields of study, which are reviewed and explored in Chapter 2. The field of cognitive linguistics provides the theoretical background to the understanding of polysemy. In cognitive linguistics, the account of semantics in general, and polysemy in particular, is informed by cognitive principles (Croft & Cruse, 2004). These principles lend themselves to the study of L2 knowledge of polysemous words because L2 language learning involves many competing cognitive factors. I also turn to psycholinguistic research on how native speakers process and store polysemous words. Finally, I review the research on L2 learner knowledge of polysemous words and the research on teaching L2 learners the different meanings of polysemous senses.

In subsequent chapters, I report how, based on previous research, I developed a new instrument for testing L2 knowledge of polysemous words. In psycholinguistic research, a common method for investigating a speaker's lexical knowledge is through judgements of semantic acceptability. This type of task asks the participant to judge whether a lexical item is used acceptably in a given linguistic context. The task can be used not only to assess whether the participant knows the lexical item or not, but the items can also be designed to measure different factors affecting the participant's judgements. In the instrument designed for this research, the items were designed to measure not only the influence of semantic similarity on L2 polysemous word knowledge, but also other competing influences, namely frequency of occurrence and L1 influence.

Using this instrument based on judgements of semantic acceptability, I gathered data from Arabic learners of English who were studying at a college in the country of Qatar. By using participants with similar profiles across the studies, I was able to maximise the comparability of the findings. The development of the instrument and collection of the responses are presented in Chapters 3 - 7.

In Chapters 8 - 10 I report how, having created the instrument and used it to collect responses from the L2 learners, I analysed the responses for the influence of four factors: intraword frequency of the polysemous senses, L2 proficiency, semantic

similarity between the senses, and the influence of the learners' L1. The thesis ends with a discussion of the limitations of the findings and their value for three issues: semantic similarity in learning a second language, the influence of the L1 on L2 conceptual knowledge, and the distinction between vocabulary size and depth of vocabulary knowledge.

Chapter 2: Literature Review: L2 learning of polysemous senses

The investigation of polysemous words is of interest to a number of fields of enquiry because it allows the researcher to question how semantic content is associated with word forms. For the field of second language acquisition, polysemous words raise questions about how the meaning of a word develops in relation to the learner's L1 and her general learning environment. For the theoretical linguist, the question is whether the semantic content can be separated into distinct senses and how the boundaries between senses are to be identified. In psycholinguistics, polysemous words raise questions about how senses are represented in the mental lexicon and how they are processed.

Each of these fields of study has had an influence on the research which specifically investigates what knowledge L2 learners have of polysemous words and how polysemous senses can be taught in a second language. The following literature review will follow this research and end at more detailed reviews of several studies dealing directly with the knowledge and teaching of L2 polysemous words.

2.1 Polysemy as an aspect of L2 word knowledge

In the literature on L2 learning, knowledge of polysemous words is considered to be an aspect of depth of word knowledge. This is because the question is not whether the L2 learner knows the polysemous word form, but how she develops her knowledge of the word's different senses. In their attempts to understand how L2 learners develop their knowledge of polysemous words, L2 researchers have turned to the insights from cognitive linguistics. This research has attempted to describe how the meaning of polysemous words can be described within the framework of general cognitive principles. In this section, I will first consider the treatment of polysemous words inside the field of L2 vocabulary acquisition. Then I will review how the field has been informed by research from cognitive linguistics for its understanding of how polysemous words form a distinct semantic category.

2.1.1 Depth of L2 vocabulary knowledge

In the literature on second language learning of vocabulary, knowledge of the polysemous senses of a word has often been incorporated into a framework of depth of vocabulary knowledge, which considers how well words are known. This is placed in contrast to breadth of vocabulary knowledge, which considers how many words are known (Anderson & Freebody, 1981). This distinction has also been characterised as size vs organisation, whereby size refers to the number of different word forms and organisation to the interconnections of the lexicon (Meara & Wolter, 2004). Depth of vocabulary knowledge is seen as important because as learners become more advanced in their proficiency, they will need to understand and use a wider range of senses, collocates, and associations in a wider range of contexts (Bogaards, 2000; Schmitt, 2014; Wesche & Paribakht, 1996).

A number of frameworks have been proposed to characterise depth of vocabulary knowledge and knowledge of polysemous meanings within these frameworks. An early and well-cited framework was developed by Richards (1976). He proposed eight assumptions about knowledge of a word including the assumption that, "Knowing a word means knowing many of the different meanings associated with the word," (Richards, 1976, p. 82). Based upon Richards' work (1976), Nation (2001) constructed a more systematic framework. He divided knowledge of a word into three broad categories of meaning, form, and use. In this framework, the aspect of *concepts and referents* includes the range of different meanings listed in a dictionary. Some of these meanings will bear no relationship between each other and the words can be considered homonymous. However, there is semantic similarity between other meanings, and while Nation doesn't use the term, the meanings which express this type of relationship are usually characterised as polysemous (Taylor, 2012).

In their frameworks, Richards and Nation attempted to provide a comprehensive description of the different aspects of depth of vocabulary knowledge. In terms of testing, it has not been possible to measure a learner's knowledge of all the aspects in a single comprehensive assessment. The frameworks are quite complicated and it

may be impossible to design an instrument to measure all the aspects. One early attempt to measure some aspects of depth of vocabulary knowledge was the Vocabulary Knowledge Scale (VKS), (Wesche & Paribakht, 1996). The VKS presented learners with five questions about their knowledge of individual words. Each question demanded increasingly sophisticated knowledge about the word. The VKS has been characterised as developmental (Read, 1997) because it assumes knowledge develops in stages. Schmitt (1998) has questioned this approach because of the difficulty of dividing the learning process into discrete stages. He prefers a dimensional approach which measures knowledge of different aspects of the word, such as how many polysemous senses a learner knows or whether the learner can spell the word.

In contrast to the VKS, the Word Associates Format (WAF) is a test based on the dimensional approach to word knowledge. The WAF is interesting because one component of the test targets knowledge of different meanings of a stimulus word. When considered in this light, the WAF provides an example of how L2 knowledge of polysemous words can be measured.

The WAF is the most widely used type of test for assessing a learner's knowledge of different aspects of a word, specifically paradigmatic knowledge of alternate words for the same syntactic role and syntagmatic knowledge of how the word should be sequenced with other words. This format was first developed by Read (1993, 1994, 1995, 1998), and while different variations of the test exist, they all present the learner with a target word and eight other words that are either associates or distractors. The following example is from the 1998 version of the test.

convenient

easy fresh near suitable	experience sound time vegetable

The four words in the left-hand box target paradigmatic knowledge. The associates *easy* and *near* are synonyms of *convenient* and they refer to different senses of the

11

target word. The four words in the right-hand box target syntagmatic knowledge. The associate *time* is used in the collocation *convenient time*. Correctly selecting this associate indicates knowledge of that collocation.

The paradigmatic measure of the WAF can be understood as a measure of polysemous word knowledge because if a learner knows distinct synonyms of a word, then she must also know the different senses of the stimulus word. The WAF has been used in a number of studies which investigate the relationship between breadth and depth (see Schmitt, 2014 for a review). It is useful to consider how well the WAF functioned as a research instrument. This provides a way of assessing its effectiveness as a measure of polysemous word knowledge.

Qian (1999, 2002) and Akbarian (2010), used the WAF to measure knowledge of a number of aspects including L2 knowledge of polysemous senses. Qian found a strong correlation between the measurement of breadth and depth for advanced learners of English ($r = .70$, $.78$, and $.82$, across the different population samples measured). Furthermore, Qian also found that the measurement of depth (WAF) accounted for a greater amount of the variation within the learners' results on a test of reading comprehension. The scores on the WAF were designed to be a combination of the results from both the paradigmatic and syntagmatic components. Knowledge of polysemous words is part of the overall correlation results; however, the strength of the effect of that knowledge remains unclear.

Akbarian divided his participants into higher and lower proficiency groups based on their results on the breadth of knowledge test. He too found a strong correlation between breadth and depth ($r = .90$ for the high proficiency group and $r = .68$ for the low proficiency group). However, his participants' results on the WAF are problematic because they could have got the same scores by guessing. There are four associates and four distractors on the WAF. A test taker would have to get more than 50% correct to do better than chance. Akbarian's higher proficiency group had a mean result of 52%, just above chance, and his lower proficiency group a result of just 27%.

Guessing may be a general issue for the WAF and not one specific to Akbarian's participants. Dronjic and Helms-Park (2014) found that even when native speakers took the WAF, there was a lack of consistency in the responses which indicated a high degree of guessing. It may be that the effect of guessing was more pronounced with Akbarian's lower proficiency learners.

Dronjic and Helms-Park (Dronjic & Helms-Park, 2014) also found that native speakers achieved less than 90% consistency in their responses on the WAF. They argue that without higher consistency of response, there is no warrant to use the test with L2 learners. There was greater consistency in responses among native speakers for the paradigmatic component than the syntagmatic component. However even then, Dronjic and Helms-Park point out that the paradigmatic component has a design issue of its own. Since the WAF identifies knowledge of different meanings using synonyms, the test may fail to accurately identify a learner's knowledge if she knew the meaning of the stimulus word but was unfamiliar with the target synonym.

In sum, the research using the WAF has been promising because it has shown how depth of vocabulary knowledge, including knowledge of polysemous words, is important for reading comprehension. However, there have been issues with the reliability of the test in general. In particular, as a measure of polysemous word knowledge, the test may not be an accurate enough measure for learners who do not have enough breadth of vocabulary knowledge.

2.1.2 Investigation into L2 learning of polysemous senses[3]

As discussed above, the score of the WAF is only partly influenced by the learner's knowledge of polysemous words. However, there has been other research which has looked more precisely at polysemous word knowledge. Kellerman (1986) investigated L2 learners' knowledge of polysemous words. Kellerman looked at transfer of L1 meanings to L2 word forms. Specifically, he asked his participants how acceptable they felt it was to transfer a list of meanings from their L1 to their L2. Unknown to the learners was that all the meanings were acceptable in the other language; however, the learners were reluctant to transfer metaphorical meanings or meanings that were generally removed from the core meaning of the word form.

In contrast to Kellerman's study of L1 transfer, Schmitt (1998) and Crossley, Salsbury, & McNamara (2010) conducted longitudinal studies where they followed L2 English university students over the course of one academic year in their development of learning a small number of words (eleven for Schmitt and six for Crossley et al.). The words they chose were polysemous, and they tracked how the number of senses developed over time. Schmitt found that the learners used more senses at the end of the semester, while Crossley et al. found that the learners consistently increased their knowledge of polysemous senses over the year. Crossley et. al. also found that the learners developed their knowledge systematically from the core meaning to peripheral meanings in the later part of the academic year.

There has not been a great deal of research on L2 learner knowledge of polysemous words. This makes the research of Kellerman, Schmitt and Crossley et al. all the more important and worthy of more detailed discussion. However, there are two

---

3. One could argue that polysemy is breadth not depth— that is, you know more words. This argument would be strongest for homonyms where the meanings are inconsistent with each other. As polysemes become more semantically similar, then the case for depth is stronger, in that the speaker knows more aspects of the same word. I think it is more helpful to consider that the most authentic measure of breadth is a test that measures correct recognition of word forms (Meara, 1992). Any knowledge beyond word form can be considered an aspect of depth, be it meaning, register, variety, etc.

important topics to address before moving into that detailed discussion. First, it is useful to establish a model of L2 vocabulary learning and the place of learning polysemous words within that model. Secondly, it also useful to describe in general the theory behind polysemy. After addressing these topics, I will return to the studies on L2 learner knowledge of polysemous words.

2.2 Models of L2 vocabulary learning

Knowledge of L2 polysemous words can be understood as a subordinate class to L2 vocabulary knowledge in general. As such, the way in which these words are learned can be related to more general models of L2 vocabulary learning. Two such models will be considered here: Jiang's (2000) psycholinguistic model of L2 vocabulary learning and Wolter's (2009) meaning-last model. As a point of departure, it is useful to briefly sketch an L1 model of vocabulary learning, as this can act as a contrast to the L2 models. Aitchison's (2003) description of L1 learning will serve this purpose.

Aitchison describes three stages to L1 vocabulary learning. In a *labelling* stage, the child attaches semantic meaning to a word form. In the *packaging* stage, the child establishes the semantic boundaries for how the word form can be used. For example, boundaries may include which four legged animals can be called *dog* and which can be called *Toby*. Finally, in the *wiring* stage, the word is established in the developing mental lexicon. There are at least two important differences in the adult L2 learning situation (Jiang, 2000). First of all, in his or her L1, the adult learner has an existing linguistic/conceptual system in place which will likely greatly influence the L2 learning, either to facilitate or inhibit learning. Second, the L2 learner, especially in a classroom learning context, has very limited exposure to the L2. This means there is likely a limited amount of meaningful input and a limited opportunity to practice language use. These two points of difference motivate changes to the structure of the L2 vocabulary learning model and the rate of time a learner takes to progress.

In contrast to the L1 model of vocabulary learning, Jiang takes as a first stage in L2 vocabulary learning the association of an L2 word form with an L1 translation equivalent. The translation equivalent is considered a combination of the lexeme (the lexical item's phonology/orthography and its morphology) as well as its lemma (the semantic and syntactic information). At this stage, the knowledge of the L2 word, beyond the word form itself, is entirely dependent on the L1 translate equivalent. As the learner's experience with the L2 develops, a stronger link is made between the L2 word form and the L1 lemma, so that the L1 lexeme no longer acts as an intermediary. L2 vocabulary is still dependent upon the conceptual structure of the L1. Jiang comments that much of an L2 learner's vocabulary will not progress beyond this stage.

The final stage is when the L2 word form is associated with an L2 lemma that is distinct from the L1 lemma. Whether an L2 learner reaches this stage depends very much on the quality of the vocabulary item. Jiang presents three types of L2 lexical items to illustrate this stage of learning. He refers to the first type as *strangers*. These are words in the L2 which have no direct translation equivalent in the L1. These lexical items have the greatest chance of being established as distinct L2 lemmas because the learner must account for the lack of support from the L1 conceptual structure. In contrast, *friends* are lexical items which share nearly identical lemmas across the L1 and L2. There may never be an occasion where the L1 lemma proves inadequate for comprehension, and thus further elaboration to form a distinct L2 is least likely for this type of word. The final type of lexical item are *false friends* where there is only partial overlap between the L1 and L2 lemmas. Polysemous words fit into this category because while some senses of the word may be shared across languages, other senses may be distinct to only one. Jiang comments that a learner is more likely to form a distinct L2 lemma for these words because she may notice where the L2 use diverges from the conventions of the L1 use, which provides an opportunity for developing new semantic content particular to the L2.

In both Aitchison's L1 model of vocabulary learning and Jiang's L2 model, the lexical item is largely learned as an isolated unit. The role of linguistic context in Jiang's

16

model appears to specify to the learner whether the L2 semantic content is congruent with that of the translation equivalent. The role of linguistic context takes a far greater role in Wolter's meaning-last model of vocabulary learning. This model is largely restricted to learning vocabulary incidentally through reading. Wolter argues that in this situation the word may initially be learned as a larger lexical chunk or collocation. It is at a later stage of development that the L2 learner will isolate the individual word from its typical context and develop its semantic content. Indeed, Wolter argues that the learner may use the the lexical item productively as part of collocation before isolating the item from the co-text and elaborating its semantic content.

Taken jointly, Jiang's and Wolter's respective models provide two possible scenarios for L2 learning of polysemous words. At the outset, the opportunity to learn a polysemous word is only considered once the learner encounters a second, semantically similar sense. According to Jiang's model, the learner is likely to notice the sense if it arises in a context that is different from the conventions of the L1 translation equivalent. Noticing the difference provides an opportunity to develop semantic content for the word distinct from the L1 equivalent. However, developing this new semantic content may take repeated encounters, and, according to Wolter's meaning-last model, the learner may develop collocational knowledge of the word prior to developing distinct semantic knowledge. The development of both semantic knowledge and collocational knowledge are considered aspects of vocabulary learning.

What might distinguish polysemous words for L2 learners is the similarity an unfamiliar use bears towards the existing semantic content associated with the L1 lemma. This factor of semantic similarity may facilitate why an L2 learner may develop semantic knowledge for the unfamiliar sense. Semantic similarity will be a key factor in the subsequent investigation of how learners develop their knowledge of polysemous words. The following section addresses in more detail how polysemous words are understood as a category of semantically related senses.

2.3 An approach to polysemy from cognitive linguistics

It is important to review how polysemy is characterised in cognitive linguistics, because this research has informed a number of studies on teaching polysemous senses to L2 learners (for example, Csábi, 2004; Huang, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003) . The researchers in these studies have taken the claim that polysemous words are represented in the mental lexicon as radial categories, whereby different senses of a polysemous word are semantically related to a core sense. In these studies, the researchers have attempted to exploit the semantic similarity between the polysemous senses to facilitate L2 learning of new senses.

A word is understood to be polysemous if its senses are distinct and yet related to one another. In the examples below, *hand* refers to three different referents: in (1) a BODY part, in (2) assistance or HELP, and in (3) WORKERS.

(1)   She waved her hand to the crowd.
(2)   Can you give me a hand to lift this?
(3)   The farm hands wake up at 5:00 in the morning.

The Oxford Dictionary of English (ODoE) identifies the core meaning of *hand* as the one used in (1). This makes intuitive sense because the BODY sense is the most common use of the word and the one that was probably learned first. The other two senses can be said to derive their meaning from the core meaning. We use our hands to assist people, which relates to the sense of HELP in (2), and farm workers use their hands to do their work, which relates to the sense of WORKERS in (3).

The extension of meaning from a core sense to a peripheral sense is characterised as radial. It seems reasonable to suppose that this radial relationship resulted from the etymological history of the word. However, researchers in cognitive linguistics go beyond this to claim that this radial structure is in some way mirrored in the mental lexicon of mature speakers (Lakoff, 1987).

In cognitive linguistics, polysemous senses, and their relationship between each other, are explained through five basic theoretical constructs: concepts, domains, construal, image schemas and categories (Clausner & Croft, 1999). It is useful to review how polysemous senses can be described through these constructs, because this description forms the linguistic argument for why the radial relationship of polysemous senses is said to explain how they are structured in the mental lexicon.

2.3.1 The construal of polysemous senses

Three different senses of the word *hand* were presented above. These senses are considered distinct from one another because they each represent different concepts. Following Clausner and Croft (1999, p. 2), a 'concept' is understood to be the basic unit of mental representation. It equates a single linguistic expression with meaning. It can correspond to categories such as *birds* or to individuals such as *Graham Norton*[4]. Thus, *hand* [WORKER] is said to be a different concept from *hand* [BODY].

Since both the WORKER sense and the BODY sense share the same word form of *hand*, the question to ask is how one concept is distinguished from the other. Perhaps not unsurprisingly, concepts are comprehended against the context of background structures (Langacker, 1987). This is a necessary condition because a concept can not be understood without the background of its context. For example, a KILOMETRE, a standardised distance between two points, can only be understood against the background of SPACE. The claim is that if we profiled the concept against a different background, then the concept would change. For example, a standardised duration between two points when profiled against the background of TIME, would refer to the concept of an HOUR. These background structures are referred to as domains. Following Langacker (1987), a concept is said to be profiled against its domain when it is comprehended by the language user.

---

4. Croft and Cruse (2004) comment that non-linguistic categories may exist, but they restrict their discussion to linguistic examples, and their examples are all nouns.

The relationship between concepts and domains can be understood as a PART-WHOLE relationship because, just as one can't understand a FINGER without knowledge of a HAND, one can't understand a concept without knowledge of its domain. Domains have different levels of complexity. There is a distinction made between a basic domain and a domain matrix. A basic domain is rooted in fundamental human bodily experiences such as SPACE and TIME (Lakoff, 1987), whereas a domain matrix is a combination of domains based on the person's knowledge and experience of the world. A domain matrix can be quite a complex context of experience, such as BASEBALL or FUTURES TRADING (Clausner & Croft, 1999, p. 21). Most concepts are profiled against domains that are more complex than simply TIME or SPACE. Moving forward, I will use domain to refer to these more complex domain matrices or what Verspoor and Lowie refer to as "domains of experience" (Verspoor & Lowie, 2003, p. 555).

Even when complex however, domains do not carry all the detail of a real world context. Lakoff (1987) has described the domain of some words with a wide range of use as an Idealized Cognitive Model (ICM). The ICM is a limited background understanding of the world that explains why some concepts, such as *bachelor*, do not adequately deal with all real world situations. A *bachelor* is understood to be an unmarried man, but it would be wrong to say that the Pope is a bachelor (Fillmore, 1975, 1977). The *bachelor* example shows that mental representation of a domain can be left somewhat underspecified to accommodate the variety of real-world examples.

If the claim is that a speaker can distinguish between the concepts of *hand* [BODY] and *hand* [WORKER] based on the domain, then there needs to be a cognitive process by which that distinction is made. This process of profiling a concept against a domain is referred to as construal. Construal is potentially important for L2 learning of polysemous senses because it shows how a learner can distinguish between two different senses of a word.

In the process of construal, the semantic representations in the speaker's mind - the concepts and domains - are related to the speaker's experience of the world. The following two sentences (taken from Clausner & Croft, 1999) show how the polysemous word *pour* is construed differently depending on its domain:

(4)   He poured the juice through the sieve.
(5)   The fans poured through the gates.

The two concepts of POUR in these examples both refer to something moving freely from one location to another. The difference between the two senses is that in (4) the concept of POUR is profiled against the domain of LIQUID and in (5) the concept, POUR, is profiled against the domain of PEOPLE. The process of construal allows the language user to fit the concept to the domain based on his experience of the world. In this way, *pour* in (4) refers to a liquid mass falling downwards, while *pour* in (5) refers to a large number of individual people walking along a walkway.

As presented here, construal seems satisfactory as a cognitive process for discerning the meaning of senses which the listener already knows. If the listener already knows that *pour* can refer to PEOPLE or that *hand* can refer to WORKERS then the process of construal is useful for explaining how the known concept is combined with the known domain. One question to ask about construal is whether the process can lead the listener to understand novel senses. Without prior knowledge, it must be strange that *pour*, a word for LIQUIDS, should be used be used for PEOPLE, let alone that *hand* should refer to WORKERS.

This question of discerning the meaning of unknown senses is especially relevant to L2 language learners, because they are likely to encounter unfamiliar senses on a regular basis. Whether language learners can discern the meaning of unknown senses will be pursued through empirical investigation in the experiments presented later. At the moment however, it is useful to consider the cognitive principles that might allow an adult language learner to infer the meaning of an unknown sense.

2.3.2 Image schemas and the relationship between senses

In order to correctly infer the meaning of an unfamiliar sense, the listener must rely on his existing knowledge of the world and his knowledge of the language. The situation becomes more complicated when we consider the case of an L2 learner, because not only will she have limited knowledge of the language and but potentially limited knowledge of the culture of her L2 as well. At the outset then, it is useful to be conservative and ascribe to the L2 learner only fundamental knowledge of the world.

In cognitive linguistics, the image schema is presented as the basic building block of experience for semantic comprehension of language. The idea of an image schema is presented by Lakoff (1987) and Johnson (1987) as a mental structure that arises out of our basic bodily experience of the world. As such, it is considered to be as close to a universal as experience of the world will allow [5]. Image schemas are considered universal because evidence of their development is taken from preverbal infants (Mandler, 2005). Furthermore, Mandler also maintains that development of image schemas informs language development. Some of the earliest acquired verbal concepts such as PATH, CONTAINMENT, and SUPPORT, are all preceded by equivalent image schemas acquired during preverbal development (Mandler, 2005).

The mental representation of an image schema is similar to the abstract idea of a geometric shape, such as a TRIANGLE. The idea of a TRIANGLE can accommodate any specific triangle, be it isosceles, equilateral, obtuse, etc. However, as an abstraction, the idea of the TRIANGLE is not rich enough in detail to be visualised itself (Lakoff, 1987, p. 453). Likewise image schemas are abstract mental structures which are used in the process of forming meaning but are not rich visuals in themselves. One example of an image schema is a CONTAINER. This image schema is said to be derived from the experience of our bodies. We experience being inside a 'container' of bounded space, but we also experience our bodies as 'containers' themselves, with

---

5.   Bowerman and Choi (2001) have shown how there are differences between languages even at the level of image schemas.

an inside and an outside. The CONTAINER image schema gives rise to basic concepts, such as IN, OUT, FULL, and EMPTY. The resulting concepts can be both imagistic, 'I'm in the room,' and non-imagistic, 'I'm in trouble.' Clausner and Croft (1999, p. 15) present Table 2.1 as an inventory of image schemas that have appeared in cognitive linguistics literature:

Table 2.1
Inventory of image schemas

SPACE	UP-DOWN, FRONT-BACK, LEFT-RIGHT, NEAR-FAR, CENTER-PERIPHERY, CONTACT
SCALE	SOURCE-PATH-GOAL
CONTAINER	CONTAINMENT, IN-OUT, SURFACE, FULL-EMPTY, CONTENT
FORCE	BALANCE, COUNTERFORCE, COMPULSION, RESTRAINT, ENABLEMENT, BLOCKAGE, DIVERSION, ATTRACTION
UNITY/MULTIPLICITY	MERGING, COLLECTION, SPLITTING, ITERATION, PART-WHOLE, MASS-COUNT, LINK
IDENTITY	MATCHING, SUPERIMPOSITION
EXISTENCE	REMOVAL, BOUNDED SPACE, CYCLE, OBJECT, PROCESS

These image schemas are useful for identifying distinct polysemous senses. For example, the distinction between the two senses of *pour* in (4) LIQUID and (5) PEOPLE is based on the MASS-COUNT image schema. This image schema is under-defined to the extent that it can accommodate both a MASS quantity, like a LIQUID, or a COUNT quantity, such as PEOPLE. The contention here is that a listener who understood the concept of POUR when profiled against the domain of a MASS quantity, would not have difficulty in profiling POUR against the domain of COUNT quantity. It is claimed that since this transformation is based on a basic experience of the world, it is available to language learners independent of their cultural background.

It is obvious however, that image schemas do not provide enough propositional content to understand just any linguistic utterance. In the *pour* examples, the listener also needs knowledge of *juice* [LIQUID] and *fans* [PEOPLE]. Moreover, Lakoff has

claimed that the relationship between polysemous senses can also be made through propositional content. To use Lakoff's example, a glass *window* on the side of a house is quite different from a hole cut into the side of building, or a sheet of glass in a shop, and yet all three can be called *windows.* The house *window* expresses [+PANE], [+FRAME], [+OPENING], while the cut-hole *window* expresses [-PANE], [-FRAME], [+OPENING], and the pane-of-glass *window* expresses [+PANE], [-FRAME], [-OPENING]. In order to understand all three senses of *window*, the listener needs to understand the propositional content of what a window is, as in the first sense.

While it is not too much of a stretch to assume that L2 learners would know the propositional content to understand these senses about windows. Other types of propositional knowledge can be much more sophisticated. For example, let's say an L2 learner knew the verb *to gut* in terms of cleaning a fish. In order to interpret the phrase "I will be *gutted,* if I loose*,"* she would need to understand that *gut* was used metaphorically to refer to disappointment.

I will discuss the nature of figurative language later in the thesis. At the moment, it is important to point out that polysemous senses can be related to one another in different ways, by image schemas, by propositional content based on knowledge of the world, or by propositional content based on cultural conventions like many metaphors. Some of these relationships seem quite intuitive and available to the L2 learner, while others appear much more opaque. It is an empirical question to see whether learners can infer the meaning of polysemous words based on these relationships.

Before turning to some empirical research on this topic, it is best to consider a final component in the description of polysemy. In order for a learner to infer the meaning of an unfamiliar sense, it is assumed that L2 learner already has knowledge of one sense of the word. Not all senses are equal, but according to cognitive linguistics, some senses lend themselves to extension better than others. This factor is potentially another reason why an L2 learner is able to successfully infer the meaning of polysemous senses.

2.3.3 Polysemous words as semantic categories

In the examples given so far the relationship between senses has been shown to go in one direction. From *pour* [LIQUID] to *pour* [PEOPLE], from *hand* [BODY] to *hand* [WORKER], and from *gut* [FISH] to *gut* [DISAPPOINTMENT]. This direction of extension was not arbitrary but based on the semantic category structure described in cognitive linguistics.

A semantic representation is said to have a category structure because a word or a phrase can be used to identify any number of real world examples. In the category of FRUIT, for example, the different members are referred to as *apples*, *bananas*, *grapes*, and so on.

The fact that these real world examples are not all alike results in language users identifying certain instances to be better examples of the category than others. An APPLE may be said to be a better example of a FRUIT than an OLIVE. In order to explain this phenomenon of graded category membership, semantic categories are often said to have a prototype or radial structure, whereby a central core example of the category is said to express all the defining qualities of the category. This core example may be an idealisation if no single example fully meets all its criteria. (Wittgenstein's discussion of *game* (1980) is a well-known example of how the core example is an idealisation). Non-core examples of this category are said to be peripheral extensions of this core example. The non-core examples express some but not all the qualities expressed by the core example.

Polysemous words represent a special type of semantic category because each member of the category is referred to by the same phonological word form. Rather than having a core example like FRUIT: APPLE, a polysemous word is said to have a core sense, such as HAND: *hand* [BODY]. Like other semantic categories, polysemous words are said to express a radial or prototype structure whereby membership is gradient based on similarity to a core sense. The Oxford Dictionary of English defines the core meaning as the following:

> Core meanings represent typical, central uses of the word in question in modern standard English… The core meaning is the one accepted by native speakers as the most literal and central in ordinary modern usage. This is not necessarily the same as the oldest meaning, because word meanings change over time. Nor is it necessarily the most frequent meaning, because sometimes the most frequently used modern sense of a word is a figurative or extended one. (Stevenson, 2010, p. xi)

Elsewhere, the dictionary uses the term "core sense", and so 'meaning' and 'sense' appear to be used interchangeably. The core sense can also be defined as the 'logical' central use of the word. This application provides coherence to the whole category of senses in that there are clear semantic relations between the senses and the central use (Geeraerts, 2007; Verspoor & Lowie, 2003).

A polysemous word is said to form a semantic category because the members of the category, referred to as senses, are both similar enough to be categorised together and yet distinct from one another to be understood as separate concepts (Lakoff, 1987; Taylor, 2012). Exactly how similar or how distinct the senses need to be from one another is a matter of debate. Crossley et al. (2010) have argued that it is best to see polysemy as a continuum between vagueness at one end and homonymy at the other. Vagueness refers to the case where a single general meaning can accommodate the differences in sub-cases. Tuggy (1993) gives *aunt* as an example of vagueness, because an *aunt* can refer to both FATHER'S SISTER and MOTHER'S SISTER. These types of meanings are usually considered too similar to be distinct polysemous senses. On the other end of the continuum is homonymy, where two meanings share the same word form, but there appears to be no semantic connection between them. The textbook example of a homonym is *bank* [RIVER] and *bank* [FINANCIAL INSTITUTION]. While there is a case that these meanings may share some etymological origin (Hanks, 2013), homonyms are usually understood as having different etymological origins. However, when identifying polysemous senses, the important distinction is whether the meanings are related at the synchronic level in the mind of the language user. In later sections, homonyms will be referred to as

meanings that are inconsistent with the core sense of the polysemous word. This distinction is made without reference to whether the senses has a historic connection to the core sense or not.

The stronger point of contention between linguists is how broadly or narrowly should polysemous senses be defined. Depending on their approach to identifying distinct senses, linguists have been described as either lumpers or splitters (Taylor, 2012, p. 223). In the example from Lakoff (1987), given above, a *window* [+PANE], [+FRAME], [+OPENING], was proposed to be distinct from a *window* [-PANE], [-FRAME], [+OPENING]. This level of fine distinction is the approach taken by splitters. Lumpers would reject this distinction because a single general sense of *window* could be vague enough to accommodate both senses of the word. The lumper approach would identify senses based on contextual differences. For example, a *window* [COMPUTER INTERFACE] would be considered distinct from *window* [HOUSE]. There is even the monosemy approach taken by Ruhl (1989) who proposes all senses of a single word form are subsumed under a single vague meaning.

One principled attempt to resolve the lumper/splitter debate can be seen in the position taken by Wierzbicka (1996), Goddard (2000) and their colleagues. They have proposed that concepts have necessary and sufficient conditions, and the words that refer to these concepts have clear, unambiguous meanings. Clearly, polysemy, which is a category of ambiguous words, poses a challenge to this model. As evidence of this proposition, they argued that words and their concepts can be defined using a language-independent metalanguage, referred to as a natural language metalanguage (NSM). While complex concepts can be described using simpler concepts, this metalanguage is made up of semantic primes which, it is argued, are concepts that cannot be reduced to anything simpler. Examples of English words referring to some of the semantic primes are I, YOU, WANT, FEEL, KIND OF, and PART OF (Wierzbicka, 1996).

While the NSM approach does accept that words can be polysemous, both Wierzbicka and Goddard argue that most dictionaries present too many meanings.

The method they recommend is to start with a single definition of a word and add other definitions only when a single meaning fails. For example, in French the word *fille* can refer to both a GIRL and a DAUGHTER (Goddard, 2000). A single definition fails in this case because a middle-aged woman is not a girl but can still be a daughter. This is a clear example of polysemy according to the NSM approach because the two concepts of GIRL and DAUGHTER are distinct but related.

As a consequence of taking the position that concepts have necessary and sufficient conditions, the NSM approach is in contrast to the idea that concepts have fuzzy borders and that there can be better and worse examples of a concept. In opposition to Wittgenstein's idea of family resemblances (referred to above), Wierzbicka argued that the concept of GAME can be described using a single definition. She laid out her definition with nine conditions (1996). However, as Hanks pointed out, there are many cases where examples of games don't fit the conditions of the definition (Hanks, 2013, p. 326). One condition is that *games* are "many kinds of things that people do," however, don't animals also play games? Another condition is that they are done "for pleasure", but its doubtful that professional athletes play games fun or pleasure. Furthermore, children often play games that are not "goal oriented", and games like ring-a-rosie are not "unpredictable". One could argue that the exceptions are not 'true' games but are 'like' games. However, such a position is open to criticism for making the concept suit one's purposes despite evidence to the contrary.

In sum, the NSM approach to polysemy is beneficial from a lexicographical perspective because it offers a principled way of identifying the different senses of a word. This is useful for the investigation of L2 learner knowledge of polysemous words because if a learner knows one sense, one does not want to assume she has knowledge of other senses. This assumption cannot be maintained if there is too much overlap between definitions. However, the NSM approach may make claims about the nature of the mental conceptual system which appear idealistic and are not easily maintained with certain examples of language use.

Hanks (2013) has argued that the lumper/splitter debate is probably unresolvable. What is more important for the question of L2 learning is how research from the field of cognitive linguistics can be incorporated into an empirical design. Indeed, there has been a certain amount of research into teaching L2 learners new senses of polysemous words (Boers, 2013; Boers & Lindstromberg, 2007; Boers & Lindstromberg, 2008; Csábi, 2004; Khodadady & Khaghaninizhad, 2012; MacLennan, 1994; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). In this research, the theory from cognitive linguistics has been used for two purposes. First, it has been used to identify senses, but it has also been used to support the claim that L2 learners can be taught new senses of polysemous words based on semantic similarity expressed by the category of the polysemous word.

2.3.4 The role of context in sense identification and word interpretation

The preceding section discussed why one polysemous sense can be considered distinct from another. However, another issue to be addressed is how a specific sense is identified from another in a given context. This is not always as apparent as it might seem, as can be seen in the following sentence,

(6)   She is from the same class as me.

The word *class* is ambiguous because it could refer to either a teaching period or a social class. In this case, the linguistic context of written or spoken information does not distinguish between the two senses. However, if the sentence were spoken by someone as part of a conversation, then the meaning might be easily understood through the situational context of information received physically, socially or pragmatically. The two senses of *class* in (6) are said to express attentional autonomy because one sense exists independently of the other due to the listener's focus of attention (Croft & Cruse, 2004).

However, in the situation of assessing L2 learners' vocabulary knowledge, it would be rare to rely on situational context to differentiate between different senses of a word. The question then is how will a learner make use of the lexical context when

she encounters a word. The following sentence, presented in the first chapter, will serve for a discussion of the learner's process of interpretation:

(7)   Let's go outside and get some fresh air.

The linguistic context in this sentence provides enough information for the meaning of *air* to be unambiguous. As such, the sense of *air* in this sentence is said to express compositional autonomy. There are two reasons why an L2 learner would correctly identify this sense of *air* as GAS. First, the learner might interpret the meaning at the conceptual level, perhaps through translation. The concepts associated with the other words and phrases in the sentence might engage with a portion of the total meaning of *air* in order for the correct sense to be selected and the other senses inhibited[6]. However, the learner might also interpret the sentence at a lexical level. She might have encountered many of these words together in her L2 on other occasions, leading to strong lexical associations which might facilitate the correct selection of the concept.

It is perhaps more straight forward to address interpretation due to lexical associations because we can turn to a corpus as evidence of the frequency of the collocations between the words. The phrase *fresh air* is the second most frequent collocation in the British National Corpus (BNC) of 100,000,000 words, occurring 480 times[7]. However, *fresh air* could be construed differently in the following idiom,

(8)   She was a breath of fresh air.

In this sentence, *fresh air* is part of an idiom referring to a person's vivacity. Be that as it may, the idiomatic interpretation is unlikely in sentence (7) because of other linguistic context. The word *outside* and the phrase *get some* occur 100 and 25 times in the BNC within four words of the target word air. These collocations would

---

6.   Interpretation at the conceptual level is returned to in the General Discussion in Section 11.3 with reference to the Revised Hierarchical Model (Kroll & Stewart, 1994).

7.   The most frequent collocation is *air force*; however, its frequency is due to the large number of news texts used in the corpus.

distinguish the sense from the idiom in (8) and help lead the learner to the correct interpretation.

One concern with strong collocations is that they may form a separate lexical unit that is independent of the senses associated with the specific words. Such a pattern of learning would be consistent with Wolter's meaning-last model of L2 learning, whereby individual words are first learned as larger lexical chunks and only later isolated and attributed with specific semantic content (Wolter's model was discussed in Section 2.2). The meaning-last model is likely applicable to the idiomatic use of *air* in sentence (8). However, it is a question whether the same idiomatic conditions apply to following sentence as well,

(9)   Air travel was growing rapidly.

The words *air travel* form a strong collocation, occurring 107 in the BNC. One could argue that the collocation refers to a concept that is separate from the sense of *air* referring to PLANES. However, there are a number of other strong collocations also referring to the sense of PLANES: *air force* (occurring 1033 times), *air transport* (occurring 114 times), *air crash(es)* (occurring 81 times), and *air fare(s)* (occurring 72 times). Given the number of strong collocations associated with this sense, it is likely that a learner would form some type of semantic rule for interpreting *air* as a modifier with the sense of PLANES.

While the context provides the information for distinguishing one sense from another, it doesn't directly distinguish between homonyms and polysemes. This is because both types of senses can express attentional autonomy. In order for senses to be considered polysemous, there needs to be a semantic connection between the words. An example of a polysemous word with attention autonomy was presented using the word *class* above in (6). In this sentence, the two senses are related by the shared idea of a *class* as group of individuals having some properties in common. Despite being semantically related by a shared idea, the two senses are distinct at the

31

level of attentional autonomy. A homonym will also express attentional autonomy, as in the following example taken Croft and Cruse,

(10)   We finally reached the bank. (margin of river, financial institution)

In this case, the two senses of bank are from different etymological roots and are not semantically linked by a shared idea. Thus, both polysemes and homonyms can express attentional autonomy, but they are distinguished at the conceptual level, as described in the preceding sections.

The senses of *air* and *class* presented above have all been examples of what are referred to in cognitive linguistics as full senses (Croft & Cruse, 2004). A full sense has a clear boundary separating it from other senses, so that when a language user is attending to one full sense, other senses are inhibited. Not all senses are full senses, however. In some cases, two senses of the same word can exist in the same utterance, unified under a general Gestalt. For example, the sense of *fresh air* in (7), referring to GAS, is different from the sense of *air* in the following sentence, referring to SPACE,

(11)   He threw the ball through the air.

While these two sentences differ from one another, they can be unified in a single sentence,

(12)   The air was so polluted I couldn't see through it.

In this sentence, *air* refers to both GAS and SPACE without ambiguity. Croft and Cruse refer to senses at this level of autonomy as facets: in some contexts facets are distinct from one another, while in others they are unified. However, facets are less interesting from the perspective of L2 learning because if L2 learners knew one facet-sense, it would be surprising if they didn't know the other.

A final point to make about lexical context is that it can also add semantic detail to the meaning of a lexical unit without distinguishing a distinct sense. The context in the following sentence identifies *my cousin* as a woman,

(13)   My cousin kept her maiden name.

Whether a cousin is male or female is not considered an essential characteristic of the word's meaning. According to the cognitive linguistic approach, certain semantic content of a lexical unit is stored in order to distinguish one distinct sense from another. Other semantic content which is not involved in distinguishing one sense from another, as in the gender of a cousin, isn't specified in the lexicon. Croft and Cruse refer to contextual modulation as this process of adding detail after the sense has been accessed. Elsewhere this level of semantic specification is referred to as vagueness (Geeraerts, 1993).

In this chapter, I have reviewed many of the models from cognitive linguistics about how polysemous senses are related to one another. However, much of the evidence for these claims comes from the introspective judgements of the linguist researchers and not from empirical evidence. Not surprisingly, the validity of these judgements has been strongly debated (Croft, 1998; Sandra & Rice, 1995; Tuggy, 1999). Subsequent to this debate there has been a fair amount of research from the field of psycholinguistics on the the storage and processing of polysemous senses with native speakers and bilinguals. The results of these studies can lend support or qualify the claims made in the field of cognitive linguistics.

Since the research into L2 teaching of polysemous senses is also based upon the claims from cognitive linguistics, the research from psycholinguistics can also support or qualify the validity of the L2 teaching methodologies. In the next section I will review the psycholinguistic research into polysemous senses with the aim of supporting research into L2 learning and teaching of polysemous senses.

2.4 Psycholinguistic perspectives on polysemy

Psycholinguistic experiments into the nature of polysemous words in the mental lexicon can be organised into two broad categories: one group investigates psychological process of using polysemous words and the other group deals with the psychological structure or representation of polysemous senses. As Sandra and Rice explain, "Whereas a process refers to a principle of cognitive functioning (i.e., a way of interacting with information), a structure refers to information that is stored in memory," (Sandra & Rice, 1995, p. 100).

This distinction is important when assessing the relevance of the results to L2 learning of polysemous words. Experiments into how polysemous words are structured are conducted online and the results are used to show whether polysemous senses have separate or shared representation in the mature mental lexicon of the native speaker. The structure of the L1 mental lexicon becomes either a benchmark for the developing L2 lexicon or a point of difference between L1 and L2 speakers.

Experiments into how polysemous words are processed are conducted offline and involve more explicit decision making on the part of the language user. Using these experiments, researchers can investigate whether general language users recognise the same similarities between polysemous senses as the expert linguists. Experiments into language processing are also used to see if semantic similarity between polysemous senses can facilitate learning novel language. The results of the language learning experiments with mature L1 speakers allow for a comparison with language learning experiments with L2 learners.

2.4.1 The representation of polysemous senses

Researchers in the field of psycholinguistics have employed three different methods for identifying polysemous senses: a dictionary based method, a method based on native speaker similarity judgements and a method based on linguistic principles. In other words, the psycholinguistic research provides two other methods of identifying polysemes in addition to the method employed by cognitive linguistics

34

reviewed above. The research from psycholinguistics offers possible insight into two issues, whether there is empirical evidence for shared mental representation of polysemous senses and whether the linguistic method is the preferable method for identifying polysemous senses.

The first method of identifying polysemous senses was based on separate listings in the dictionary Oxford English Dictionary (OED). This method was employed by Klein and Murphy (2001) and Foraker and Murphy (2012). In the OED, there are superordinate and subordinate levels of senses. Different polysemous senses were identified if they were listed under different superordinate levels. For example, the first two superordinate sense levels for the word *paper* are "I. Senses relating to the material" and "II. Paper bearing writing, illustrations, etc." By this classification *paper* [MATERIAL] and *paper* [NEWSPAPER] are different senses sharing the same core sense. The researchers did not compare senses at the subordinate level, such as *paper* [ESSAY] and *paper* [NEWSPAPER].

Klein and Murphy (2001) based their experiments on the claim that if polysemous senses share the same core sense, then when the core sense is primed, these senses should be accessed in the same amount of time. If the priming benefits only one sense and not the other, then this is taken as evidence of separate representation for polysemous senses. They used a sense judgement task with two categories, same senses ('daily paper' vs. 'liberal paper') or different senses ('daily paper' vs. 'wrapping paper'). Participants were primed with the first phrase of the pair ('daily paper') and then they made a sense judgement on the second phrase ('liberal paper' or 'wrapping paper'). The different-sense category was responded to slower than the same-sense category, which was taken as evidence against a shared core representation.

Foraker and Murphy (2012) investigated reading comprehension time using the same choice of polysemous senses as Klein and Murphy. They included the factor of dominance in their design; the dominant sense is the sense language speakers most often produce for a polysemous word form. In Foraker and Murphy's study,

participants read stimuli sentences that biased either a dominant, subordinate or neutral context. They were then measured on their reading speed for second sentences that biased either the dominant or subordinate context. Participants read the dominant-supported sentence faster, and Foraker and Murphy argue this also is evidence against a shared core representation.

After the dictionary-based selection of senses, other studies have identified polysemous senses using similarity ratings provided by native speakers (Brown, 2008; Klepousniotou, Titone, & Romero, 2008; Rodd, Gaskell, & Marslen-Wilson, 2002). By this method, a pair of polysemous senses rated more similar to each other would be identified as polysemous (highly or moderately-overlapping) and pairs rated less similar would be identified as homonymous (non-overlapping). Klepousniotou et al. (2008) replicated Klein and Murphy's methodology (2001), but instead of senses based on dictionary entries, they used highly, moderately and non-overlapping senses, based on similarity ratings provided by native-speakers. They found faster judgements for polysemous word pairs with highly overlapping senses compared to ambiguous words with moderately to non-overlapping senses. This indicates that there is evidence for more shared representation for two senses which are more similar to each other. Brown (2008) found similar results using four categories of sense pairs: unrelated, distantly related, closely related, an same sense. Brown also found response times were faster as the pairs were more similar. Rodd, Gaskell, and Marslen-Wilson (2002) also used similarity ratings to distinguish between senses. Rather than comparing polysemous senses for the same word, Rodd et al. compared the response times for polysemous words to the response times for homonyms and unambiguous words in a lexical decision task. When compared to unambiguous words, participants responded significantly slower to homonyms but not to polysemous words.

All three of these studies used similarity ratings to categorise how similar senses were to one another. All three found faster response times for senses with greater similarity scores or words whose senses expressed greater similarity. In light of the results of these studies using the similarity-judgement method, the conclusions to

the studies using the dictionary selection method become less clear. Both Klein and Murphy (2001) and Foraker and Murphy (2012) argued that slower response times for the different senses indicate that they don't share the same mental representation as the core sense. This conclusion is binary, either the senses share the same representation or they don't. However, the results of the studies using the similarity-judgement method are best understood as gradient with response times increasing as the senses are rated more different from one another. The case for shared or separate representation isn't clear from the results of these studies. What the studies do show is that explicit judgements of similarity are corroborated by evidence of more implicit knowledge.

After the dictionary entries and similarity judgements, the third method of sense categorisation is based on the descriptions proposed by cognitive linguistics. Sandra and Rice (1995) also conducted an experiment to measure the effect of priming on polysemous prepositions. They used senses defined by Lakoff (1987) according to semantic similarity based on image schemas and propositional content. Participants were asked to judge whether a prepositional phrase was acceptable or not. Half of the participants were primed with the core sense of the preposition and half were not. The participants made more errors in the primed condition, and the error rate increased for the more extended senses. The results indicated that the core sense increasingly inhibited the activation of the senses as the senses became more extended. These results support the proposition of the radial structure because the extended senses are proposed to share fewer and fewer similarities with the core sense.

This radial structure was further supported by Klepousniotou (2002) and Klepousniotou and Baum (2007) who also conducted online experiments with polysemous words. In their case, they took nouns as their polysemous words and again the senses were identified by linguistic principles. Specifically they compared homonyms and two categories of polysemes (metonyms and metaphors). The metonyms and metaphors expressed specific relationships. An example of the metonym category were senses based on a COUNT/MASS relationship as in *chicken*

[ANIMAL] (COUNT) vs. *chicken* [FOOD] (MASS), and an example of the metaphor category were senses based on a PART/OBJECT relationship as in *mouth* [BODY] (PART) vs. *mouth* [CAVE] (OBJECT). Klepousniotou (2002) presented her participants with an auditory prime sentence that biased one of the senses. The participants then made a visual lexical decision on whether the target word was a real word or a non-word. The researcher found that there was an advantage for polysemous words over homonyms, and within the class of polysemy, metonyms showed an advantage over metaphors. This is evidence that polysemous senses share greater representation in the mental lexicon than homonymous meanings. Whether there is a core sense or not is left in question. Klepousniotou and Baum (2007) conducted auditory and visual lexical decision tasks without priming. In the auditory condition, the researchers found a polysemy advantage for words in both the metonym category and the metaphor category; while in the visual condition, they found an advantage for metonyms. These studies support Klepousniotou's earlier findings (2002).

The evidence from Klepousniotou's and Klepousniotou & Baum's studies is rationalised according to the following argument. The responses to the homonyms are slower because they express two distinct representations of meaning. When the participant makes his lexical decision, these two representations are in competition with one another, which slows the participant's decision. The responses to the polysemous senses are faster because they share in the same representation and so there is less competition between these senses.

The research from psycholinguistics has shown that when polysemous senses are identified by similarity-judgements or linguistic principles there is evidence for shared representation of polysemous senses. Whether that shared representation is a specific core sense or a more distributed representation is unclear. Nevertheless the psycholinguistic research does support how similarity between senses is identified by the cognitive linguistic method of linguistic description.

For research with L2 learners, it is preferable to identify polysemous senses by linguistic principles, rather than by similarity judgements. While native speakers can

be expected to know most of the available senses for a polysemous word, this is not the case for L2 learners. As a result, L2 learners couldn't be expected to provide similarity judgements themselves. The judgements of native speaker could be used in researching L2 knowledge; however, these results would only indicate how similar the learners were to the native speaker norm.

In contrast, by identifying senses according to linguistic principles, the researcher is able to make claims that are falsifiable. The linguistic principles allow the researcher to say that one sense is easier to learn than another because it is more similar to the core sense than the other. It may be possible for a well designed experiment to confirm or refute this claim.

2.4.2 How native speakers process polysemous senses

The psycholinguistic research reviewed up to now has focused on adult L1 speakers as participants in online experimental tasks. These tasks have focussed on the representation and storage of polysemous senses. In contrast to L2 learners, the mental lexicons of the adult native speakers are stable to the extent that they have already learned the polysemous words in question. Since the mental lexicons of L2 learners are still developing, there is a question whether the L1 mental lexicon is comparable.

In contrast to the online experiments, research into how native speakers process polysemous senses in offline experimental tasks can offer different implications. The offline experimental tasks are designed to be successfully completed if the subject finds the relationship between senses meaningfully similar. Since the judgements in these experiments are more conscious and explicit, the results may be more applicable to L2 learners.

Sandra and Rice (1995) conducted two offline experiments using spatial prepositions as the experimental polysemous words. They investigated whether native speakers judged the similarity of prepositional phrases in accordance with the radial network described in cognitive linguistics. In a sorting task, participants organised prepositional phrases along granular distinctions that were consistent with the

proposed radial category. In a second experiment, two groups of native speakers rated prepositional phrases for similarity to either a central spatial sense or a central temporal sense. The results showed that more prototypical senses were rated more similar to both the central spatial and the central temporal senses. The more extended senses, however, were often rated similar to only one of the central senses, but not both.

These studies show that the non-expert, general language user can explicitly recognise the semantic similarities between senses that had been proposed by expert linguists. Given this finding, L2 language learners may also be able to recognise the semantic similarity between senses when processing the meaning of polysemous words. One caveat to this application to language learners is that the native speakers have the benefit of already knowing the meaning of the senses. Other experiments into the processing of polysemous senses have focussed on whether native speakers can learn novel polysemous senses which were created using linguistic principles. Since these experiments focus on language which the native speakers do not know, these studies are more comparable to the task faced by the language learner.

Murphy (1997) investigated native speaker learning of "new words"and Rodd, Berriman, Landau and Lee (2012) investigated native speaker learning of "novel language". While "new words" refer to imaginary word forms that were created for the study and assigned with different meanings, "novel language" refers to the apparently implicit process by which existing words in a language obtain new meanings. These studies attempt to show that learning new polysemous senses is possible for speakers with a developed mental lexicon.

Murphy created two experiments using new words to look at whether novel use of language could be explained by extension from a core sense. Murphy doesn't give details of his participants, but we can assume that he used English native speakers. In both experiments, the participants were presented with paragraphs that used new words in a rich context to establish the words' meaning. The following quotation is

the paragraph context for the core sense to one of the new words. In the materials a picture is included with the paragraph,

> When there is excess water in the soil due to excessive rain, Quinese people use a *wift* to help remove it. The pipe is inserted in the ground, and the swirling blades pull excess water into the pipe. The small tube is an air intake valve. Thus, the *wift* is a kind of dehydrator for soil. (Murphy, 1997, p. 250)

The first paragraph established the core sense meaning of the new word and subsequent paragraphs established meanings extended from the core sense.

In the first experiment, the core sense of the new word referred to a tool for some imagined purpose. There were three different conditions for the experiment. In the Close condition, the participant read the core-sense paragraph, an extended-sense paragraph (A), and a further-extended-sense paragraph (A') which was semantically related to the extended sense. The participant rated the acceptability of the further extended sense (A') on a scale of 1-7. In the Distant condition, the participant read the core-sense paragraph, an extended-sense paragraph (A), and a further-extended-sense paragraph (B') which was **not** semantically related to the extended sense. Again the participant rated the further extended sense (B') on the same scale. In the No-Earlier-Use condition the subject read the core-sense paragraph and the further-extended-sense paragraph (A') and the participant rated the further extended sense (A').

 The mean acceptability rating for the Close condition was 4.6 out of 7. The further extended sense was rated more acceptable in the Close condition than in the Distant condition (3.7 out of 7) or the No-Earlier-Use condition (3.6 out of 7). The difference in the acceptability ratings indicates that speakers develop polysemous categories by extending senses based on semantic similarity.

In the second experiment, Murphy again used new words, but here he compared conventional extensions of polysemous senses to unconventional extensions. Researchers working with cognitive linguistic theory have identified certain

41

conventional ways in which senses are extended from core senses (Croft & Cruse, 2004). Murphy looked at five of these conventional extensions of polysemous senses. An example of one of the conventional extensions was the COUNT/MASS distinction, by which one can both "raise chickens" and "eat chicken." In the unconventional extension, *chicken* would refer to the place where chickens are kept. Real words, like *chicken*, were not used in the study; as in experiment 1 new words were used to control for the effect of the extension. In the results, the mean acceptability rating for the conventional extensions was 6.2 out of 7, whereas the mean rating for the novel extensions was only 2.9 out of 7. The effect in this experiment was far greater than the effect of relatedness in the first experiment. The results provide evidence for the acceptability of extensions along lines of conventions identified by linguists. In other words, the results show that native speakers without a linguistics background find conventional extensions more meaningful than other extensions.

The results from the second experiment corroborate and extend the findings of Klepousniotou (2002) and Klepousniotou & Baum (2007) who found greater evidence for shared mental representation of conventional metonyms with the core sense than for other types of relationships. The strength of results for the second experiment using the same type of conventional metonyms used in Klepousniotou (2002) and Klepousniotou & Baum (2007), raises the possibility that extensions which show greater evidence of shared representation also lend themselves to greater ease of learning. This is just a conjecture, but it is worth noting as a possibility. On a more critical note, there is a question of how relevant conventional extensions are to L2 learners. The senses for many and perhaps most polysemous words are not conventional metonymic extensions. Take for example the word *line*, whose core sense is "any long two dimensional mark," and includes extensions such as "a line of music as in 'a bass line'," (2011). Even though the relationship between the core and extended sense is apparent, the relationship doesn't fit the type of conventional extension examined in experiment two, and yet this is the type of extension L2 learners are expected to learn. It is this more common type of polysemous extension

that Rodd et al. (2012) investigated, again with native speakers learning novel language.

The research of Rodd et al. (2012) is relevant to this strand of research for two reasons. First, like Murphy (1997) they investigated native speaker learning of novel polysemous language, but unlike Murphy, they looked at metaphoric extension of language. Second, like the studies on the representation of polysemous meaning (Foraker & Murphy, 2012; Klein & Murphy, 2001; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008) their research includes a lexical decision task. In their experimental design, Rodd et al. presented a word with a single dominant meaning, such as *ant*, to L1 participants in a paragraph-length context. They wrote the paragraph so that the word was used with a novel meaning. For example, the word *ant* was used in the context of a tiny recording device. These novel meanings are referred to as related items. The authors contrasted these items with unrelated items which were created by paragraphs where the novel meaning of the word had no relation to the semantic properties of the dominant meaning of the word. For example, as an unrelated item, *ant* was used in the context of lines painted on the face. The authors investigated the effect of relatedness by comparing results of related and unrelated items across three tasks: subjective semantic ratings, a cued-recall of semantic properties, and a non-priming lexical decision task. The lexical decision task was used to see if participants identified a word with a related novel meaning more quickly than a word with an unrelated meaning. Faster response times for the related condition were taken as evidence that the novel meaning had been incorporated into the existing semantic representation in the mental lexicon.

The lexical decision task was conducted across two experiments. In one experiment, the participants read the paragraphs once a day for six days and then conducted the lexical decision task on the 7th day. In the other experiment, the participants completed semantically engaging worksheets once a day for four days and conducted the lexical decision task on the 5th day. The participants responded significantly faster for the related items in the semantically-engaging condition (p < .05), whereas there was no significant difference in response time in the daily-

reading condition ($p < .9$). These results indicated that semantic consolidation benefits the learning of related meanings for the same word form over unrelated meanings. The results also supported the findings of Klepousniotou (2002) and Klepousniotou and Baum (2007) who reported faster responses to words with related ambiguous meanings over words with unrelated meanings. Furthermore, the findings of those studies also included metaphoric extensions of the core sense just as the related senses in this study are metaphoric extension of the dominant meaning.

The results of Rodd et al. (2012) lend more support to the hypothesis that semantic relatedness to existing knowledge of the word facilitates learning polysemous senses. Murphy (1997) found that conventional extensions supported strong learning over distant or unconventional senses. In contrast, Rodd et al. (2012) did investigate the metaphoric extensions of the novel language. This metaphoric language is closer to the variety of extension that L2 learners will encounter when learning L2 polysemous words.

Both Rodd et al. (2012) and Murphy (1997) dealt with native speakers learning novel language in their own L1. Learning a second language poses other complicating factors upon the influence of semantic relatedness. One of these factors is the influence of the learner's L1 on L2 acquisition of new senses. When learning new senses in one's L1, a speaker has a very broad knowledge of the variety of contexts and meanings of a different word form. With this broader knowledge there is more opportunity to make connections between the known uses of the word and the novel context. In learning a new sense in one's L2, the learner may have a limited knowledge of the variety of contexts and meanings in which the word form is used. This means she'll have less knowledge of how the word can be used from which she could make a novel extension.

Another possible factor affecting L2 learning is the influence of the learner's L1. In the L1, there is a more direct link from the word to the concept and its semantic features. In contrast, access to the concept from an L2 word form may travel by way

of a translation path through its L1 equivalent. If the L1 equivalent does not share the same extended sense as the L2 word, then the learners' direct translation strategy would have failed in the attempt at comprehension. There has been a strand of research which has looked at whether L2 learners have direct access to semantic concepts or whether the access is mitigated by a translation path through an L1 equivalent.

2.4.3 L2 psycholinguistic perspectives: Translation ambiguity studies

The research reviewed on the representation of polysemous senses deals solely with native speakers and not with second language learners. In psycholinguistics the majority of research into L2 ambiguity has focussed on interlingual homophones (Degani & Tokowicz, 2010b). This refers to the situation where one word form is shared across two languages. Interlingual homophones have been used to investigate the independence of the L2 lexicon from the L1 lexicon. However, these studies have necessarily dealt with related languages sharing the same orthography. In the experiments presented later in this thesis, the first language of the L2 learners is Arabic. Since Arabic and English are unrelated languages with different orthographies, bilingual studies between these two languages do not lend themselves to research using interlingual homographs. More relevant has been research into translation ambiguity. Translation ambiguity occurs when two concepts represented by a single word in one language are represented by two words in another language. There are two directions for translation ambiguity, either the L1 word form is ambiguous and the L2 is represented by two word forms, or the L2 word is ambiguous and the L1 is represented by two word forms. The studies dealing with L2 ambiguity are most relevant to this thesis; however, the studies of L1 translation ambiguity offer insight into the effect of semantic transfer from the L1 to the L2.

2.4.3.1 The effect of L1 ambiguity on L2 comprehension

Jiang (2002, 2004) investigated whether there was evidence that L2 learners develop new semantic content for L2 words rather than simply mapping the semantic content of the L1 (the L1 lemmas) to L2 word forms. Jiang looked at L1 translation

ambiguity with advanced Chinese and Korean learners of English. He presented the learners with three groups of English word pairs. The first group of English word pairs were semantically related and ambiguous in the learners' L1; for example, 'problem' and 'question' translate as *wenti* in Chinese. The second group of word pairs were semantically related and unambiguous in the learners' L1; for example, 'painter' and 'artist' translate as *huajia* and *yishujia* in Chinese. The last group of English word pairs were semantically unrelated and unambiguous in the learners' L1. The experimental task was for the learners to decide whether the English word pairs were semantically related or not. The results for correct responses were recorded in milliseconds.

The results indicated that both the Chinese and Korean learners of English were faster at identifying the word pairs as semantically related if they translated into the same L1 word form. Jiang concluded that this was evidence that the L1 lemma continued to influence the semantic decision even among very advanced learners.

Jiang also recorded the response times of monolingual native English speakers as a control group. The native speakers did not know Chinese or Korean, and their results can act as a check on whether the semantic relatedness between word pairs were equally balanced. The native speakers responded significantly faster to the English word pairs which translated into a single Chinese word than to those pairs which translated into two different Chinese words. As the native speakers did not know Chinese, this result indicated that the semantic similarity between the two groups of word pairs may not have been equally balanced. The native speakers' response times to the word pairs for the Korean participants were not significantly different, indicating a balance in semantic relatedness.

In comparison to the native speakers, the learners committed more errors for the semantically related English word pairs that translated into different L1 word forms. This meant they did not recognise the semantic similarity which the native speakers had identified. This result indicated a limitation of learners' semantic understanding of L2 lemmas in contrast to native speakers. Thus, while L2 learners can map form to

meaning independently of their L1 mapping structure, their understanding may be incomplete for even advanced learners.

As expected, the English language learners took relatively more time to identify that a word pair was not related. However, the native speakers of English also took more time to respond to unrelated word pairs. Jiang didn't give inferential statistics for the response time to the unrelated word pairs, but the proportion of extra time which the learners needed to respond that a pair of words were unrelated was comparable to the proportion of extra time the native speakers needed.

In sum, Jiang found that L2 learners recognised semantic similarity faster between two L2 words which translate into the same L1 word form. This result indicates that L2 learners might find it easier to learn L2 polysemous senses if the word is similarly polysemous in the learner's L1. In addition to this benefit, the learners were still able to recognise semantic relatedness for words that translated into different L1 word forms, albeit they showed less understanding of semantic relatedness than the native speakers did.

Following Jiang (2002, 2004), Elston-Güttler & Williams (2008) also investigated the influence of L1 translation ambiguity on L2 processing. Specifically, they looked at how ambiguous L1 words affected the acceptability of word use in the L2. For example, the German word *Blase* translates into both *bubble* and *blister* in English. Elston-Güttler & Williams presented German-learners of English with English sentences using an unacceptable translation, such as 'His shoes were uncomfortable due to a bubble.' Their stimuli included four groups of ambiguous translations, highly/moderately related translations and nouns/verbs. (See Table 2.2 for examples.)

Table 2.2
Examples of critical items (Elston-Güttler & Williams, 2008, p. 173)

German word	Related sentences	Control sentences
Highly related:		
Noun: Blase	His shoes were uncomfortable due to a bubble.	She was very hungry because of a bubble.
	With the chewing gum the child made a blister.	With a desk chair the secretary made a blister.
Verb: tragen	The heavy suitcase was difficult to wear.	The hard work was difficult to wear.
	In winter a heavy warm coat is practical to carry.	The new car was comfortable to carry.
Moderately related:		
Noun: Schlange	A frightening thing to see in the forest is a queue.	A nice thing to see hanging on the wall is a queue.
	She had to wait several minutes due to the snake.	She worked very hard due to the snake.
Verb: reiben	Mary's back hurt so the massage expert began to grate.	Mary's hair was too long so the woman at the salon began to grate.
	The hard cheese was easy to rub.	The new dining room table was easy to rub.

In an anomaly detection task, the participants were measured for speed and error in response to whether the target word was used acceptably. The researchers found that the learners responded to all but the moderately related nouns with significantly more errors and longer response times than native speakers. Surprisingly, the researchers don't discuss why the moderately related condition should be different for nouns. In general they argue for a continued influence of L1 lemmas for this category of ambiguity. I would argue that highly related nouns have a shared representation in the L1 while the moderately related nouns do not. This structure might result in greater online disambiguation for highly related nouns, and shorter, more direct retrieval for moderately related nouns, resulting in shorter response times. In contrast to nouns, verbs may always be more difficult for learners in acceptability tasks because verb meanings have more language-specific features,

making conceptual overlap less likely between languages (Elston-Güttler &
Williams, 2008, p. 183).

While Jiang (2002, 2004) and Elston-Güttler and Williams (2008) provided evidence
that L1 lemmas continue to influence L2 comprehension, these studies employed
advanced language learners who had already learned the L2 word forms under
investigation.

2.4.3.2 The effect of L2 ambiguity on L2 comprehension

The translation ambiguity studies investigated so far have looked at ambiguity in the
learner's L1. These studies have shown the continued influence of the L1 lexicon on
the semantic development of the L2. A few studies have also investigated L2
translation ambiguity, one of which, Gathercole and Moawad (2010), investigated
the effect of ambiguity in both the L1 and the L2. Specifically, they investigated the
influence of three categories of translation ambiguity for Arabic and English
monolinguals and early and late bilinguals with Arabic L1. The three categories
were classical (*fingers* and *toes = asabie*), homophones (*sun* and *son*) and radial (*cap* of
a pen and baseball *cap*). They investigated translation ambiguity in both the L1
(Arabic) and in the L2 (English), making for three categories of ambiguity with two
levels (Arabic and English). Their method of investigation was a picture
categorisation task using the English and Arabic words as prompts. The participants
had to decide which of several pictures could be labelled by a given term. They
found that classical categories expressed the greatest interlanguage influence.

Gathercole and Moawad's research is limited by the number of test items the
researchers used in their instrument. In total they only used 18 test items, which
meant that there were only 3 items to each category by language. The items which
are most relevant to learning L2 polysemous senses are only those items where the
L2 (English) expresses a single word form and the L1 (Arabic) expresses two
concepts, and of those items, primarily the radial and classical categories. This
comprises only 6 items, which is a small sample from which to generalise. One
problem with such a small sample is that other factors, such as part of speech or type

of figurative extension, may have influenced the results. For example, the radial concepts with English ambiguity included both nouns and verbs (*hand*, *cap*, and *to smoke*) as well as both metaphoric extension (*hand:*BODY and *hand*:CLOCK) and alternation in transitiveness (*'smoke* cigarette' and 'he *smokes')*. It is unclear to what extent these secondary factors might have played in the results.

It might be reasonable to generalise from the classical category because the examples provided can be understood through the logic of meronymy: *fingers* and *toes* are all digits, *a watch* and *a clock* are both time pieces, and *fishing* is specific type of *hunting*. The difficulties for learning one example in the classical category are likely to be similar to learning another example in that category. There is less consistency in the radial category because the examples express a wider variety of relationships. A *stovetop burner* seems to be related to an *eye* by its shape; in contrast, *to pay* seems to be related to *to push* by the action of giving money across a shop counter. This type of relationship has been described as *motivated* but not predictable (Lakoff, 1987). It is a question whether the motivated relationship that links two concepts can facilitate L2 learning (Boers, 2013). It is difficult to be confident about the results for this category which includes only three items. The radial category has been identified as the structure underlying the relationship between polysemous words. A wider investigation of this type of relationship is needed because of the variety of semantic relationships expressed by different examples.

Degani and Tokowicz (2010a) conducted a study with L1 English speakers who had no prior experience with Dutch, the L2 language. Degani and Tokowicz developed three groups of English-Dutch translation pairs: unambiguous pairs, meaning-ambiguous pairs and form-ambiguous pairs. Examples are presented in Table 2.3:

Table 2.3
Example stimuli and definitions by condition for Degani & Tokowicz (2010a, p. 300)

Condition	English word	Definition(s)	Dutch translation(s)
Form-ambiguous	sky	1. the region of the clouds or the upper air	1. lucht
			2. hemel
Meaning-ambiguous	change	1. the result of alteration or modification	1. verandering
		2. coins of small denomination	2. wisselgeld
Unambiguous	arrow	1. a mark with a pointed end used to indicate a direction or relation	1. pijl

The researchers conducted three training sessions and three testing sessions. In each training session the English speakers learned the Dutch words by reading the word form and an English definition. Each Dutch word form was repeated four to eight times. The learners were tested by orally producing an English translation to a Dutch word and by visually deciding whether a Dutch word was a correct translation of the English word.

Degani and Tokowicz found that unambiguous words were responded to with fewer errors than the ambiguous words. However, the differences, while significant within participants ($p < .05$), were not that substantial. In the third testing session the unambiguous words were responded to 8% and 11% more accurately than the meaning and form-ambiguous words in the English translation task. Furthermore, there was no conclusive difference in the reaction times between the meaning-ambiguous and unambiguous words in the final delayed test ($p > .10$). This study indicates that ambiguous L2 words are slightly more difficult to learn than ambiguous L2 words. However, what this study doesn't address is whether L2 learners can correctly interpret the meaning of an L2 word in context. This is potentially a major challenge for L2 learners and the research of Paribakht (2005), discussed below, is related to this question.

Paribakht (2005) investigated how lexicalisation of a concept in the L1 facilitates inference of meaning in the L2. She presented Farsi-speaking learners of English with paragraph length texts containing target words which the participants were not expected to know. Half of the target words were lexicalised in Farsi and half were not. Using think-aloud protocols to investigate their inference processes, Paribakht found that participants were three times more successful at inferencing lexicalised target words over non-lexicalised target words.

While this study does not specifically deal with translation ambiguity, it does support the evidence from those studies of continued influence from the L1 lexicon. What is of more interest to this thesis is the 'homonymy' inference strategy employed by the participants. This strategy is identified when the learner makes a semantic connection between the form of the target word and another word with a similar form. Paribakht gives the following example:

> To my mind [**towing**] means (to tiptoe). The word [tiptoe] means 'to walk on your toes', but here I took the word [**towing**] to be a verb meaning (to tiptoe). (Paribakht, 2005, p. 743)

By this strategy, the learner is applying the semantic content of a similar word form to make a meaningful inference given the context of the paragraph. It should be noted that the homonymy strategy was only employed in 1.7% of the inference cases. In the large majority situations (60.1%), the learners employed a sentence meaning strategy. Moreover, the term 'strategy' might be inappropriate because this is really a classification of a type of error. This is the type of error one can expect a learner to employ when she encounters a known polysemous word in a novel context.

Degani, Prior and Tokowicz (2011) compared the similarity ratings of bilinguals and monolinguals for concepts that express L2 translation ambiguity. The researchers were able to investigate bidirectional transfer in this study by including both Hebrew-English and English-Hebrew bilinguals, with English monolinguals as a control group. This study contributes to the question of whether similarity between

senses in L2 polysemous words can lead L2 speakers to some independence from their L1 lexicon.

The participants were asked to rate the similarity between pairs of words on a scale of 1-7. The word pairs were organised by two variables, each with two levels: translation type (shared translation and different translation) and relatedness (related and unrelated). The word pairs with a shared translation translated into the same word in Hebrew and the different translation pairs translated into two different words in Hebrew. Relatedness was defined according to semantic ratings of native English speakers. Related pairs, such as *arch* and *rainbow*, were rated above 4 on a 7-point scale by native speakers, and the unrelated pairs, such as *brother* and *fireplace*, were rated below 3.

In Degani et al. (2011) shared-translation pairs were rated significantly higher than different-translation pairs *(p < .01)* by both bilingual groups, for English-Hebrew and for Hebrew-English, but there was no corresponding difference for the English monolinguals (*F* < 1). Degani et al. argue that not only was Hebrew affecting the semantic ratings as an L1, but Hebrew was also affecting the ratings as an L2. The researchers proposed that the higher similarity ratings indicated that for the bilinguals, the "pairs that share a translation become more strongly interconnected than pairs with different translations," (Degani et al., 2011, p. 24) and as a result are co-activated when either concept is accessed.

One result from the study was not predicted. The bilinguals also rated the unrelated word pairs for shared translations more semantically similar than the unrelated, different translations. The authors illustrate the result with the unrelated word pair *tool-dish* which translates into the Hebrew word *kli*. Unrelated word pairs with shared translations such as this can often be described as homonyms. The higher similarity ratings for these word-pairs run counter to research involving lexical decision tasks, where bilinguals responded more slowly to homonyms (Elston-Güttler & Friederici, 2005). This slower response to homonyms was attributed to

separate semantic representations. Thus, it was unexpected that homonym-pairs should be judged as similar in Degani et al.'s research.

Degani et al. contend that that their rating task tapped into later stages of processing in contrast to a lexical decision task. They argue that the inhibition of unrelated concepts, which would occur at an earlier stage of processing, had dropped by the later stage. They go on to claim that both unrelated concepts remain activated, and that "this coactivation leads to an association between the two meanings and/or lexical representations," (Degani et al., 2011, p. 24).

Amongst bilinguals it appears that if two words expressing different concepts are linked by a word form in either the speaker's L1 or L2, then the speaker is encouraged to find semantic similarity between the two words. As this finding involved speakers whose mental lexicons were mature across their L1 and L2, it is a question whether the result applies to L2 learners. It may be that L2 learners would be encouraged to look for similarity if an L2 word form referred to two separate concepts. However, since English monolinguals did not recognise the similarity, it seems unlikely that the similarity the bilinguals saw would be available for learners.

The motivation for reviewing the literature on translation ambiguity was to look at the evidence from psycholinguistic research on the influence of the L1 on L2 learning. The studies which investigated L1 ambiguity (Degani & Tokowicz, 2010a; Elston-Güttler & Williams, 2008; Jiang, 2002; Jiang, 2004) provided evidence for continued L1 lemma influence, even among advanced L2 learners. Elston-Güttler & Williams and Jiang both employed online tasks in their experiments. This indicates that they were investigating the representation of L2 words which the subject had already learned. The methodology employed in the L1 ambiguity studies perhaps limits their relevance to L2 learners. The subjects in these studies made quick semantic decisions which probably tapped into implicit cognitive processes relating to how information is stored (See Sandra and Rice's discussion of structure vs process on page 34.)

The more relevant direction of translation ambiguity is when the L2 word form is ambiguous, as this ambiguity includes L2 polysemous words. The tasks employed by these studies allowed the subjects more time to make a decision and thus likely tapped into how the senses were processed offline. Gathercole and Moawad employed a picture categorisation task which employs conscious decision making but is likely limited to concepts which can be represented visually. Paribakht conducted a think aloud strategy which can be used for more abstract concepts. However, this task did not investigate ambiguous words directly, so it remains a question whether this methodology can tap into the processing of L2 polysemous words. Perhaps the most promising method was the rating task employed by Degani, Prior and Tokowicz. It is likely that learners of different L2 proficiencies could conduct a rating task. The task could be applicable to the study of learning L2 polysemous words, because the learners' similarity ratings could be compared with a measurement of their knowledge of L2 polysemous words. This method as well as others are considered in Chapter 4.

The psycholinguistic research into L2 translation ambiguity has proven quite thin on the ground. Paribakht (2005) gave evidence for the difficulty L2 learners have in making a correct meaning inference for words that are not lexicalised in their L1. However, this study doesn't include difficulty involved with L2 ambiguity. Of those that do, there isn't clear data on how L2 learners deal with L2 ambiguity. The results from Gathercole and Moawad (2010) were inconclusive about the difficulty of L2 ambiguity for early and late bilinguals, likely because too few items were tested. The results of Degani et al. (2011) were less applicable to learners as the subjects in their study were mature bilinguals rather than L2 learners.

While there has been little research in psycholinguistics on how L2 learners process polysemous words, there has been some research on L2 knowledge and learning of polysemous words. This research is the most relevant to how L2 learners learn the different polysemous senses and it will be treated with greater detail than the research discussed up to now.

2.5 L2 knowledge of polysemous words

In the psycholinguistic studies reviewed above, the focus was on very specific problems of how language is processed and represented. It is difficult to obtain from this research a more descriptive picture of the individual learner and how his or her language knowledge develops over the course of extended study. For this reason, it is useful to turn to longitudinal studies which can present evidence of language development with students studying in language courses. This evidence can be analysed for the factors affecting the learning of L2 polysemous words.

Three studies investigated the knowledge and factors affecting L2 learning of polysemous senses. Schmitt (1998) and Crossley et. al. (2010) conducted longitudinal studies to investigate how time spent in general language learning is a factor in learning more polysemous senses. Kellerman (1978, 1986) took a different approach to the development of polysemous word knowledge in an L2. He investigated the factors affecting transfer of an L1 polysemous sense to an equivalent word in the L2. Together a review of this research provides insight into the factors affecting the development of L2 polysemous word knowledge.

2.5.1 Longitudinal studies on the development of L2 polysemous word knowledge

Schmitt (1998) and Crossley et. al. (2010) conducted longitudinal studies where the development of polysemous knowledge was measured over the course of many months. In both studies the learners demonstrated knowledge of more polysemous senses as time went on.

Schmitt (1998) tracked the acquisition of 11 words by three learners through a longitudinal study over a year. Schmitt's purpose for conducting this study was to focus on the acquisition stages of particular words. By focussing on individual learners over a year, Schmitt was able to describe degrees of acquisition and loss in the learners' lexicon, with the limitation that only a few words could be investigated. Schmitt took a dimension approach to the development of word knowledge (Nation, 2001; Richards, 1976). In this approach knowledge is understood by different components which do not necessarily develop in unison. I will restrict my discussion

to the component Schmitt refers to as 'knowledge of different meanings', because it has the most relevance to the question of learning polysemous senses.

Schmitt's three participants were post-graduate students with different first languages, studying at a British university in non-language departments. The target words were taken from the University Word List (UWL) and the 4000-5000 frequency range of the Brown Corpus[8]. All the words were polysemous with three or more senses. After piloting the list of words on other international students the following words were chosen: two unknown words, *brood* and *spur*; four well known words, *abandon, dedicate, illuminate* and *suspend;* and five in between, *circulate, convert, launch, plot* and *trace.*

Schmitt identified the most common meanings of the polysemous target words by consulting dictionaries, dominant meanings from pilot studies, and corpus data. The number of meanings per target word ranged from four to eight and the median number was six. To assess the participants on their knowledge of the different meanings, Schmitt first asked them in an interview what meanings they knew; he then used prompt words to elicit other meanings. Schmitt considered the unprompted responses a demonstration of productive knowledge and the prompted responses a demonstration of receptive knowledge. The participants were scored 2 points for each unprompted meaning, 1 point for each prompted meaning and no points for unknown meanings. One-on-one interviews were conducted three times at ½ year intervals (T1, T2, T3). After T2 and T3 the meanings were explained to the participants.

Schmitt compared the results of the L2 learners to native speakers whose vocabulary knowledge had been assessed by the same procedure. He saw an increase in the

---

8. The Brown Corpus (Francis & Kucera, 1979) is comprised of 1,000,000 words taken from 500 samples of English texts (fiction, non-fiction and journalism). The frequency measurement associated with these words was used in the development of the University Word List (Xue & Nation, 1984) and in the revision of the General Service List (Bauer & Nation, 1993). These list were widely used in L2 research at the time of Schmitt's writing.

number of senses the participants knew. Initially, the three participants knew the following proportions of the total number of senses: .33, .35 and .25. By the end of the study their scores had increased to the following: .56, .36 and .44. As we can see, this number did not grow as far as complete knowledge of all the senses. The words that Schmitt (1998) investigated were low frequency words with few senses. This was appropriate to the high proficiency of his participants; however, learners with a lower proficiency would be learning polysemous words of a much higher frequency. It is a question whether the results from Schmitt's study can apply to lower proficiency learners.

Crossley, Salsbury and McNamara (2010) conducted a study to investigate L2 learners' development of polysemous word use in English. Crossley et. al considered a polysemous word to be one with multiple meanings that are connected to a core meaning. The researchers viewed polysemy through a lexical network model and found polysemy interesting as an index to the learner's developing conceptual knowledge of the L2 lexicon.

Crossley et al. attempted to find out if learners increased their production of frequent polysemous words as their proficiency in English developed overall. At the outset, the researchers used two databases, WordNet and CELEX, to establish two measures for identifying polysemous words and their senses. WordNet is a database that associates word forms with concepts and it was used to assign polysemy values. CELEX was used to assign frequency values to the corpus of 99 transcripts collected through interviews with the participants.

Crossley et. al. found there was an initial increase in the number of senses produced over the first couple of months, and then there was a plateau in the number of senses produced. The WordNet polysemy values produced significant difference, $p < .001$, for the 2nd, 4th, 16th, 32nd, 50th,and 52nd weeks of learning. A pairwise comparison showed the values increased significantly between the 2nd and 16th weeks. However, there was no significant correlation between time spent learning and WordNet polysemy values ($r = .09$, $p > 0.05$).

These results show that initially the learners' use of polysemous words increased with the time spent learning English, but after the first four months there was a levelling out of polysemous word use. This result was similar to Schmitt's finding, who also found that his learners increased in their knowledge of polysemous senses but that they were unlikely to demonstrate complete knowledge of all the senses.

2.5.2 Longitudinal studies: Implications of the evidence

In these longitudinal studies, both Schmitt and Crossley et al. used interviews to gather evidence of their participants' developing language knowledge; however, their interview methods are markedly different, which resulted in qualitatively different evidence of L2 polysemous word knowledge.

In his interviews, Schmitt asked specific questions to determine what his participants knew and did not know about specific words. He first asked the students what meanings they knew of the target word. He then attempted to elicit the remaining meanings by using keywords that were chosen for their strong semantic associations. The first step in the interview was taken to measure productive knowledge and the second step to measure receptive knowledge. I would argue that both steps in the interview are in fact measures of productive knowledge. The first step will most likely elicit the more dominant meanings and the second step meanings that are more subordinate. Receptive knowledge is usually associated with the skills of listening and reading (Nation, 2001), but in this example the knowledge was assessed through the productive skill of speaking. Knowledge associated with correct inference would not be elicited through Schmitt's receptive knowledge protocol. We can take the example of a learner who encounters a familiar word form in an unfamiliar context. Using her background knowledge of other meanings of this word, she may correctly infer this unfamiliar sense of the word. This ability to correctly infer the meaning of a word should be considered an aspect of receptive knowledge because the inference is based on knowledge of the word. However, the researcher would not be able to elicit this sense without presenting the learner with the context.

It is important to measure receptive knowledge of vocabulary, because as Schmitt himself points out (Schmitt, 1998, p. 287), there is a movement from receptive to productive knowledge, and measurement of the receptive will help better understand this continuum. A straightforward method of assessing receptive knowledge would be to use a semantic decision task as used in psycholinguistic studies. The task would require the participant to answer yes or no as to whether or not a word can be used in a given sentence context. Such a question would indicate whether the student would be able to construe novel uses of the word. Considerations of receptive knowledge of polysemous senses will be taken up more thoroughly in Chapter 4.

In contrast to Schmitt, Crossley et al. used a range of elicitation materials in their interview method to prompt naturally occurring discourse. They gave special attention to the places in the data where a learner used a polysemous sense for the first time. They used these occurrences as evidence of two developments of learning. First, the increase of new polysemous senses indicated expanding knowledge of polysemous senses over time, and second, they argued that the knowledge of the new senses developed from the knowledge of the previously known senses.

While the evidence does show that a greater variety of senses are produced over time, it doesn't provide strong evidence explaining why that happens. In addition to inter-sense similarity, other factors may influence language learning. The factors of frequency, cognitive salience, and L1 influence might play a role in what sense the learner acquires. This is to say that a sense may be learned earlier if it is more frequent in the language the learner hears, if it is more important for carrying or interpreting the learner's message, or if the corresponding concept to the sense has an important parallel in the learner's L1. In itself, the evidence provided by interview discourse isn't able to confirm what role one factor of influence plays in relation to other factors.

The study of polysemy is challenging for both theoretical and methodological reasons; namely, it is difficult to clearly identify one distinct sense from another and

also to identify its occurrence in a large corpus. It would be a difficult for any single research construct to adequately deal with these challenges. It is useful to see how the two longitudinal studies by Schmitt and Crossley et. al. dealt with issues of distinguishing and identifying distinct polysemous senses. The following discussion presents a more detailed review of these studies.

2.5.3 Longitudinal studies: The factor of semantic similarity

One challenge for research with polysemous words is to identify distinct senses. This challenge was addressed earlier in reference to the lumper vs splitter debate over how similar or distinct individual senses should be from one another (cf. Section 2.3.3). In these two longitudinal studies, we can see how the theoretical debate between linguists also has methodological implications for empirical research.

Schmitt discussed this point as a limitation of the study, saying he, "sometimes found it difficult to determine the students' knowledge of the subtle differentiation between similar meaning senses without actually giving away those differences … Future studies of this type should probably use only clearly distinguishable meaning senses," (Schmitt, 1998, p. 309).

The similarity between senses was also an issue in Crossley et al.'s study. They used the computer based models of WordNet to identify senses, but it is not entirely clear whether these WordNet senses are identifying instances of polysemy or vagueness. As Crossley, Salsbury and McNamara point out, polysemy lies between a continuum of vagueness and homonymy, with vagueness expressing only a single core meaning and no related senses, and homonymy being distinct senses where one sense is inconsistent with the meaning of the other sense. I believe there is a question whether these results record modulations of vagueness. learning new senses or whether they record the construal of the same sense in slightly different contexts.

Crossley et al. provide the following examples as evidence of how a learner develops his knowledge of a new sense. In the first semester a student was using *think* to mean SUPPOSE: "I think he from Chicago." By the third semester the same student was using *think* to mean PONDER: "I'm think I'm going be Christian," (pp 598-599).

Senses like *think* [SUPPOSE] and *think* [PONDER] seem very fine-grained and it is not clear whether they are examples of polysemy or vagueness. According to Geeraerts (1993), vagueness refers to the distinction of meaning based on contextual specification, so that whether a *neighbour* is male or female is considered vague. In contrast, a lexical item is polysemous if its referent can be both true and false, so that you may be *covered* (by an insurance policy), but not *covered* (from the rain). To return to the example in the study, *think* can be considered an example of vagueness. It seems illogical to say that one can *think* (SUPPOSE) but not *think* (PONDER). Rather, it appears that *think (*SUPPOSE) and *think* (PONDER) are the same mental activity under different contexts. Thus it is better to consider *think* as an example of vagueness and not polysemy.

In both studies, polysemous senses were identified as a way of quantifying the knowledge a learner had of a particular word. This quantification is used to support the claim that L2 learners develop their knowledge of polysemous words as their length of time studying increases. However, Crossley et al. also argued that polysemous senses are radially organised from a core sense to peripheral senses. This organisation established a possible learning progression. They proposed that if a learner learned a core sense first, then the semantic similarity of the core sense with other senses would lead her to learn these other senses.

As evidence of this progression, Crossley et al. argued that their participants learned the core sense of the word earlier and the peripheral senses later. However, it is not clear that the evidence convincingly supports this claim. As evidence, Crossley et al. use the example of *think*. The claim is that *think* [SUPPOSE] is the core sense and that *think* [PONDER], among other senses, is peripheral. As discussed above, it is not clear how distinct the two senses are from one another; however, it is also not clear why *think* [SUPPOSE] should be considered the core sense. Crossley et al. (2010) claim that the core sense is more concrete or more literal than other senses (p. 577), and that it is usually acquired first. This idea is more complicated than presented in the article. The idea of a core sense is primarily one that is used in cognitive linguistics and is derived from the prototype theory of semantics (Lakoff, 1987). In cognitive

linguistics, the core sense is the one which provides conceptual coherence to the category described by the polysemous senses (Geeraerts, 2007). The Oxford Dictionary of English defines the core as the most typical, literal and central use of the word (Stevenson, 2010). The fact that the core sense may be more concrete than other senses is a corollary of the core sense and not a defining aspect. Moreover, many words do not lend themselves to being categorised as concrete. It is difficult to know why *think* [SUPPOSE] should be considered more concrete than *think* [PONDER]. Finally, it is also difficult to decide which of these two senses is the most literal sense.

It is important to establish a framework that identifies three components of polysemous words: how is the core sense defined, what makes one sense distinct from another, and how is the semantic extension between senses established. Such a framework would support the learning progression claimed by Crossley et al. For example, one could propose that if a student understood the phrase "*circulate* the blood," she might be expected to understand, "*circulate* the air" but potentially less likely to understand "*circulate* at a party". The proposed framework would be able to establish first, that *circulate* [BLOOD] was the core sense; second, that *circulate* [AIR] was a distinct sense; and finally, that *circulate* [BLOOD] was more similar to *circulate* [AIR] than *circulate* [PARTY].

2.5.4 Longitudinal studies: The factor of frequency

In the literature on polysemous senses, the term of 'frequency' refers to different constructs and applications. First of all, 'frequency' can refer to either the frequency of occurrence for the word form or it can refer to the intra-word frequency of the individual senses. For example, the word form *air* is among the 1000 most frequent words in English according to the General Service List (West, 1953)[9], but *air* in the

---

9. The General Service List is an old resource; however, at the time of this research (~2011-2012), it provided the best coverage of the most frequent words in fiction, non-fiction and academic texts (Coxhead, 2000). Since this research was completed a New General Service List has been completed which offers more up-to-date coverage for future research (Browne, Culligan, & Phillips, 2013).

sense of MUSIC would have very low intra-word frequency because this sense only occurs in very specific contexts.

In the research on L2 vocabulary acquisition, it is the frequency of the word form which has received the greatest amount of attention. According to Nation, one of the main factors for learning vocabulary is repetition of occurrence (2001). For this reason, frequency of the word form is one of the primary considerations researchers take when constructing a research instrument for L2 learners. For example, Schmitt (1998) used word form frequency to select target words for his longitudinal study so that his participants would likely know some words well and others less well.

The frequency of the word form becomes important for another reason in L2 learning of polysemous words. According to Zipf (1945), there is a positive correlation between the frequency of a word form and the number of senses it expresses, so that typically the most frequent words in the language are also the most polysemous. It is the frequency of polysemous words which serves as one of the strongest warrants for further research into how L2 learners develop their knowledge of L2 polysemous words. The more frequent a word form is in a language, the more important it is that a learner knows this word for both comprehension and production. However, if the word form has a number of different meanings, then we need to ask whether the learner must learn each individual meaning separately or whether knowledge of some meanings can lead her to correctly infer unfamiliar meanings.

The correlation between polysemy and frequency was taken into consideration by Crossley et al. They used frequency of word form as one of their measures for how the learners' use of L2 polysemous words grows over time. They found that CELEX word frequency values produced a significant difference at the 4th and 16th weeks, $p < .001$. These were the same weeks which saw a significant increase in the use of polysemous senses according to the WordNet values. In this case there was a significant correlation between time spent learning and frequency values $r = .37$, $p < .001$. There was also a significant positive correlation between the frequency and

polysemy values $r = .59$, $p < .001$. These findings of Crossley et al. support the notion that as the L2 learners learn more polysemous senses, they also tend to use the word form more frequently.

From the perspective of how L2 polysemous senses are learned, intra-word frequency is probably more important for research than frequency of word form. In general, frequency of occurrence is one of the most common predictors of learning, so it is reasonable to expect that if a learner encounters the specific sense of a word frequently enough, she will likely learn that sense. It follows that if another factor were proposed to be involved in learning L2 polysemous senses, the researcher would first need to demonstrate that intra-word frequency could not account for the evidence. Only then could he or she make a credible claim for the influence of the other factor. For example, it would be difficult to claim that a learner learned an L2 polysemous sense through semantic similarity without evidence that the sense was infrequent in the language.

Unfortunately, to my knowledge, there isn't a large scale corpus in the public domain with data on intra-word frequency. Both Schmitt and Crossley et al. were researching productive use of polysemous words, and they were able to identify when their learners first produced an L2 polysemous sense. Schmitt went further and compared what senses his learners knew to a list of other available senses. Many of these other senses would only occur in quite restricted contexts: *'abandon* to despair', 'a *brood* of ducks', '*circulate* at a party', or 'a *conversion* in football'. Data on the intra-word frequency of these senses would be an indicator of how likely or unlikely a learner would encounter them.

I have proposed here that in order to investigate how L2 learners acquire new polysemous senses the factors of intra-word frequency and semantic similarity must be taken into consideration. Kellerman's (1986) research is of interest for this reason. In a study of L2 learners, Kellerman attempted to construct a model of language transfer from the learners' L1 to their L2 based on these two factors frequency and semantic similarity.

2.5.5 Kellerman's model of L1 transfer for polysemous senses

Kellerman investigated the likelihood that an L2 learner would transfer an L1 polysemous sense to her L2. For example, in English we can use *head* in the sense of 'head of a table'. Kellerman investigated the likelihood that English speaker would transfer this sense of *head* to an equivalent L2 word form. He proposed a model of semantic transfer based on the frequency of the polysemous sense and its similarity to a proposed core sense. Specifically, he asked his participants to provide subjective decisions about the frequency, prototypicality and translatability of L1 senses of polysemous words.

Kellerman attempted to operationalise his investigation on transferability to the exclusion of other performance variables, presumably factors such as age, context, or familiarity. He stated that, "it is the claim of this paper that transferability can indeed be established entirely on the basis of the learner's knowledge of his native language, and that the establishment of these probabilities will have validity for any given L2," (Kellerman, 1986, p. 37).

In an earlier study, Kellerman found that ratings for transferability and prototypicality correlated, $p < 0.05$ (Kellerman, 1978). In that case prototypicality was measured by a card sorting task to establish the semantic distance to a previously identified core sense. However, in that study Kellerman was unable to determine what factors contributed to prototypicality. Subjective measures of concreteness did not correlate with the transferability judgements, and subjective measures of similarity to the prototypical sense only correlated to transferability at Spearman's *rho* $= 0.41$, $p = 0.05$, a weaker correlation than that between transferability and prototypicality.

His hypothesis in the later study (Kellerman, 1986) was that the transferability of a sense is a function of its similarity to a core sense and its subjective frequency, and relative to the transferability of other senses. To investigate the combined role of similarity and frequency, Kellerman devised what he termed the 'eye' experiment. The polysemous Dutch word *oog* ("eye") expresses many concrete extensions and

Kellerman reports that judgements of subjective frequency are easier to make with concrete nouns. He used six senses of *oog,* identified by the following English phrases: 'the human eye', 'the eye of a potato', 'an electronic eye', 'the eyes of a peacock's tail', 'the eye of a needle', and 'the spots on dice'. He felt that none of his subjects would have been explicitly taught the extended senses in English, meaning that intuitions should be at work.

The subjects were 35 Dutch first-year students of English at university. The six senses of *oog* were presented in defining phrases. The phrases were presented in pairs so that each sense was paired along side each of the other five senses in forced-choice preference tests. The subjects performed three tests. First, a translation test asked which sense is more likely to be rendered by *eye* in English. Second, a similarity test asked which sense is more similar to the *human eye* sense. The *human eye* sense was excluded in this test. Third, a subjective frequency test asked which sense is more 'frequent' in everyday language.

When results of the tests for similarity, frequency and transferability were compared separately to each other using a Chi-square analysis, there was a significant difference between each of the three tests. This indicated that neither subjective frequency nor perceived similarity predicted transferability when considered separately. Next, Kellerman compared expected transferability to observed transferability. The expected transferability scores were based on the interaction of similarity and frequency results and calculated using Luce's Choice Theory. The observed transferability scores were the results of the translation test. When the results of the observed transferability scores were compared to the expected transferability scores (similarity x frequency), the analysis found that the expected transferability scores did not differ significantly from the observed transferability scores ($p > .05$). This was taken to indicate that in this experiment transferability could be described as an interaction of subjective frequency and perceived similarity to the prototypical sense. However, chance could have played a role, as a subsequent experiment did not replicate these results.

After reporting this analysis, Kellerman then presented his replication of the "eye" experiment with English speaking learners of French and German and substituting the Dutch *oog* phrases for English phrases using eight senses of *head*: 'a head of steam', 'the head on a boil', 'the head of a poppy', 'the head on a glass of beer', 'the head of a nail', 'the head of a table', 'the head of a golf club', and 'the head of a piece of paper'. The subjects were 89 English secondary school students (aged 16-17) learning French and German. Again, expected scores were calculated based on the interaction of the frequency and similarity results. These expected scores were again compared using a Chi-square analysis to the observed scores established by the results of the translation test. In contrast to the *eye* experiment, a comparison between the expected and observed transferability scores was found to be significantly different in the *head* experiment. This indicates that in this experiment transferability cannot be explained by subjective frequency and perceived similarity to the prototypical sense.

Kellerman noted that even though the results of "eye" experiment were not replicated in the "head" experiment, there was a strong correlation between observed and theoretical scores. Kellerman maintained that it would be premature to dismiss the model since the correlation analysis confirmed the hypothesis in contrast to the Chi-square analysis.

2.5.6 A Critique of Kellerman's findings for ecological validity

Kellerman's research was focussed solely on semantic transfer of polysemous senses from the learner's L1 to her L2. His approach is appealing because of its narrow focus on a single psycholinguistic aspect of language acquisition. However, there are a number of reasons why the ecological validity of the research can be called into question. First, in a normal L2 learning process, the learner's receptive understanding of the L2 plays a large role in her productive knowledge the language. Furthermore, Kellerman's model of transfer is based on an idealised core sense and the core sense may be more irregular in practice. Furthermore, one should

also take into consideration how the similarity between the L1 and the L2 may affect transfer and how the learner's L2 proficiency may also be an influence.

The question Kellerman investigated was whether an L2 learner would transfer a polysemous sense from her L1 if she had no knowledge of its use in her L2. The normal assumption is that receptive knowledge precedes productive knowledge and so it is more likely that a learner will learn a polysemous sense receptively first and then learn to use it productively. Kellerman's research describes the situation where the learner's L1 influences her to produce a sense of the polysemous L2 word without L2 evidence of that sense. While L2 learners are likely to produce some senses without evidence from the L2, other senses may need a lot of evidence. It might be more reasonable to reframe the question to ask what makes a polysemous sense easy or difficult to learn. In this way equivalency of the polysemous sense in the L1 would be one influencing factor among others.

The most straightforward way to measure how much evidence an L2 learner receives is through the frequency of linguistic occurrence. However, instead of using a measurement of linguistic frequency, Kellerman collected data on subjective frequency. This measurement had it limitations, as Kellerman addressed in his discussion. He noted that subjective frequency could refer to the experience of linguistic occurrence (the number of times *eyes* of a 'potato' occurs in a corpus), the experience of the objects of the sense (not all potatoes have *eyes*), or the experience of the objects associated with the sense (potatoes are more common than peacocks). In this way it is ambiguous what exactly is measured by subjective frequency. Intra-word linguistic frequency, as measured by a corpus, would be the most reliable factor in learning polysemous senses in the L2. If a sense appears frequently in a corpus, it is more likely to appear in the learner's experience of the language.

The construct of the core sense was also problematic. In Kellerman's hypothesis, L1 transfer is a combination of subjective frequency and semantic similarity to a core sense. While there are difficulties in measuring subjective frequency, the influence of core sense, while commonly identified, runs the risk of being overly simplified. The

similarity to the core sense has been proposed elsewhere as a factor in learning L2 polysemous senses (Crossley et al., 2010; Csábi, 2004; Verhallen & Schoonen, 1993; Verspoor & Lowie, 2003). This factor presupposes that the core sense will be learned first and that the core sense is a clear semantic concept, neither of which can be taken for granted. Kellerman attempted to simplify the factor of similarity by choosing categories which centre on human body parts: the *eye* and the *head* in his two experiments. It is likely that the learners would know the sense of a HUMAN EYE before the sense of POTATO EYE. This is not necessarily the case with other polysemous words such as *branch* or *form.* For these words it is likely that senses of RETAIL SHOP or DOCUMENT could be learned before the core senses of TREE and SHAPE. Kellerman's choice of *eye* and *head* also means that the core senses are easily identified. In contrast, other research into semantic categories has focussed on words, such as *over* or *lie,* where the identity of the core sense is much more ambiguous (Brugman & Lakoff, 1988; Coleman & Kay, 1981).

Assuming the learner has a clear idea of the core sense, there is still a question of what establishes the similarity. In the case of *eye*, the similarity is usually based on the feature of SHAPE for most of the peripheral senses ('eye of a needle', 'the eye of a potato'). As Kellerman points out, only one sense, 'electronic eye', expressed FUNCTION as a second feature. In the *head* experiment, the features of similarity might be described by SHAPE and POSITION. In Kellerman's experiments, the factor of similarity appears fairly reliable because the core sense is unambiguous and there are only two features of similarity for each. This is not representative of polysemous words, however, where the core sense is often ambiguous and the features of similarity are more numerous. In such cases, there would likely be a great deal more variation between participants in their judgements of similarity.

Finally, Kellerman does not consider proficiency as a factor in L1 transfer of polysemous senses. His only comment on proficiency is that, "there does not appear to be any obvious effect either of nominal proficiency or of the particular pairing of target and source languages," (Kellerman, 1986, p. 43). This comment is curious since no independent measure of proficiency was taken and the possibility of an

effect of proficiency was not considered when comparing the differing results of the 'eye' experiment to the 'head' experiment. The learners are presumed to have some proficiency in the second language because the test asks if the sense of a word could be translated into a specific language (English, French or German). It is worth noting that the Dutch learners were probably more proficient in their L2 than the English students in theirs. The Dutch learners of English were at the university level while the English learners of French or German were in secondary school. Also, English is considered to be the de-facto second language in the Netherlands (Skutnabb-Kangas, 2000), which would encourage a higher level of proficiency, while German or French have very little currency in the UK and would lack the equivalent encouragement.

Given the expected differences in proficiency and the corresponding differences in the results of the two experiments, I believe that it is reasonable to say that proficiency could be more of a factor in the transfer of polysemous senses from the L1 to the L2 than Kellerman maintains. A learner with greater proficiency could be expected to know more polysemous senses than a learner with lower proficiency. By knowing more senses, the learner has a greater opportunity to recognise the relationship between those senses, and a greater likelihood of knowing when an L1 sense could be transferred and when it couldn't. Proficiency might also relate to more general confidence about the language because it means the learner can handle other things in the text, leaving more scope to work out the troublesome item. This is just conjecture, but I believe that it does warrant a consideration of proficiency as a factor in learning L2 polysemous senses.

In sum, Kellerman's research provides evidence that conceptual transfer from the L1 to the L2 can be described by the interaction of similarity to the core sense and the subjective frequency of the object in question. However, the factor of similarity to a core sense is often much more ambiguous than in the examples Kellerman used in his experiments and the factor of subjective frequency proved to be difficult to obtain valid measurement. Also, the importance of proficiency in the L2 was not considered for its effect on the transfer of conceptual knowledge. Finally, from the review of Kellerman's research, there appears to be a need to address how L2

learners develop their knowledge of polysemous senses with evidence from learners working in their second language. This will be a main aim of later experiments.

2.5.7 Discussion: Four factors influencing L2 learning of polysemous senses

Kellerman's study (1986) can be taken into consideration alongside the longitudinal studies of Schmitt (1998) and Crossley et al. (2010). In the review of these studies, four factors have been discussed as possible influences on L2 learning of polysemous senses: intra-word frequency of the sense, the semantic similarity of the sense to a core sense, the proficiency of the L2 learner, and the influence of the learner's L1 on the L2.

As discussed above, intra-word frequency would seem like a self-evident reason why a learner would learn a specific sense. However, since there isn't available data on the corpus frequency of senses, the researchers have had to turn to other methods. Both Schmitt and Crossley et al. consulted the frequency of the word form and then analysed knowledge of the senses through other qualitative means. In contrast, Kellerman attempted to elicit the learners' understanding of subjective frequency; however, this measure proved to be unreliable. There appears to be a need for some measure of intra-word frequency, at even some approximate level.

Semantic similarity to a proposed core sense was investigated as a factor by both Crossley et al. and Kellerman. However, there is a concern over how the core sense should be identified. Crossley et. al. identified the core sense as the most literal or concrete sense, yet it wasn't apparent whether this description could clearly identify the core sense of many polysemous words. Kellerman restricted his research to words with very clear, concrete examples of the core sense. It is important to establish a principled method of identifying the core sense for both concrete words and more abstract ones, because a reliable identification of the core sense is a necessary prerequisite to any investigation of whether semantic relatedness facilitates learning.

The factor of L2 proficiency was different in each of the three studies reviewed here. Schmitt's participants were advanced learners, while those in Crossley et al.'s study

were beginners. One should be cautious when extending the results of one group to the other. However, it is reasonable to expect that the ability to make inference or the confidence in transferring a concept from one's L1 would change depending the learners' L2 proficiency. Kellerman attempted to model L1 transfer outside of the influence of factors such as L2 proficiency, but such a position might prove to be untenable.

Finally, L1 influence was not addressed in the research of Schmitt and Crossley et al., whose learners came from a variety of L1 backgrounds. In contrast, Kellerman provided evidence that L2 learners are more likely to transfer prototypical senses from their L1 to their L2. However, other factors may also affect L1 transfer: Are learners more likely to transfer senses if they feel their L1 and L2 are similar, and as the learners' L2 proficiency increases, will they become more confident about transferring from their L1?

I have proposed that these four factors are possible influences on L2 learning of polysemous senses: intra-word frequency of meaning, semantic similarity to a core sense, the proficiency of learner in her L2, and the influence of the L1 on the L2. One way to investigate the validity of these factors is to see if they are accounted for in studies which attempted to teach L2 learners unfamiliar senses of polysemous words. Several relevant teaching studies are reviewed in the next section .

2.6 L2 teaching of polysemous words

The different senses of polysemy have been described as forming a radial category of meaning (Lakoff, 1987) whereby certain semantic similarities between senses connect the wide variety of meaning into a unified whole. Drawing upon the theory of prototype effects in semantic categories (Rosch, 1978), researchers have also proposed that a certain sense in a polysemous word category will express common semantic features that make it the central or core sense (Caramazza & Grober, 1976; Geeraerts, 2007; Nunberg, 1979). Kellerman (1978, 1986) drew upon this theory to investigate learners' willingness to transfer L1 senses based on factors of similarity to a core sense and subjective frequency. Also, the theory of a radial category informed

Crossley et al. (2010) in their investigation of how learners develop their knowledge of polysemous senses.

Several researchers have proposed that cognitive linguistics has the potential to inform L2 teaching methods for aspects of language such as metaphor, polysemy, phrasal verbs and idioms (Boers, 2000b; Boers, 2013; Boers & Demecheleer, 1998; Boers & Lindstromberg, 2007; Boers & Lindstromberg, 2008; Lindstromberg, 1996; MacLennan, 1994). The argument is that cognitive linguistics can provide insight into the semantic similarities that underlie aspects of language like polysemy; teachers and researchers can use these insights to develop learning strategies for language students. Several authors have developed methods based on cognitive linguistics to teach a number of different L2 language aspects: metaphor (Beréndi, Csábi, & Kövecses, 2008; Boers, 2000b; Boers, 2000a; Gao & Meng, 2010; MacArthur & Littlemore, 2008), idioms (Boers, Demecheleer, & Eyckmans, 2004; Kövecses & Szabó, 1996), and phrasal verbs (Condon, 2008; Tyler, Mueller, & Ho, 2010). A few studies (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003) have also investigated the effectiveness of teaching L2 polysemous senses using a method informed by cognitive linguistics. A review of these studies will show how well L2 learners develop their knowledge of polysemous words under quasi-experimental conditions.

2.6.1 Verspoor and Lowie's use of the core sense for inferencing unfamiliar senses

Verspoor and Lowie (2003) investigated how learners develop their understanding of polysemous meanings. The guessing method for learning new vocabulary had been widely promoted and yet there was still little known about the actual processes which learners undertake when they guess at new meanings. Verspoor and Lowie proposed that one way to achieve this goal would be by developing a learning strategy whereby a learner would use the core sense of a polysemous word in order to guess the meaning of extended senses.

Verspoor and Lowie investigated the core sense inferencing strategy with Dutch learners of English. They selected 18 polysemous target words which were identified

as unknown in a pilot study with different learners of the same L2 English proficiency. These 18 target words had at least three distinct meanings, organised radially from the core sense (S1). Thus the second meaning (S2) is semantically derived from the first meaning (the core meaning) and the third meaning (S3) is derived from the second. The following three items were used for the target word *boost*.

(S1) Boost me up this tree and I will get you an apple. (*boost = duwtje omhoog geven*)

(S2a) The landlord will boost the rent. (*boost = bevorderen*)

(S3) The tax cut can boost the economy. (*boost = verhogen*)

Crucially, each English target word is represented by at least three different word forms in Dutch, the participants' L1. Representative sentences for each target word were selected from the New York Times. A second sentence was selected for the intermediary sense (S2b) for testing purposes:

(S2b) The victory boosted Enqvist's chances of a gaining a spot in the Tennis Master's Cup. (*boost = bevorderen*)

The study was conducted with 78 students from three classes in two Dutch pre-university courses. All participants had three or more years of English study. Test 1 involved a guessing exercise while the Tests 2 and 3 measured short term and long term memory respectively.

In Test 1, the participants were presented with two English sentences for each target word. A Dutch translation was provided for the meaning of the target word in the first sentence but not in the second. The English target word was used in a different meaning in the second sentence and the participants were required to write the Dutch translation of this second meaning, using the meaning of the word in the first sentence as an aid in guessing. There were two different versions of Test 1, which were randomly distributed to the participants, so that half wrote one and half wrote the other. In both tests the second sentence used the target word in the intermediary sense (S2a). In one test, the first sentence used the target word in the core meaning

(S1). Thus the participants writing this test would use the core meaning of the word to guess the intermediary sense in the second sentence. In the other test, the first sentence used the target word in the third, extended meaning (S3). Thus these participants would use the extended meaning to guess the intermediary sense.

After writing the test, the participants were given a sheet with the correct translations for memorisation. After memorising, they were given an unannounced short term recall test (Test 2). In this test the 18 target words were presented to the participants in 18 sentences, using the second intermediary senses, but in different sentences from the ones used in the first test (S2b). The participants were again asked to provide Dutch translations of the target word in this meaning. A third, long-term, test was given to the participants unannounced 2 to 3 weeks later. This test (Test 3) was identical to the second test.

 The data were analysed by comparing the means of the group of participants who used the core sense as a clue to guessing to the group who used the extended sense as a clue. First, t-tests were conducted for each of the three tests. The core sense group performed significantly better than the extended sense group on Test 1 (core sense mean score = 10.7, $n = 18$; extended sense mean score = 7.8, $n = 18$; ; $p < .01$). There was no significant difference between the groups for Test 2 (core sense mean score = 17.3; extended sense mean score = 17.2; $p > .05$). This was to be expected, since the participants had reviewed the translations shortly before. The core sense group again performed significantly better on Test 3, the long-term retention test (core sense 13.0, $n = 18$; extended sense 10.7, $n = 18$; $p < 0.01$).

The results indicated that the L2 learners could use the meaning of a core sense to successfully infer the meaning of a related sense.

2.6.2 Verspoor and Lowie (2003): Concerns about balanced items

Verspoor and Lowie's investigation of polysemy potentially offers insight into the way learners develop their knowledge of vocabulary beyond a simple one-to-one translation from their L1. However, there is a question whether the items expressing the extended senses were more difficult than those expressing the core senses. The

76

target words were presented in the context of single isolated sentences. This was necessary because the different senses of the polysemous words are only construed from their context. The meaning of that context, however, needs to be transparent to the participant. If this is not the case, and the learner fails to correctly construe the sense of the target word, then we can't know why the learner failed. Was it due to the difficulty of the abstract/metaphorical sense of the word? Or was it because the learner didn't adequately establish the context and thus didn't know what the sentence was really about at all?

The context sentences can be difficult for a number of different reasons. The words themselves can be unknown to the subject, the syntactic structure can add difficulty, and the topic of the sentence can be something the learner is unfamiliar with. First, it is useful to look at a sentence the participants should have no difficulty with. The following sentence is taken from the extended (S3) sense for *boost*:

(14)   The tax cut can boost the economy.

The words *tax* and *economy* are fairly common words (although the compound noun *tax cut* might be more difficult) and the grammar is in simple present using a modal, which is acceptably familiar grammar.

The extended-sense sentence for the word *grapple* is far more difficult:

(15)   But those explanations have not made it any easier for them to *grapple* with the stark reality of losing four young men in such a terrible fire.

If the subjects don't know words like 'boost' or 'grapple' then can we expect them to know 'stark reality'? The sentence structure also adds to the difficulty of the sentence. The verb form is present perfect negative, and the complement to the main verb contains two embedded verb phrases. This structure is considerably more difficult than that in the 'boost the economy' example. Also, there are several anaphoric references to context outside of the sentence, marked by the words "but", "those explanations", "not easier", and "such". This creates a degree of ambiguity which might confuse the participant.

The topic of the sentence can also increase its difficulty. For example, one of the *bulge* sentences in the intermediate-sense is about basketball:

(16)   A breakway [sic] dunk by Raheed Wallace ended a 12-0 run by the Bullets that gave them their 5-point bulge.

This sentence uses a very specialised context and a participant who hasn't an interest in sports might find it difficult to establish its meaning.

Items such as these establish the extended-sense sentences as more difficult than the core-sense sentences. This is because the core-sense is usually more concrete and so their contexts will also be more concrete. While the core sense is not necessarily the most concrete sense, it often is because it refers to real world objects and actions. Unless consideration is taken to normalise the context of the different sentences, there will be a marked difference in difficulty between the core-sense and extended-sense sentences. For example, the core-sense and extended-sense sentences for 'gut' are as follows:

core-sense: "My mother hates gutting fish."

extended: "But while President Clinton fought against the gutting of environment laws, he offered little to no resistance on civil liberties."

There is an obvious inequality between the contexts of the two sentences. There is a question whether the learners were more successful at inferencing from the core sense because the core sense sentences provided a clearer, less complicated context than the extended sense sentences.

The English proficiency of the subjects may also have been a substantial factor affecting the results. The researchers must have felt that their students could cope with the language difficulties posed by the test items. This indicates that some of the students had quite a high level of English. However, their general level of English proficiency was not established. It would be good if their proficiency could be measured according to a standardised test, such as TOEFL or IELTS, or described with reference to a standard scale such as the Common European Framework of

Reference for Languages. Nevertheless, the students were probably of a very high level, some near fluent in English, and it brings into question whether these results are applicable to learners whose proficiency is high-intermediate and below.

2.6.3 Comparison between schema-based and translation-based teaching methods

While Verspoor and Lowie looked at the benefit L2 learners could gain from using the core sense for learning new senses, three studies (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007) designed methods of teaching L2 polysemous senses based on cognitive schemas that underlie the senses of polysemous words. In each of these three studies, the cognitive-schema method was compared to a translation-based method. The similarity between these studies is compelling because chance is less likely a factor if the results are consistently replicated.

In each of the three studies, L2 learners were taught the polysemous senses of two L2 target words: Khodadady and Khaghaninizhad selected the French words *arriver* and *sur* for their Iranian L1 students; Morimoto and Loewen, the English words *break* and *over* for their Japanese students; and Csábi, the English verbs *hold* and *keep* for her Hungarian students.

Morimoto & Loewen and Khodadady & Khaghaninizhad organised their students into three groups: one group received instruction through a schema-based method, the second through a translation-based method, and one group acted as a control and received no instruction. Csábi did not employ a control group, but did employ two proficiency levels, so that separate high and low proficiency groups received the schema-based instruction and translation-based instruction.

In all three studies, the cognitive-schema teaching method was developed using insights from the field of cognitive linguistics. Morimoto & Loewen and Khodadady & Khaghaninizhad employed image schemas in their experimental method. As discussed in Section 2.3.2 on page 22, image schemas are cognitive constructs informed by basic experience with the world. Morimoto and Loewen argued that an

image-schema based method would allow teachers to present polysemous words more systematically than by presenting each different sense individually.

Csábi employed the cognitive schemas explained by Talmy's (1988) force dynamics. This theory describes language according to an experiential basis in a similar way to the image schemas. The difference is that force dynamics are more suited to verb phrases, whereas image schemas are more suited to nouns and prepositions.

Unlike Morimoto & Loewen and Khodadady & Khaghaninizhad, Csábi expanded her schema method to also explain how phrasal verbs and idioms could be integrated into the polysemous category through conventional metaphors. For example, the metaphor OBSTRUCTION IS UP is used to explain the phrase *hold up*, as in "The whole thing was held up about half an hour"; and the metaphor KNOWING IS SEEING is used to explain the idiom "Keep something under your hat."

Khodadady and Khaghaninizhad's study was a replication of Morimoto and Loewen's, and the same method was employed across both studies. For both the schema-based and translation-based instruction, the L1 equivalent of the target word was presented in three example sentences and the instructor discussed the differences between the target word and the L1 equivalent with the students in their L1. Then the experimental treatments were explained to the students. For the schema-based instruction, the students read a summary sheet explaining the core meaning of the target word and its extensions. For the translation-based instruction, the students read a list of meanings of different senses of the target word, similar to what one would find in a dictionary. Finally, both groups translated English sentences that used the target word into their L1. The schema-based instruction group translated five sentences, while the translation-based instruction group translated ten sentences. The increased translation was to balance out the total time on task between groups.

The participants wrote two identical sets of post-tests: the first was two days after the treatment, and the second was 14 days after the treatment. One post-test measured receptive knowledge through an acceptability judgement test. The other

post-test measured productive knowledge. The students were prompted with a picture and noun phrase to write an English sentence using the target word.

In Csábi's study, the teaching procedure for the schema-based instruction groups was broken into three sections: one for senses, another for phrasal verbs and a third for idioms. In each section, the students were explained the motivations behind the meanings of both *hold* and *keep* in their L1. After each section's explanation they were instructed to memorise the meanings. They were then given a gap-fill completion task as a test.

In contrast, the translation-based instruction groups in Csábi's study were explained the L1 translation for each of the meanings for *hold* and *keep*. These were written on the board. As with the schema-based instruction groups, the translation-based instruction groups were instructed to memorise the meanings and then to complete the gap-fill completion task as a test. The presentation of the meanings was also broken into three sections for senses, phrasal verbs and idioms. The lower proficiency groups took the gap-fill completion task again as a post-test one day later, while the higher proficiency groups wrote the same post-test two days later.

While the three studies were comparable in their methods, there was less consistency between the studies in their results. The results for Morimoto and Loewen's study were mixed. In the translation-based instruction, students performed significantly better with *break* but not *over*. In the schema-based instruction, by contrast, the students performed significantly better with *over* but not *break*. Furthermore, neither treatment was significantly better than the control in the second receptive post-test.

In contrast, Khodadady and Khaghaninizhad found that the participants who had received schema-based instruction performed better for all three conditions: the two post tests (acceptability and production), the two time intervals (2 days and 14 days after intervention), and target words (*arriver* and *sur*). However, a comparison of the mean results for the schema-based instruction and the translation-based instruction were only significant in the first acceptability post-test. For the word *arriver,* the

mean score for the schema-based group was 10.71 (*SD* 1.14, *n* = 15) and 9.48 (*SD* 1.38, *n* = 15) for the translation-based group, *p* < .05. For the word *sur*, the mean score for the schema-based group was 12.38 (*SD* 2.18, *n* = 15) and 11.15 (*SD* 1.65, *n* = 15) for the translation-based instruction group, *p* < .05. The results for schema-based and translation-based groups were better on all tests than the control group, who had received no instruction and only wrote the tests.

Csábi scored her results in two ways, as keyword-type answers or entirely-correct answers. Keyword-type answers were marked as correct if *hold* and *keep* were chosen correctly; entirely-correct answers were marked correct if the student wrote the other words in the phrase or idiom correctly. In Table 2.4 below, the results for the keyword-type answers are presented as percentages by group for both the test and post test.

Table 2.4
Results of the test and post test by group (standard deviations were not provided) Csábi (2004), p 253.

	Low Proficiency			High proficiency		
	Experimental	Control		Experimental	Control	
Test	96.15%	88.10%	*	95.71%	88.00%	†
Post-test	86.17%	70.60%	*	87.70%	71.20%	*

* Significant at *p* < .01, † Significant at *p* < .02

Csábi found there were significant differences between treatments for the results of low proficiency learners and high proficiency learners for all tests. She also calculated the results for the entirely-correct answers but for some reason did not compare them for statistical significance. These results were comparable to the keyword-type results except for the idiom session when high proficiency learners were compared. In that case, the translation-based instruction group performed better on the second post-test than the schema-based instruction group. In all, the students performed better in the schema-based treatment - when the motivation behind meanings was explained to them - in three of the four tests.

When the studies are taken together, the results are mixed over the benefit of the schema-based instruction. While the schema-based instruction was better in each

study for some of the tests, the translation-based instruction also proved effective in some cases. It's worthwhile to consider whether the mixed results were due to the research design or due to the underlying rationale of the schema-based instruction.

2.6.4 Controlling for deeper language processing

Morimoto and Loewen (2007) gave four reasons in favour of the schema-based instruction method. First, the underlying common meaning could serve as a basis for comprehending novel usages and could prevent a tendency of the learner to over- and under-generalise the use of the word. Second, they argued that schemas would be an intuitive aid to memory, which was the same argument also given by Csábi (2004). Third, they argued that an image schema has the potential to overcome the limitations of an L1 translation equivalent of the L2 polysemous word. Often translation equivalents across languages do not align with all meanings and uses. Finally, schema-based instruction would allow for deep processing of words, which was seen as an essential component to vocabulary acquisition (Craik & Lockhart, 1972; Ellis, 1997).

Their last point about deeper language processing is potentially problematic. Both schema-based instruction and translation-based instruction have the potential to engage deep language processing. For this reason, an attempt must be made to balance the two instructional methods for the amount of cognitive effort that they demand of the students. However, there are reasons why the experimental design in the three schema-based experiments may be unbalanced in this regard.

Success in language learning has been attributed to activities which require greater cognitive effort on the part of the learner (Craik & Lockhart, 1972; Ellis, 1997). The schema-based groups in Csábi's study were required to make semantic links between words used in widely varying contexts. Not only this, but schematic drawings and interpretive drawings were used to explain the phrasal verbs and idioms. In contrast, the translation-based groups were presented with the translation equivalent of the different meanings, written on the board. In comparison to the schema-based condition, the translation-based instruction is a much more passive

mode of learning. The extra cognitive effort required of the schema-based instruction group could well be a factor in their stronger results on the keyword based tests. In order for the translation-based condition to be a true comparison to the schema-based, it would need to involve a training task of equivalent engagement.

In contrast, there is reason to believe that Morimoto and Loewen's research design benefited the students in the translation-based condition. For the schema-based condition the students translated five English sentences into Japanese using a summary sheet and the researcher explained the core meaning of the target word. However, for the translation-based condition, the students translated ten sentences with, "an inventory of meanings associated with the target word," (p. 354). This means that the translation-based instruction group translated twice as many sentences as the schema-based group. This was done to accommodate the longer time spent on the image-schema instruction. As a consequence, the schema-based instruction group spent more time passively listening to the teacher, while the translation-based group spent more time on the active learning of translation. While the overall time was kept equal, the translation-based instruction group had an advantage in time spent on deep processing. This extra time spent on active learning could have been a factor in the results, which saw, against expectation, the translation-method group performing about equally to the schema-method group.

If Morimoto and Loewen's design was unbalanced then we need to question why the students in Khodadady and Khaghaninizhad's study, which used the same research design, did not produce comparable results. Instead, the students who received the schema-based instruction treatment performed better than the translation-based instruction group. Other factors could have played a role, such as differences in L2 proficiency or differences involved in the first and second languages: Japanese and English compared to Persian and French. Be that as it may, there do appear to be concerns with the research design which make it difficult to interpret the results confidently.

Across the three studies, schema-based instruction was generally more successful than translation-based instruction. This success was qualified by Morimoto and Loewen's results, which were mixed. As a possible explanation, the authors question whether or not the learners would naturally use an image schema even when doing the translation task. They comment that the same image schema for *over* was also available in the Japanese equivalents (*-wo koete, -no ueni, -wo ootte)*. This would make the image-schema method redundant because the learners were already engaging the image schema through translation. They also comment that application of the image schema in metaphorical domains might have been too difficult for the learners to grasp, such as in "You cannot *break* the contract." This is problematic because according to their two explanations either the image schema instruction is not necessary or it is not effective enough.

In contrast, the experimental methods presented in Verspoor and Lowie (2003) and Csábi (2004) engaged the students with instruction that more directly tested the application of insights from cognitive linguistic theory. In Verspoor and Lowie the students who made an inference based on a core sense translation were compared to the students who made an inference based on an extended sense translation. As predicted by cognitive linguistics theory, the core sense method was more successful. In Csábi's study (2004), the students had to decide whether to use *keep* or *hold* in the gap fill exercise. There is a great deal of similarity between these two words. This similarity would have engaged the students to make more subtle semantic distinctions that would require more close consideration of the cognitive linguistics based instruction.

2.6.5 Discussion: General concerns in the experimental methods

Three common concerns arose in the reviews of studies on teaching L2 polysemous senses: the question of a balanced design, the differences between the purposes and results of the studies, and finally the makeup of the participants. The two design issues discussed in the reviews were the balance of the item types and the balance of the learning tasks. In Verspoor and Lowie (2003), the participants were prompted with either a core sense item or an extended sense item. The core sense items were

seen to be more simple to understand than the extended items because they used more frequent and common words and less complicated syntax. Similarly, in the research of Csábi (Csábi, 2004), Morimoto and Loewen (2007) and Khodadady and Khaghaninizhad (2012) there was an imbalance between the schema-based instruction and the translation-based instruction. In Csábi's design, I argued that the cognitive-linguistic based treatment offered the participants greater depth of processing than the translation treatment. In contrast, in Morimoto and Loewen's study, as well as in Khodadady and Khaghaninizhad's replication, the participants spent more time-on-task in the translation treatment than in the cognitive-linguistic based treatment.

There were also differences in both the purposes and results of the studies. Verspoor and Lowie designed their method to develop an inference strategy, so that when the learners encountered a novel polysemous sense, they would be able to infer what its meaning was based on their knowledge of the core sense. By contrast, in the other three studies the participants were required to memorise a set list of polysemous senses. The method employed by Verspoor and Lowie has the obvious benefit of being applicable to a wider range of language and it also has the potential to be used by the learners outside of a predefined lesson in the context of unstructured language learning.

The methods employed in these studies are experimental and would benefit from replication. The only replicated method was the image-schema treatment employed first by Morimoto and Loewen and then by Khodadady and Khaghaninizhad. The results of these studies were not comparable overall. While Khodadady and Khaghaninizhad found a benefit for the cognitive linguistic treatment for both prepositions and verbs across the post tests, Morimoto and Loewen only found a benefit for the preposition in the first post test and no benefit for the treatment when the verb was taught.

Finally, these studies used participants from different L1 backgrounds and probably of different proficiencies. Morimoto and Loewen proposed that the learners' L1 may

have affected results. The researchers argued that if there was strong translation equivalency between the languages, then the image-schema treatment might not be necessary; however, even if there wasn't that strong translation equivalency, then the metaphoric concepts expressed by the L2 polysemous senses might be too difficult for the learners to grasp.

The second factor affecting the participants was their proficiency. I discussed above about the difficulty of many of the items in Verspoor and Lowie's instrument. The fact that many of their participants were able to complete the study draws into question whether their proficiency was so high that the results might not be applicable to learners who have a less advanced proficiency.

In sum, these four studies found some evidence of benefit from learning L2 polysemous senses through teaching methods informed by cognitive linguistics. However, this benefit is mitigated by questions of instrument design, replicable methods and the role of the learners' proficiency and L1. Factors such these should be further investigated for their effect in L2 learning of polysemous senses.

2.7 Conclusion

As discussed earlier, there is a correlation between the frequency that a word form occurs in the language and the number of senses it has (Zipf, 1945). While it is generally accepted that it is important for L2 learners to know the most frequent words in the language (Nation, 2001), there is a question about what knowing a word entails. For polysemous words, a major challenge is whether knowing only one sense constitutes word knowledge considering that the word is used in many other ways.

Several studies reviewed here have shown that L2 knowledge of polysemous senses develops over time (Crossley et al., 2010; Schmitt, 1998); however, there is less evidence whether the learners learn these senses separately from each other, or whether they are able to infer the meaning of an unfamiliar sense based on their existing knowledge of other senses. Some evidence indicates that this might be the case. Kellerman (1986) proposed that similarity to a core sense was a key factor in L1

transfer of polysemous senses; however, his research design did not involve evidence from learners using their L2. Crossley et al. (2010) relied on computer-based delineations to identify polysemous senses. However, these senses did not align well with natural expectations of what a sense is. It remains inconclusive whether L2 learners develop their knowledge of polysemous words based on semantic extension or whether this process is mitigated by other factors, namely L1 influence or intra-word frequency.

There is some evidence from the research on teaching L2 polysemous words that students benefit from inferring the meaning of unfamiliar senses based on some existing knowledge of the word's semantic content. This has been analysed based on the extension of the core sense or an underlying cognitive schema. However, in most of these studies, there was an imbalance between the schema-based instruction and the translation-based instruction (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007).

Verspoor and Lowie's method (2003) was more directly designed around the factor of semantic similarity. In both the core sense treatment and the extended sense treatment, the learner was presented with a translated polysemous sense to use as inference prompt. Of all the studies reviewed in this chapter, Verspoor and Lowie's holds the most promise for fully engaging with the learning potential of the semantic similarity between senses. However, since there were reservations with the participants' proficiency level and with the item design of this study, the results cannot be attributed solely to the experimental design. In the next chapter, Verspoor and Lowie's study will be replicated with participants from a different L1 (Arabic) and whose English proficiency is likely different from those in Verspoor and Lowie's original study.

Based on the literature reviewed here, the following research questions have emerged which will be pursued in the thesis though a series of experiments and subsequent analysis of the results:

- *To what extent is semantic similarity between polysemous senses a factor in how learners develop their knowledge of L2 polysemous words?*

- *What is the effect of the following factors on L2 knowledge of polysemous words: the intra-word frequency of different senses, the language learner's L2 vocabulary size, and the influence of the learner's L1?*

The research questions are based upon the psycholinguistic and meaning-last models of vocabulary learning as presented in Section 2.1.2 (page 15). Jiang's psycholinguistic model takes the position that increased processing of an unfamiliar sense can lead to storage of new semantic content. Processing is understood to include the act of construal referred to in the literature from cognitive linguistics. Knowledge of a polysemous sense is understood to include both the stored semantic content and the ability to process that stored information. Learning is considered to be the act of processing that leads to knowledge of the new sense.

The above research questions will be addressed using a new test, the Test of Polysemous Meanings (TPM). Chapter 4 presents how the test was designed to elicit judgements of semantic acceptability from L2 learners on the use of polysemous L2 words. The test items present the words in sentence length contexts. The test includes both acceptable items where the word is used acceptably and distractor items where the word is used unacceptably. In contrast to the longitudinal studies of learning presented earlier (see Section 2.5.2), the TPM is designed to be used in a cross-sectional approach, whereby acceptability judgements are elicited from a large sample of learners. A greater number of acceptable responses to the acceptable items is taken to indicate greater knowledge of those item types.

In order to investigate the role of semantic similarity addressed in the first research question, the test includes an item type referred to as logical distractors. These are items where the target word is used unacceptably but logically and meaningfully (see Section 4.2.4 on page 115 for more information). If learners judge these items as acceptable more often than illogical distractors (where the target word is used non-sensically), then the reason for the greater acceptability is taken to be the semantic

similarity that the polysemous words bear to the learners' existing knowledge of the target word.

After presenting how the TPM was designed in Chapter 4, the test was piloted with Arabic-speaking learners of English in Chapter 5. One of the main purposes of the pilot was to see if these learners of English judged the logical distractors as more acceptable than the illogical distractors. Based on the results of the pilot study, the test items were further normed for native speaker acceptability in Chapter 6. Then the test was given to a larger cohort of Arabic-speaking learners of English. The results of which are presented in Chapter 7..

The TPM was also designed to investigate the factor of intra-word frequency, which is addressed in the second research question. Intra-word frequency refers the frequency of occurrence of different senses of a polysemous word relative to other senses in a large corpus. For each target word on the TPM, there are three acceptable items representing a different level of intra-word frequency. The responses to these items can be compared between the three different levels to investigate intra-word frequency. These design considerations for intra-word frequency were also presented in Chapter 4 in addition to those for semantic similarity. The results of this analysis of intra-word frequency are also presented in Chapters 5 and 7.

The results first presented in Chapter 7 are further used to address the research questions in Chapters 8 through 10. The second research question addresses the effect of L2 vocabulary size. In order to investigate the role of this factor, a separate test of vocabulary size was given to the participants who took the TPM. In Chapter 8, the results of this test are used to re-analyze the results of the TPM for the effect of vocabulary size.

In Chapter 9, the effect of semantic similarity is returned to through a further a re-analysis of the TPM items for the semantic relationship of the polysemous senses to a core sense of the target words. While semantic similarity had been investigated

earlier through the use of the logical distractors, the objective in Chapter 9 was to see if semantic similarity was a factor in judgements for the acceptable items as well.

Finally, in Chapter 10, the effect of L1 influence is addressed through a translation of the polysemous senses used in the TPM into the learners' L1. The results of the TPM are reanalyzed based on this translation to see if senses which translate into the same equivalent word in the learners' L1 are judged more acceptable than senses which translate into different L1 words..

The development of the Test of Polysemous Meanings grew out of a replication of Verspoor and Lowie's (2003) study. This replication is presented in the following chapter. The replication aims to determine whether Arabic learners of English, at an intermediate proficiency, can guess the meaning of a polysemous sense more correctly when using a core sense as a cue rather than an extended sense.

Chapter 3: Replication study of Verspoor and Lowie

This chapter presents a partial replication of Verspoor and Lowie's *(2003)* study on L2 inferencing of unfamiliar polysemous senses. The replication differs from the original study in two important ways: the participants were L1 speakers of Arabic in contrast to the Dutch speakers of the original, and the participants were intermediate L2 learners of English and not upper intermediate. The instruments used in the replication had to be adapted for Arabic speakers. In most cases this simply involved changing the Dutch translations to Arabic translations, but in certain cases the new tests items had to be created. The replication test was given to 55 Arabic speaking learners of English, studying in their first year at university. The results of the study are presented and discussed.

3.1 Introduction

As noted in Chapter 2, research on the L2 learning of polysemous words can be organised under two broad thematic headings. The first theme was the knowledge learners have of L2 polysemous words (Crossley et al., 2010; Kellerman, 1978; Kellerman, 1986; Schmitt, 1998; Schmitt, 1999) and the second theme was teaching learners L2 polysemous senses (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). Of the studies in the teaching theme, Verspoor and Lowie directly investigated the effectiveness of learning a new sense based on similarity to a core sense in comparison to similarity to an extended sense. Verspoor and Lowie's study is important because similarity between senses is the feature which distinguishes polysemous words from homonyms or words where the meanings are inconsistent with each other. Verspoor and Lowie stated that beginners could benefit from learning a core sense first, and moreover,

> Intermediate learners might also benefit from a brief introduction into the way that the different senses of a polysemous word may be related to each other and to a core sense, so that they can discover meaningful links among the various senses. (p 570)

By these comments on beginner and intermediate learners, the researchers claimed that the results of their study were applicable to learners at these proficiency levels. The original participants were pre-university Dutch learners of English. I felt that there was a warrant to replicate the study with pre-university Arabic learners of English. Verspoor and Lowie's study provided robust results but they were restricted to Dutch-speaking English language learners. To show that their results are applicable to language learners in general it is important to investigate their claims with learners from different L1 backgrounds. In the replication, the Arabic-speaking learners of English were comparable to the original participants in all but their first language and geographic location. They were studying English for equivalent purposes and they were studying intensive English in a pre-university course at a university in Qatar. A replication of Verspoor and Lowie's results with these learners would support their claim that intermediate learners can benefit from a core-sense inferencing strategy.

Verspoor and Lowie found evidence that learners were more successful at guessing the meaning of an unknown polysemous sense if they had knowledge of the word's core sense rather than knowledge of an extended sense. The study was based around the claim from cognitive linguistics that polysemous words express a radial structure whereby a core sense is linked to extended senses through shared semantic similarities. The semantic distance is said to increase as the extended senses become less similar to the core sense. The idea of the radial structure was discussed in Section 2.3.3.

The researchers proposed that their learners benefited from recognising how senses were related to each other through the core sense; however, in my critique of this study (see Section 2.6.2), I questioned whether their participants were representative of typical pre-university L2 English learners. My critique was based upon an assessment of the item difficulty in the tests and not upon empirical evidence. Such evidence could be obtained from replicating the study with comparable participants. The participants in the replication were similar to those in Verspoor and Lowie's research. They were studying in an English medium university and taking a course

in English for academic purposes. Using these participants, the replication study attempted to confirm Verspoor and Lowie's finding that a core sense facilitates guessing an intermediary polysemous sense better than an extended sense. Two research questions are addressed in this chapter:

RQ1   *Do L2 learners guess the correct meaning of a polysemous sense more often when using a core sense as a cue rather than an extended sense?*

RQ2   *Are the results of post-secondary Arabic-learners of English comparable to post-secondary Dutch-learners of English also studying in an English-medium university program?*

3.2 The method of the original study

Verspoor and Lowie identified 18 English polysemous words which their participants were not expected to know. Each word expressed three different senses. Importantly, these three senses translated differently into Dutch. Native English speakers confirmed that the intermediary sense (S2) was more semantically similar to the core sense (S1) than the extended sense (S3).

There were three tests in their research design. In the first test, each of the 18 polysemous words was presented in two English sentences. In the first sentence the polysemous word would be glossed into Dutch, the participants' L1. This sentence would act as a cue for the meaning of the polysemous word in the second sentence. The task of the participants was to write a translation of the target word as it was used in the second sentence. The first sentence would either present a core sense of the target word as a cue or an extended sense of the target word as a cue. An example of an item with a core sense cue (S1) is as follows:

Core Sense Cue (S1)	<u>Boost</u> me up this tree and I will get you an apple.	boost is 'duwtje omhoog geven'
Target Item (S2)	The landlord will <u>boost</u> the rent.	boost is _____

This is in contrast to an item with an extended sense (S3) as a cue:

Extended Sense Cue (S3)	The tax cut can <u>boost</u> the economy.	boost is 'bevorderen'
Target Item (S2)	The landlord will <u>boost</u> the rent.	boost is _____

Half the participants took the test using the core senses as cues and the other half using the extended senses as cues. The participants were then given the same test with the answers completed. They were asked to memorise the translations for the intermediary senses (S2). Finally, a post test was given to them at two intervals to measure their retention of the target senses, one immediately after the memorisation task and one two or three weeks later. This test asked the participants to translate sentences containing the target words used in the intermediary sense. These sentences were different from the original test, but the sense remained the same. These post-test sentences were referred to as (S2b) and those on the first test as (S2a).

3.3 Modifications to the instrument

The original study contained Dutch glosses of the target words. These glosses had to be adapted for the Arabic L1 learners in the replication study. The adaptation had to meet two conditions. First, the Arabic translation had to be different for the three different senses (S1, S2 and S3) for each target word. Second, the Arabic translation had to be the same for the intermediate sense as it was used across two sentences (S2a and S2b).

A native Arabic speaker with an MA in translation translated the target words into Arabic as they were presented in the four context sentences. An example of these four sentences with their Dutch and Arabic translations is presented with the target word *rake* in Table 3.1.

Table 3.1
Arabic adaptation of the instrument from the Dutch original

Condition	Example sentence	Dutch equivalent	Arabic equivalent
S1 (core sense)	The gardener set fire to the piles of weeds he had raked up.	harken	يدمّ، يجمع أعشاب
S3 (extended sense)	We have been raking through all her papers.	zorgvuldig doorzoeken	يفتش/يبحث بدقة

S2a (intermediary sense) test 1	Relief foundations raked in $13 million last year.	inzamelen	يجمع (أموال طائلة) , يدر(أرباح طائلة)
S2b (intermediary sense) tests 2 and 3	But Federal prosecutors here say prison walls have not stopped Mr. Hoover from overseeing an illegal narcotics business that raked in $100 million a year.	inzamelen	يجمع (أموال طائلة) , يدر(أرباح طائلة)

The translator identified one word from Verspoor and Lowie's original list, *sprawl*, which translated into one single Arabic word for senses S2 and S3. This target word was excluded from the study because no substitute sentence could be found to distinguish between the two senses. For four other words - *boost*, *gut*, *skim*, and *smother* - one of the intermediary-sense sentences (S2a or S2b) translated into the same Arabic word as either the core sense (S1) or the extended sense (S3). New sentences were selected from the Brown Corpus or the British National Corpus, so that the target word in the intermediary conditions (S2a and S2b) translated into the same Arabic word. Table 3.2 presents the four original sentences from Verspoor and Lowie and their replacements:

Table 3.2
Alterations to the Original Study.

Item	Original Sentence	Replacement Sentence
1.2b	The victory boosted Enquist's chances of gaining a spot in the Tennis Master's cup.	Perhaps advertising might help them boost their sales.
4.2b	Seven days a week, from morning till evening the five hulking vessels of the Delaware Rivers and Bay Authority skim across the water, from Cape May to Cape Henlopen and back.	Moments later the [airplane] skimmed across the landing strip edging closer and closer to a touchdown, then in a streamer of dust it landed. ('airplane' replaces 'V1')
15.2b	Mr. Mobuto ran his country like a dictator, smothering political opposition, torturing rebels, and violating international treaties.	He's as anxious as you and I to smother things up.
11.2b.	A Shell station near Hamburg was gutted early today.	In 1838, a devastating fire gutted their small shop, and soon thereafter David Brown moved west to Illinois settling on a land grant in his declining years.

3.4 Replication Instruments

In total, 17 target words were investigated in the replication study. After the four modifications to Verspoor and Lowie's original items, the tests used in the replication study remained the same as in the original. Table 3.3 lists the different instruments used in the replication study:

Table 3.3
Instruments used in the replication of Verspoor and Lowie (2003)

Instrument	Purpose	Method	Appendix
Test 1 - S1	Core sense condition	Written by half the participants.	Appendix 1.1, p. 317
Test 1 - S3	Extended sense condition	Written by half the participants.	Appendix 1.2, p. 321
Worksheet - S1	Intermediate sense answers	Studied by participants in the core sense condition.	Appendix 1.3, p. 325
Worksheet - S3	Intermediate sense answers	Studied by participants in the extended sense condition.	Appendix 1.4, p. 328
Test 2	Short term post-test	Written by all participants on the same day.	Appendix 1.5, p. 331
Test 3	Long term post-test	Written by all participants 2-3 weeks later.	Appendix 1.5, p. 331

3.5 Participants

In the original study, Verspoor and Lowie selected participants enrolled in a pre-university course who had at least three years of English, but no quantitative measure of their proficiency was provided. In the replication the participants were first year university students with a score equal to or higher than 450 on the TOEFL or 5.5 on the IELTS. Fifty-five first-year university students wrote the tests; they were from five classes studying English for academic purposes.

3.6 Procedure

The tests were conducted in classes that lasted 90 minutes. Verspoor and Lowie did not detail how their participants were instructed in the test procedure. The procedure is not self-explanatory, so the following preparatory steps were made to familiarise the participants with the task. First, a short presentation was given to the participants in which the idea of polysemy was explained using examples not part of

the study. Then the participants were told that the test contained cue sentences with an Arabic gloss and test sentences with no Arabic gloss. They were directed to write an Arabic gloss for the target word in the test sentence. They were told that the sense of the target word was different from the glossed sense in the cue sentence. The explanation and directions for the task lasted 5-10 minutes. Within each class the participants were randomly assigned to either the S1 (core sense) presentation or the S3 (extended sense) presentation. The participants then wrote the first test (Test 1 - S1 or Test 1 - S3) which took approximately 15-20 minutes.

There were then 15 minutes allotted to memorisation and review. Verspoor and Lowie did not explain the procedure for the memorisation task except to describe the worksheet that was provided to the students. In order to standardise the memorisation task, the students spent 10 minutes reading aloud the translations for each target sense as a class and discussed the meaning of that Arabic word in English (Worksheet - S1 or Worksheet S3). This procedure ensured that each student would review each item. For the next 30 minutes the teacher taught a lesson on essay writing, which was considered to be unrelated to the study.

Finally the students wrote the unannounced short-term recall test (Test 2) for the final 10-15 minutes of class. The third test (Test 3) was given to the student two weeks later. They wrote this test at the beginning of the class for 15 minutes. Participants were rejected from the study if they left 12 or more of the 17 questions unanswered on any of the 3 tests. This left 55 participants who satisfactorily completed the study.

3.7 Marking

The tests were marked by five different test markers. Each marker was given an approximately equal number of Tests 1, 2 and 3. While Verspoor and Lowie used two test markers per test, only one was used per test in this study. This was done in order to minimise the amount of work asked of volunteer translators. All test markers were Arabic L1 bilinguals who were teachers of English at the university. The markers were given a marking guide which included the Arabic translations

from the worksheet which the participants studied after Test 1. The markers were directed to mark the translations as correct or incorrect. The marking guide was used to indicate the general translation, but the markers used their own discretion to decide on the appropriateness of any given translation. If they felt a second marker should look at any of the translations, the markers put a question mark beside their decision. These questionable items were marked by a second marker who also provided the researcher with a gloss translation in English for the item.

3.8 Results

Table 3.4 presents the mean scores of the original study in comparison with the results from the replication. Standard deviations are presented in parentheses. There were 18 items in the original study but only 17 in the replication. In the original study, there were 40 participants in the core sense condition and 38 in the extended sense condition. In the replication study, there were 26 participants in the core sense condition and 29 in the extended sense condition.

Table 3.4
A comparison of the mean scores for the original and replication studies

| | | Test 1 | | Test 2 | | Test 3 | |
Study	No. items	Core	Extended	Core	Extended	Core	Extended
Original	$n = 18$	10.7 (2.7)	7.8 (2.8)	17.3 (1.2)	17.2 (1.3)	13.0 (3.2)	10.7 (3.8)
Replication	$n = 17$	3.5 (2.4)	4.6 (2.4)	8.8 (2.9)	8.7 (4.4)	8.4 (3.1)	7.9 (2.8)

In the original study, there was a significant difference between the core and extended conditions for Test 1 ($p < .001$) and Test 3 ($p < .001$), but there was no significant difference between the conditions for Test 2 ($p > .05$).

For the replication study, t-tests were used to compare the means of the two groups for each of the three tests. In contrast to the original study, no significant results were found in any of the comparisons for the replication: Test 1 ($t(53)=1.6$, $p = .122$); Test 2 ($t(53)=0.15$, $p = 0.88$); and Test 3 ($t(53)=0.6$, $p = 0.54$).

The first research question asked whether the learners could guess the meaning of a polysemous sense more correctly when using a core sense as a cue. While the

original study of Verspoor and Lowie found evidence to support this hypothesis, the results of the replication did not. There was no significant difference between the two groups for any of the tests. The second question asked whether the Arabic learners of English were comparable in their L2 proficiency to the Dutch learners in the original study. The results indicated that the two groups were very different. The Dutch participants responded far more correctly than the Arabic participants on all three tests and across both treatments.

3.9 Discussion

The results of the Verspoor and Lowie's original study suggested that learners of English develop their understanding of polysemous words by extrapolating central concepts to more peripheral concepts. While the replication study found no statistical difference between the core sense group and the extended sense group, these results do not necessarily refute the findings of the original study. There are several possible explanations for the difference in findings between the original and the replication. First, the results were not marked with same level of rigour employed in the original study. The original study employed double-blind marking with a third marker to decide on differences. In the replication study, each test was marked by only one marker and a second marker was used only when the first marker questioned his or her initial decision. This may have had an effect on the results, but it's unclear why unreliable marking would result in lower scores. It is more likely that unreliability in marking would have caused greater variation within scores.

Second, it is not clear what instruction was given to the participants in the original study for the initial task (prior to Test 1) or the memorisation task (prior to Test 2). The original study reported very high results on Test 2, the short-term recall test, nearly twice that reported in the replication. Verspoor and Lowie explained that the participants were given time to review the worksheet in which the correct translations were provided. The researchers did not explain the methodology they used to review the worksheet. Did the participants review the worksheet as a group or individually? Were they given instructions on how to memorise the answers, or

were they simply told to read the worksheet over? The fact that both groups scored nearly perfect on the translations on the second, unannounced test indicates that the students paid very close attention to memorisation when reviewing the worksheet.

Third, the level of English proficiency of the participants in the replication study might have been lower than the proficiency of the participants in the original study. The context sentences were primarily taken from the New York Times, a newspaper which requires a high level of reading proficiency. One way of determining the difficulty of a reading text is by identifying what proportion of the words in a text occur infrequently in a larger corpus (Nation & Waring, 1997). According to Milton (2013), Intermediate English-learners at the B1 level in the Common European Framework of Reference for Languages (CEFR, n.d.) know about 2500 of the most frequent words in English. This is a reasonable estimate of the proficiency of the learners who participated in the replication. A word frequency analysis using the VocabProfiler (Cobb, n.d.) was conducted using the text of test items to show how frequently the words in the test occurred in the British National Corpus (BNC). The analysis showed that the first 2000 most frequent words in the BNC covered 79.8% of the text of the test items, and the first 3000 words covered 83.12% of the test items. By these estimates, the participants in the study were probably unfamiliar with about 20% of the words in the test. This indicates that the participants in the replication were unfamiliar with many more words in the tests beyond the target words. Another difficulty with the test is that the sentences were isolated from their original discourse context, allowing limited context to establish the meaning of unknown words. The original researchers mentioned this limitation; however, the sentences were not standardised according to length or to whether the context provided a directive or non-directive cue to the meaning of the target word.

Finally, the participants in the original study were native speakers of Dutch. This language has historical similarities to English that Arabic lacks. While the design of the study ensured that the polysemous English words were not ambiguous in the participants' L1, the Dutch learners of the original study might have benefitted from the similarities between English and Dutch. For example, the syntax might have

been easier for them to parse and there may have been more recognisable cognates in the sentence context of the items.

It's likely that no one factor is responsible for the difference between the results of the original study and the replication. However, the results are not rich enough to confirm how the factors discussed here affected the responses of the participants. For example, do L2 learners above a certain proficiency make better inference judgements? Was the simplified context more of an aid to inference than knowledge of the core sense? Do learners need to follow explicit inferencing methods to be successful? How does the learners' L1 influence their judgements?

What is needed is more understanding of how learners develop their knowledge of polysemous senses outside of an explicit teaching method. This would entail finding out what polysemous senses the L2 learners have already learned and what senses they haven't yet learned. With this type of data it may be possible to tease apart the factors that influence L2 learner knowledge of polysemous words.

3.10 Conclusion

In this chapter, a study was conducted to replicate Verspoor and Lowie's research into whether L2 learners can more effectively guess the meaning of an unfamiliar polysemous sense using a core sense as a cue rather than an extended sense. The results of the replication study could not confirm the findings of the original study. A number of reasons were discussed to explain the differences in the results between the two studies, including reliability in marking, differences in L2 proficiency, and the influence of the learners' L1. The learners in the original study likely benefitted from having a larger English vocabulary size and from speaking Dutch as an L1 because both the language and the culture were more similar to American English (the L2 culture) than the Arabic-speaking learners in the replication. These factors likely made it possible for the learners in the original study to establish the discourse context of the test items and interpret their meaning, leaving them to and focus solely on the target words. In contrast, the learners in the replication were likely confused about the discourse context of the test items and were unfamiliar with

several words in each test item, making it difficult for them to form hypotheses about the meaning of the target words

The results of the replication do not refute claims made by Verspoor and Lowie, but instead they raise questions as to how to effectively investigate L2 learning of polysemous senses. It is not clear whether or not learners of a lower proficiency can make use of the core-sense inferencing strategy proposed by Verspoor and Lowie. It's also not clear whether the research design presented in this replication could be reliably adapted to suit learners of a lower proficiency level. These issues are addressed in the next chapter in order to decide the most effective way to investigate how L2 learners develop their knowledge of polysemous words.

Chapter 4: Development of a new measure of L2 knowledge of polysemous words

The following chapter presents a new test of L2 polysemous meanings. In the following section, the reasons are discussed for designing a new test instead of adapting Verspoor and Lowie's instrument for learners at a lower L2 proficiency. The discussion moves on to describe the considerations taken to select the polysemous target words and how those words should be presented to best elicit knowledge of the polysemous senses. A decision was taken to address knowledge of polysemous senses through semantic acceptability judgements. In the creation of test items, one group were created to address the factor of intra-word frequency by selecting senses which occur at a range of frequencies in the language. Finally, a group of distractor items were created to elicit acceptability judgements based solely on their semantic similarity to the core sense of the word and not on prior encounters with the way the words are used in the distractors.

4.1 Rationale for design the Test of Polysemous Meanings

The replication of Verspoor and Lowie's (2003) research did not support the findings of the original study. Both studies compared how well L2 learners of English could infer the meaning of unknown polysemous senses using either the meaning of a core sense or an extended sense as a cue. The original study found that L2 learners were more successful at inferencing unfamiliar polysemous senses when using a core sense rather than a more abstract, extended sense. Moreover, the learners in the core sense condition were able to recall the inferred sentences better in a post test several weeks later. Unlike the original, the replication study found no significant difference between the core-sense and extended-sense groups for inferencing success.

The unsuccessful replication was attributed to a lower proficiency level for the participants in the replication study. Despite the difference in the results, three aspects of the Verspoor and Lowie's design could still be of use in the further investigation into how L2 learners develop their knowledge of polysemous senses. The design used a sentence-level context to establish the sense of the word. Such a presentation lends more ecological validity to the design than presenting the sense by a definition or gloss. Also, the design presented a number of different senses for

each word, which establishes the word as polysemous. Most importantly, the design prompted the participant to infer the meaning of a target sense based on the semantic similarity between the two senses as established by the sentence-level context. Given the positive aspects of the original design, and the aim of this thesis is to investigate whether L2 learners use semantic relatedness to develop their knowledge of polysemous senses, the question to be addressed next was whether to modify the original design to suit the lower proficiency of the participants or to design a new study that incorporates the useful aspects of the original.

If the original design had been kept, there were two options to modify it for a lower proficiency level. Either new target words could have been selected or the target words could have been retained but the sentence-level context adapted to the lower proficiency level. One reason the original design used infrequent target words was to ensure that the words would be entirely new to the participants. If the original words were replaced with more frequent word forms, some learners would likely know the word form in one sense or another, even with learners of a relatively lower proficiency. However, if the original target words were kept, there would be two other challenges. First, some of the target senses refer to real-world concepts that might be unfamiliar to the participants. For example, the intermediate sense for the word *peg* refers to fixing one currency to the value of another currency. Participants unfamiliar with this economic concept would be at an uncontrolled disadvantage to those who were familiar with it. Another challenge is that there is only a very fine distinction between many of the intermediate and extended senses. For example, *"boost the rent"* and *"boost the economy"* are listed under the same same sense in The Longman Dictionary of Contemporary English (LDOCE) (2010). This dictionary is suitable for upper-intermediate to advanced students and it would be a good source for context sentences at the learners' proficiency level. Moreover, the LDOCE is useful for identifying the intra-word frequency of different senses of polysemous words. The LDOCE was developed using the Longman Corpus Network which is a combination of the British National Corpus of 100 million words, the Longman Lancaster Corpus of 30 million words (Summers, 1996) and a Longman corpus of

106

spoken and written American English of 105 million words (Longman Corpus Network, n.d.). The texts for the corpus came from a wide range of both written and spoken texts and were balanced according to subject, medium, language difficulty and context. In the creation of the dictionary, the lexicographers identified different instances of a word form by part of speech, collocation and meaning sense, and they organised these instances by frequency of occurrence in the corpus (Summers, 1996).

When Verspoor and Lowie's target words are looked up in the LDOCE, only 9 of the original 18 target words have all three senses listed. For six of the target words, the LDOCE incorporates two senses in one dictionary listing. If the original target words and their three senses were retained, it would be a challenge to find or write context sentences to isolate the precise meaning of the senses. For example, in the LDOCE the target word *boost* has the sense of RAISE, "He boosted her up," and IMPROVE, "The win boosted the team's confidence." However, the dictionary doesn't include a third related-sense needed to establish the sequence of core, intermediary and extended senses.

Because of the difficulties of adapting the original design for lower proficiency learners, the creation of a new instrument was seen as the best course. The original design used words that the participants were not expected to know. However, as discussed above, it is a challenge to find words that the participants don't know but which still express a number of distinct senses whose concepts are familiar to the participants. As a consequence, this new instrument was designed with words with a high frequency in the language whose most common senses would likely be known to the participants. However, other, less frequent senses of these words would likely be unfamiliar with the learners.

In the turn towards familiar target words, the design of the new instrument changes from Verspoor and Lowie's focus on teaching new words and towards the measurement of learner knowledge of polysemous senses. In addition to Verspoor and Lowie's research, there have been other studies which have investigated the benefits of teaching methodologies informed by the principles of cognitive

107

linguistics (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007). However, no study has fully investigated whether L2 learners develop their knowledge of polysemous senses according to the principles of cognitive linguistics outside of explicit teaching methodologies. Chapter 2 presented detailed reviews of studies which focussed on L2 learning of polysemous words. These studies did not address the question of whether semantic similarity influences L2 learning of new polysemous senses according to principles set out by cognitive linguistics. Crossley et al. (2010) focussed on the relationship between senses, but the distinction between the senses was very fine-grained and did not correspond to the listing found in dictionaries nor according to cognitive linguistic principles. Kellerman (1986) did not use L2 learner language as evidence in his study of semantic transfer between languages. Finally, Schmitt's research (1998) did not focus on the semantic relationship between the senses. The following research question was proposed to direct the development of a test of polysemous meanings:

RQ   *Can a new instrument be designed to investigate whether semantic similarity between polysemous senses is a factor in how learners develop their knowledge of L2 polysemous words?*

The instrument is intended to provide a snapshot of a learner's knowledge for a selection of polysemous senses. Conclusions about development are made using a large sample of L2 participants. The results will show that some senses are known better than others. The assumption is that as their L2 knowledge progresses, the students will eventually know all the senses. Using this assumption, the results can be taken to show how well developed the students' knowledge is of each sense.

To reach any conclusions, other assumptions must be considered:
   a) partial knowledge is an indicator of future full knowledge
   b) knowledge development is linear
   c) the task reliably reveals what they do and don't know.

108

4.2 Design of the instrument

The design of the new instrument was based on consideration of four factors: which words to select, how the words should be presented to the learners, what task the learners would be required to perform, and how semantic relatedness should be measured.

4.2.1 Rationale for the selection of words

Studies researching L2 learning of polysemous words have focussed on different word classes: Kellerman (1986) focussed on the nouns *head* and *eye*; other researchers have looked at verbs (Crossley et al., 2010; Csábi, 2004; Gathercole & Moawad, 2010; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003); the design of the Word Associate Format (WAF) (Read, 1998) lends itself to the use of adjectives; and a few studies have conducted research with prepositions (Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007). Nouns were chosen for the new instrument because the semantic relationships between their senses may be more apparent to learners. Elston-Güttler and Williams (Elston-Güttler & Williams, 2008) report that L1 speakers were more reliable in their acceptability judgements for polysemous noun senses than verb senses. They also report that there is greater semantic overlap between senses for nouns than verbs. Furthermore, in investigating learner knowledge of words, there is an assumption that words are learned as discrete units with their own semantic content and not as part of a larger phrase where the semantic content can't be separated from the larger co-occurrence of language. When the different parts of speech are considered, nouns are the most likely to be separate linguistic units with their own semantic content. As Wray discusses (2015, p. 15), concrete nouns are most prototypical of the association of a word form with a concept and nouns associated with more abstract concepts are analogous with the prototype. Verbs however, are more strongly bound with their grammatical patterns and their concepts cannot be so easily isolated.

The selection of the target words needs to guard against both a floor effect and a ceiling effect. A floor effect would occur if the learners were not familiar with either the word form ("I've never seen this word before") or any sense of the word ("I don't

know what this word means"). To minimise the chances of a floor effect, word forms were chosen from the first 1000 most frequent words in English. These word forms are highly frequent in the language and, in general, 72% of the words in authentic English texts are from the first 1000 most frequent words (Nation & Waring, 1997). Word forms that are more frequent in the language are more likely to be understood and produced by L2 learners (Ellis, 1997). Furthermore, lower proficiency L2 learners use more frequent words than higher-level proficiency learners (Laufer & Nation, 1995). In contrast, a ceiling effect would occur if most of the learners knew all the senses of a polysemous word. To minimise the chances of a ceiling effect, the target words need to present senses that express a range of meaning frequency. The most frequent word forms in English express the greatest number of senses (Zipf, 1945). These are the best candidates for words that express senses with both high intra-word meaning frequency and low intra-word meaning frequency. It is unlikely that all the learners will know the most infrequent senses associated with these word forms. Thus, by choosing high frequency word forms, we can limit the possibility of both a floor effect and a ceiling effect.

4.2.2 Presentation of the target words

After the selection of words, the second factor in the design of the new instrument is the presentation of the words. Verspoor and Lowie presented the words in a sentence level context. This was seen as one of the benefits of their design because it provided ecological validity. Moreover, Hanks (2013) has argued that words should only be said to have meaning in transactional contexts in which they are used between a speaker and listener or writer and reader. The list of senses as presented in a dictionary should be considered 'meaning potential,' "potential contributions to the meanings of texts and conversations in which the words are used, and activated by the speaker who uses them," (p 73). For the target word to be meaningful in the test, it should be placed in the context of other textual discourse. It is a question just how much context is needed and of what kind. Huang (2004) investigated how well L2 learners could infer the meaning of a polysemous word when presented with different semantic cues. She presented L2 learners with a target word in a context

sentence. The context sentence was similar to those used as example sentences in a learners' dictionary. For three groups of learners, the context sentence was further elaborated with another semantic cue, either an elaborated-context cue, a semantic-frame cue or a meaning-chain cue. A fourth group of learners acted as a control. They were presented with the context sentence but no other semantic cue. Examples are presented below.

elaborated-context cue:
> My conversation with Jane came to a surprising end. I asked her about her life as a young kid.
> She could not <u>bring</u> herself to talk about her childhood.

semantic-frame cue:
> The gunman forced us into the room.
> She could not <u>bring</u> herself to talk about her childhood.

meaning-chain cue:
> (a) Remember to bring me a book.
> (b) What brings you here? What causes you to come here?
> She could not <u>bring</u> herself to talk about her childhood.

a control cue:
> [Nothing was provided]
> She could not <u>bring</u> herself to talk about her childhood.

After reading each of the following context cues, the L2 learner had to translate the meaning of *bring* in the sentence, "She could not <u>bring</u> herself to talk about her childhood." Surprisingly, the control cue was arguably as successful as the other cues. Huang found that there was no difference between the results between the learners who used the control cue compared to either the semantic-frame or meaning-chain cues. There was a difference in the results between the elaborated-context cue and the control cue. The elaborated-context cue prompted subjects to select the correct gloss more often than the control cue, but the two cues prompted an equal number of correct translations. While the elaborated-context cue was somewhat better for prompting correct understanding of the polysemous sense, the

control cue has the benefit of being shorter, and thus more items can be tested using this cue before the participants are affected by test fatigue.

The control cue is comparable to the example sentences found in dictionaries. These sentences should serve as satisfactory context to establish the sense of the polysemous target words. The example sentences from learner dictionaries are also a good resource because they attempt to strike a balance between being authentic and understandable. The example sentences in the LDOCE are based on real-life examples from the corpora, but adapted for clarity to readers with a lower language proficiency. One issue with Verspoor and Lowie's design was that the items were chosen from the New York Times. While this selection process ensured that the context would be authentic, it was felt that the context was too difficult for L2 learners at an upper intermediate proficiency level. For this reason, the decision was taken in the current study, to take sentence examples from a learners' dictionary. By doing this, senses of a target word can be established using authentic sentences, while the other word forms in the sentence are controlled for frequency appropriate to the learners.

4.2.3 Rationale for the task used in the test

A number of tasks have been developed that can be used to measure learner-knowledge of polysemous senses. One such task uses the interview method, whereby the researcher first asks the learner to explain all the meanings she knows of a word and then prompts the learner with a word-associate to measure knowledge of a specific sense (Schmitt, 1998). This is an effective measure for productive knowledge of polysemous senses, but it is limited by the time it takes to conduct oral interviews. The Word Associates Format (WAF) (Read, 1998) is another method which assesses knowledge of polysemous senses by observing whether the learner can correctly associate the polysemous word with its synonym. For example, the word *common* would be presented with the four associate options *complete, light, ordinary, shared.* The target associates, *ordinary* and *shared,* test the learner's knowledge of different senses of *common.* This method can assess receptive knowledge to the extent that the different senses of the polyseme can be associated

112

with distinct synonyms. However, the WAF does not provide context, which limits its ecological validity, and it also assumes that the learner will have knowledge of the associate synonym. (See section Section 2.1.1 for a more detailed discussion).

Semantic acceptability tests offer the most promise for testing the receptive knowledge of a large number of polysemous senses established in sentence-length contexts. These tests present the learner with sentences designed to elicit a specific sense. The task of the learner is to answer whether the polysemous word is used correctly or not. Stallman, Pearson, Nagy, Anderson and Garcia (1995) developed an acceptability test to measure the depth of word knowledge of L1 children. Each word was presented in five questions, as in this example for the word *toss*:

(1)   Is <u>tossing</u> a way of throwing?
(2)   Is <u>tossing</u> something you do gently?
(3)   Do <u>toss</u> like to fish?
(4)   Can a bell <u>toss</u>?
(5)   Can a person <u>toss</u> a real house?

It is unclear whether the question format would be a reliable measure of polysemous senses. For example, *face* has the sense of 'side of a mountain', which could be elicited through the question "Is a <u>face</u> something you can climb?" However, in order to reliably elicit this sense, the question might need to use the collocation "mountain face." If this is the case, then the success of the item is not because it is formatted as a question but because of its context.

Appel (1998) designed a test of polysemy, the Poltest, where the target word was presented in context in five different sentences. The following is an example for the word to *break*:

(6)   She broke the soup in small pieces.
(7)   He wanted to break with his family.
(8)   When his wife died, he was a broken man.

(9)   He wore a shirt with a broken collar.

(10)   The waves broke on the rocks.

Unlike the questions used in Stallman et al., Appel contextualised the target word in declarative sentences. Moreover, some of the sentences used the target word correctly, while others used the target word incorrectly. A test of this kind should be able to assess the test taker's knowledge of a word in an authentic context. Knowledge is assessed in this method by testing not only whether the test taker knows the different senses of the word but also whether she knows how not to use the word. However, a test of this kind needs to take into consideration creative extension of the target word. The use of *break* in (6) or (9) is not standard use, but it is understandable if taken creatively.

Appel found that the Poltest was successful in differentiating native Dutch-speaking children from children with a minority-language L1, finding that the minority children knew significantly fewer meanings of polysemous words, with a high reliability rating (Cronbach's alpha = 0.86) when compared to an interview method. This format offers the most possibilities for the assessment of L2 knowledge of polysemous words. However, Appel did not explain how the different meanings of the polysemous target words were selected, nor how the distractor sentences were constructed.

There are concerns around asking whether an item is correct or not. First, if the learner is able to make a meaningful sentence from a distractor, then it is a little obtuse to simply consider the response as an error. This concern is addressed in Chapter 5. Other tasks could elicit different types of responses. For example, the learner could be asked how confident she is that an item is correct or not on a Likert scale. However, this is a further complication of a task which is already novel for the learners. Some learners might understand confidence to refer to how often they use the sense while to others it might refer to their total understanding of the sentence. Despite its limitations, the task which asks if the use of a word is correct or incorrect has the benefit of simplicity and as such is less open to misinterpretation.

In consideration of the different options discussed above, a semantic acceptability task will be used in the instrument because it is the least likely to be misinterpreted by the participants. The items will present the polysemous target word in a declarative sentence which should provide enough context to make the meaning clear to a learner who already knows the target sense. However, the items need to be created in such a way that influence of semantic similarity can also be distinguished from the influence of intra-word frequency. This consideration about item creation is the next concern.

4.2.4 Measurement of semantic relatedness

The distractor items in the Poltest offer a method for addressing how semantic relatedness should be measured. The challenge of assessing the effect of semantic relatedness is to control for knowledge of the word due to other factors, such as frequency. A learner who responds that a semantically related distractor is correct, such as (6) 'She broke the soup into small pieces', indicates that she is extending her semantic knowledge of the word to novel sentences[10]. Since this sense is not acceptable in English, frequency can be ruled out as factor affecting her response. Other research has also promoted the use of novel language to control for different learning factors. The use of the semantically related distractor was proposed as early as 1943 by Cronbach. For example, he proposed that technical knowledge of an element in chemistry could be tested by asking students to select the correct examples from the list, *brass, iron, water, sulphur, fire,* and *oxygen* (Cronbach, 1943). In Rodd et al.'s (2012) research, reviewed in Chapter 2, they used a design to teach novel language to L1 speakers in order to show how new senses develop in fields such as technology. Furthermore, semantically related distractors may be more

---

10. Semantically related distractors provide interesting opportunities for discussion because a learner might respond that such a distractor is correct for creative reasons. Some measure of control will be provided through other distractors which are designed to be nonsensical. However, I will postpone a full discussion of the factor of creativity till the final discussion in chapter 12.

effective for assessing the knowledge of more advanced learners (Greidanus & Nienhuis, 2001).

The construction of semantically acceptable but incorrect sentences is by no means straightforward (Schmitt, 1994). One possible way of creating such items is through the use of 'false friends' in another language with a related etymology. The translation of a polysemous word to its equivalent in a related language is likely to share many of the same senses. However, in some cases the related language will use the equivalent word in senses that are not acceptable in English but are still semantically understandable. For example, the word *cover* can be translated into French as *couverture*. The French word can be used to mean ROOF while the English word cannot. Based on this correspondence between the two languages, a semantically acceptable distractor can be created:

(11)   The house was old and its cover needed to be repaired.

A learner who identifies a semantically acceptable distractor item as correct indicates that she might be construing the meaning of the target word, *cover*, based on its similarity to other senses of the word. However, there may still be other factors that could prompt the learner to find the item acceptable. In order to determine the strength of semantic similarity as a factor, the semantically acceptable distractor items can be compared to semantically unacceptable items, such as in the following sentence,

(12)   The people have a sense of cover in their neighbourhood.

If there are other factors that prompt the learner to find the item acceptable, then the learner is as likely to identify the semantically unacceptable item to be correct as they are the semantically acceptable items. If the participants identify the semantically acceptable distractors as correct more often than the semantically unacceptable items, then this can be taken as evidence of the effect of semantic similarity outside of the influence of intra-word frequency.

4.3 Creation of the Test of Polysemous Meanings

4.3.1 Procedure for the selection of target words

The first step in the creation of the study was to select the possible target words. A list of all the common nouns of the first 1000 words in the General Service List (West, 1953) was compiled. The words needed to express a wide range of senses, so that the most common sense of each word could be expected be well known to the learners, but that the less frequent senses would be less well known to the learners. The Longman Dictionary of Contemporary English (LDOCE) was used to screen potential items because the meanings in this dictionary are organised by intra-word meaning frequency, so that the word's most frequent meaning in the Longman Corpus Network is listed first (see page 106 for more information on the Longman Corpus Network). The number of meanings for each of these words was noted. 20 nouns were then selected which had a minimum number of five meanings: *air, arm, board, body, branch, case, character, class, course, cover, cut, face, form, hand, head, heart, line, order, point,* and *position.* The General Service List identifies high frequency words based on word families. For example, *class* and *classification* would be counted under the same family in the GSL. However, when creating the items, the only variation of the form of the target words across the example sentences was with the use of plurals. This was consistent with how nouns are presented in the example sentences from learner dictionaries.

The descriptive statistics for the number of meanings of these 20 words as nouns are as follows, ($M = 21.4$, $SD = 13.2$, $Min = 6$, $Max = 51$). However, these figures are slightly deceptive because the LDOCE uses two types of "signposts" for listing the different meanings of a word. The two types of signposts can be illustrated with this example of the first five entries for the word *hand* as a noun:

1. PART OF BODY
2. HELP
3. CONTROL
4. get out of hand
5. on the other hand

In the dictionary, each entry also contains a definition and other lexical information. However, in these examples, two types of signposts are clearly seen. One type is written in upper case letters and describes a general use; the other is written in lower case and presents a phrase containing the headword. While the LDOCE does not explain the difference, it appears that the lower case entries refer to uses that occur in more restricted, fixed phrases or idioms. For this reason, it is unclear whether these meanings are attached to the single headword or to the phrase as a whole. These meanings fall outside of the scope of this thesis, which considers polysemous words as separate orthographic units. In contrast, the meanings indicated by the upper case signposts appear to be more flexible in terms of linguistic contexts. The test items were created using these general meanings which were signposted by upper case letters. The descriptive statistics for the number of meanings of the 20 words are quite different when only the general meanings were counted: ($M = 10.2$, $SD = 3.9$, $Min = 5$, $Max = 22$). These statistics are more indicative of the number of senses available for each of the target words.

Three senses were selected for each target word. The senses were selected according to relative intra-word frequency, so that the first sense was the most frequent sense; the second was the middle-frequency sense (selected from one of the middle-ranking senses); and the third sense expressed low-frequency (selected from one of the low-ranking senses). As an example, the list of senses for *class* in the LDOCE is as follows: 1 SOCIAL GROUP, 2 STUDENTS, 3 TEACHING PERIOD, 4 STUDYING, 5 SAME TYPE OF SOMETHING, 6 TRAIN/AIRCRAFT ETC, 7 QUALITY, 8 STYLE/SKILL, 9 UNIVERSITY DEGREE. The list is ordered by relative frequency with 1 SOCIAL GROUP being the most frequent sense and 9 UNIVERSITY DEGREE being the least frequent sense. In the final test, 1 SOCIAL GROUP was selected as the high-frequency sense, 3 TEACHING PERIOD was selected as the middle-frequency sense, and 8 STYLE/SKILL the low-frequency sense. A full list of the senses used in the instrument is presented at the end of the chapter in Table 4.1.

4.3.2 Creation of Acceptable Items

For each target word, three senses and their example sentences were chosen: the most frequent sense, one of the least frequent senses, and one sense of middle frequency. In this case, frequency refers to the intra-word frequency of one sense relative to the other senses of the word, not frequency within a corpus representing the language as a whole. Idiomatic senses were excluded from consideration. Idiomatic senses were considered to be those senses which were highly constrained by collocation and lacking flexibility of use. For example, 'head' can be used in the idiom '[one person] goes over [another's] head,' in reference to circumventing established hierarchy. This sense has to be used with one person in opposition to another. However, a slight modification to the structure, "The lecture went over my head," and the idiom now refers to intellectually challenging subject matter. Idioms such as these were not included as senses. One exception to this rule was the most common sense of 'order', which was "in order". The LDOCE signposts the meaning of this sense as FOR A PURPOSE. This sense was chosen because of its frequency across many different contexts. In some cases, the LDOCE didn't give an example sentence. Where no example sentence was provided, sentences were chosen from the Cambridge Advanced Learner's Dictionary (Cambridge Advanced Learner's Dictionary, 2011). The list of acceptable items can be found in Appendix 2.1, on page 333.

4.3.3 Creation of Logical Distractors

So far in the process, 20 target words were selected, for each of which, three acceptable items were created representing three different senses. These acceptable items were created using example sentences taken from learner dictionaries. In addition to these three acceptable items, two distractor items were also created for each target word: a logical distractor and an illogical distractor. The logical distractor represented a semantically acceptable but incorrect use of the word, and the illogical distractor represented a semantically unacceptable use of the target word.

To create the logical distractors, I consulted Oxford foreign language dictionaries for languages with historic similarities to English: French (Oxford-Hachette French

Dictionary, 2011), Spanish (Oxford Spanish Dictionary, 2011), Italian (Oxford-Paravia Italian Dictionary, 2011) and German (Oxford-Duden German Dictionary, 2011). The participants who took part in the studies, discussed in later chapters, were Arabic L1 speakers and didn't know any of the languages used to create the distractors. The closest equivalent to the core sense of each target word was indexed from these dictionaries. 'False friends' from these dictionaries were identified for each target word and an example sentence was written using this sense. For example *cover* in French is *couverture*, which can be used to mean ROOF, as in "The house was old and its cover needed to be repaired." This sentence is referred to as a logical distractor, because while the use of the target word *cover* in the sentence is incorrect, the use could be logically extended from core sense of "something that protects or conceals". The list of logical distractors can be found in Appendix 2.1, and Table 4.1, on page 126, glosses the meaning of the logical distractors.

4.3.4 Creation of Illogical Distractors

The second type of distractor is referred to as an illogical distractor. In these sentences the target word is used in a sentence that does not make sense to a native speaker, for example, "The people have a sense of <u>cover</u> in their neighbourhood." Example sentences were chosen from the LDOCE. The look-up word for the example sentence needed to be meaningfully different from the target word. For example, *pride* and *cover* are ontologically different in their central meanings. The target word, *cover* in the example, is then substituted for the look-up word, *pride*. The resulting sentence is then checked by the researcher to make sure that no novel meaning could be construed. For example, when the target word *body* was substituted for look-up word *watch*, the resulting sentence was "He glanced nervously at the body." Since the meaning of this sentence could be acceptably construed, the sentence was altered to "He glanced nervously at the body to check the time." The list of illogical distractors can also be found in Appendix 2.1, and Table 4.1 presents the original word that the target word replaced in the logical distractors.

4.3.5 Item revision for learner proficiency and native-speaker acceptability

All the example sentences, for both the acceptable items and the logical and illogical distractors, were run through a frequency profiler, an online application which identifies which words in a text are found in the General Service List (Cobb, n.d; Heatley, Nation, & Coxhead, 1994). The General Service List (GSL) is a list of 2,000 high-frequency word families in English (West, 1953). Words outside of the GSL were substituted with GSL words. In the example above, "glanced nervously" was substituted with "looked". Proper nouns were retained. Two native speakers reviewed the sentences and changes were made based on their suggestions about the clarity and appropriateness of the distractor sentences. For example, one native speaker was unfamiliar with the use of *branch* to mean a secondary train line. This meaning was substituted with *branch* to mean a tributary river. The items were then piloted by a different group of three native speakers, two North American and one British. All three native speakers identified all the acceptable items as "Correct" and all the distractor items as "Incorrect".

4.3.6 Creation of the task for the test

The semantic acceptability task used in the test presented the participants with the five items for each target word. The participants were asked to decide whether or not the meaning of the word was used correctly in the sentence by checking a box for CORRECT or INCORRECT. In order to be confident that the participant had made a decision one way or the other a DON'T KNOW option was included to allow participants a choice if they couldn't make a decision. The following instructions were included with the test to explain the task:

Instructions

In this study, you will read sentences with one WORD in capitals. Please decide if the meaning of the WORD in capitals is correct or incorrect.

In the example, the word in capitals is LIFE. In Sentence 1 and Sentence 3, the word LIFE is used correctly, so the CORRECT boxes are checked. In Sentence 2, the word LIFE is used incorrectly, so the INCORRECT box is checked.

1. It was one of the best days of my LIFE.	CORRECT ☑	INCORRECT ☐	DON'T KNOW ☐
2. He was sitting in a LIFE.	CORRECT ☐	INCORRECT ☑	DON'T KNOW ☐
3. She was enjoying married LIFE.	CORRECT ☑	INCORRECT ☐	DON'T KNOW ☐

In all the sentences, the grammar is correct. Please think about the meaning of the WORD in capitals.

If you don't know whether the word is used correctly or not, you can check the DON'T KNOW box.

The test is designed to be marked according to signal detection theory (Shillaw, 2009), whereby an acceptable item responded to as 'Correct' is marked as a 'Hit' and as a 'Miss' when responded to as 'Incorrect'. In contrast, distractor items are marked as 'False Alarms' and 'Correct Rejections' when responded as 'Correct' and 'Incorrect' respectively.

4.4 Summary: Design of the Test of Polysemous Meanings

The Test of Polysemous Meanings (TPM) was developed with the following characteristics:

- Twenty target words were selected from the first 1000 most frequent words in the GSL.

- Each expressed at least $n = 5$ different senses.

- For each target word, three acceptable senses were selected based on intra-word frequency: high-frequency sense, a middle-frequency sense and a low-frequency sense.

- The context sentences for the acceptable items were selected from LDOCE and Cambridge Advanced Learner's Dictionary

- A logical distractor for each target word was created in a three-step process:

  1. The translation equivalent of the target word was selected from a related language to English.

122

2.   A sense of that translation equivalent was identified if it was not acceptable in English but understandable based on semantic similarity to the category of the polysemous word.

3.   An example sentence was found in the LDOCE to present the target word for the unacceptable but understandable sense.

- An illogical distractor was created for each target word by substituting the target word in an example sentence that rendered the meaning nonsensical.

- A task was created to require the participants to read each item and judge whether the meaning of the target word was 'Correct', 'Incorrect', or to select 'Don't Know' if no judgement could be made.

- Three native speakers reviewed the acceptable items for acceptability and the distractors for unacceptability and confirmed all items as suitable.

The final test of 100 items is included in Appendix 2.1. Table 4.1 (placed at the end of the chapter on page 126) presents the target words selected for the Test of Polysemous Meanings and the senses for the five categories of items. The senses are presented in capital letters. The sense description includes the headword taken from the LDOCE used to gloss the sense and, in square brackets, the domain where that sense of the target word is used. For example, the sense of *air*, as in "Trudy is always putting on airs," is glossed as BEHAVIOUR and is understood to occur in the domain of [SOCIETY]. The same presentation is used to present the logical distractors, but the illogical distractors are only presented with the word used in the substitution of the sentence. For example, the illogical distractor for *air* is "Birds lay their *air* in the spring." This sentence was taken from an example sentence for *eggs*. In the table, EGGS is presented in the logical distractor column for the target word *air*. The next chapter presents the procedure and results of the pilot of the Test of Polysemous Meanings (TPM).

The design of the TPM was motivated by the results of a replication of Verspoor and Lowie's original research (see Chapter 3). The original study was interesting because it showed how L2 learners could correctly infer the meaning of an unknown sense

based on the semantic similarity to a known sense. However, the results of the replication indicated that the original design was incompatible with learners of English at a lower L2 proficiency. As a consequence, TPM was designed to show that learners at a lower proficiency also use semantic similarity to infer the meaning of an unknown sense. Instead of a translation task, the TPM uses a semantic acceptability task for acceptable items, logical distractors and illogical distractors. It is expected that the learners will judge the logical distractors as acceptable more often than the illogical distractors; this will be taken as evidence that the learners use semantic similarity to infer the meaning of unfamiliar senses. The responses to the logical distractors will be used to address the first research question, which was presented in Chapter 2 on page 89:

To what extent is semantic similarity between polysemous senses a factor in how learners develop their knowledge of L2 polysemous words?

With the change from a translation task to a semantic acceptability task, the TPM also differs from Verspoor and Lowie's original design with a change from teaching polysemous words to a focus on knowledge of polysemous words. Knowledge is here understood with reference to Sandra and Rice's distinction between how concepts are processed in the mental lexicon in contrast to how they are structured (see page 34). While evidence for how semantic knowledge is structured relies on online tasks, such as through the effects of semantic priming, the semantic acceptability task of the TPM is conducted offline and is designed provide evidence for how semantic knowledge is processed. In this way knowledge of polysemous words is understood to include the use of inference based on stored information.

Previously, L2 learners' knowledge of polysemous words has been investigated through longitudinal studies (see the research of Crossley et al., 2010; and Schmitt, 1998, as reviewed in Section 2.5.1). In these studies, the knowledge of L2 polysemous words was shown to develop by measuring the same participant at intervals over a period time long enough for learning to take place. In contrast, the TPM is cross-sectional in its design. As such, the development of L2 polysemous knowledge will

be based on the comparison of learners at different levels of proficiency. The expectation is that the participants with a higher proficiency will also score higher on the the TPM than those at a lower proficiency. Furthermore, the assumption is that as the lower proficiency participants improve their overall language proficiency, they will also perform similarly to the high proficiency group on the TPM. The measurement of proficiency will be addressed in Chapter 8, where the scores on a vocabulary size test are presented for the same participants who took the TPM. These scores will be used to address the effect of vocabulary size on the results of the TPM.

The effect of vocabulary size on the TPM is directly referred to in the second research question, also presented in Chapter 2 on page 89:

What is the effect of the following factors on L2 knowledge of polysemous words: the intra-word frequency of different senses, the language learner's L2 vocabulary size, and the influence of the learner's L1?

The design of the TPM takes is controlled for intra-word frequency. Each target word is represented by three acceptable senses, categorised by how frequent each sense is relative to other senses of the same word. The prediction is that L2 learners will find the senses more acceptable as they express greater intra-word frequency. The other two factors addressed in the second research question, L2 vocabulary size and L1 influence, will be addressed using additional material.

Finally, Chapter 10 presents a translation of the senses used in the TPM. Using these translations, the results of the TPM are reanalysed for the effect of L1 influence.

Table 4.1
The target words and senses used in the Test of Polysemous Meanings.

	High-frequency sense	Middle-frequency sense	Low-frequency sense	Logical distractor	Illogical distractor
air	GAS [ATMOSPHERE]	PLANES [ATMOSPHERE]	BEHAVIOUR [SOCIETY]	SPACE [PHYSICAL-SPACE]	EGGS
arm	BODY [BODY]	WEAPONS [MILITARY]	FURNITURE [FURNITURE]	HORSE'S BODY [HORSE]	YEARS
board	INFORMATION [INFORMATION]	GROUP OF PEOPLE [GOVERNANCE]	MEALS [LODGING]	FLOWER BED [GARDENING]	BAG
body	[PEOPLE/ANIMALS]	GROUP [PEOPLE]	body of something [KNOWLEDGE]	SUBSTANCE [COOKING]	WATCH
branch	OF A TREE [TREES]	OF GOVERNMENT [INSTITUTIONS]	OF A ROAD, RIVER, ETC. [ROADS,RIVERS, ETC]	INDUSTRY [INDUSTRY]	FORTUNE
case	SITUATION [EVENTS]	LAW/CRIME [LEGAL]	SUITCASE [LUGGAGE]	CASH REGISTER [MONEY]	DRESS
character	ALL SOMEBODY'S QUALITIES [PERSON]	MORAL STRENGTH [PERSON]	LETTER/SIGN [ORTHOGRAPHY]	APPEARANCE [PERSON]	PRESCRIPTION
class	SOCIAL GROUP [SOCIAL HIERARCHY].	TEACHING PERIOD [TEACHING]	STYLE/SKILL [QUALITY]	MONETARY RATE [MONEY]	PIN
course	EDUCATION [EDUCATION]	PLANNED DIRECTION [TRAVEL]	MEDICAL TREATMENT [MEDICINE]	MONETARY RATE [MONEY]	SCALE
cover	PROTECTION [PHYSICAL OBJECTS]	BED [BED]	WEATHER [WEATHER]	ROOF [BUILDINGS]	SECURITY
cut	REDUCTION [BUDGETS]	HOLE/MARK [PHYSICAL OBJECTS]	SHARE OF SOMETHING [MONEY]	FEATURE (OF A PRODUCT) [COMMERCIAL PRODUCTS]	SUNLIGHT

126

face	FRONT OF YOUR HEAD [BODY]	PERSON [BODY]	CLOCK [TIME PIECES]	ATTENTION [BODY]	BUS
form	TYPE [CATEGORIES OF THINGS]	SHAPE OR CONFIGURATION OF SOMETHING [PHYSICAL OBJECTS]	PERFORMANCE [PERFORMANCE]	OVERWEIGHT [BODY IMAGE]	POTATOES
hand	PART OF BODY [PERSON]	HELP [PERSON]	WORKER [PERSON]	WAY OF DOING SOMETHING [PERSON]	FLOWER
head	TOP OF BODY [PERSON]	CALM/SENSIBLE [PERSON]	FRONT/LEADING POSITION [TABLE]	EMOTIONS [EMOTIONS]	ROOM
heart	BODY ORGAN [BODY]	EMOTIONS/LOVE [EMOTIONS]	ENCOURAGEMENT [MOTIVATION]	CORE [APPLE]	CAR
line	ON PAPER/ ON THE GROUND [PHYSICAL OBJECTS]	WAY OF DOING SOMETHING [THOUGHT]	FAMILY [ANCESTRY]	PERSON'S FIGURE [PERSON]	RADIO
order	FOR A PURPOSE [PLANS]	ARRANGEMENT [THINGS OR EVENTS]	SOCIAL/ECONOMIC SITUATION [SOCIETY]	RANGE [MONEY]	PAPER
point	A SINGLE IDEA IN AN EXTENDED WHOLE [IDEAS]	PLACE [PLACES]	SHARP END [INSTRUMENTS/ TOOLS]	A VERY SMALL PART OF SOMETHING [COOKING]	WASH
position	SITUATION [EVENTS]	PLACE WHERE SOMEBODY/ SOMETHING IS [GEOGRAPHIC AREA]	JOB [EMPLOYMENT]	BALANCE [BANKING]	VOLUME

Chapter 5: Pilot study of the Test of Polysemous Meanings

The Test of Polysemous Meanings, created in the previous chapter, is here given to a group of L2 learners as a pilot study. The objective was to see if the learners judged the acceptability of items according to two predictions. First, it was predicted that polysemous senses with greater corpus frequency would be judged as acceptable more often than the less frequent senses, and second, that the logical distractors would be judged as acceptable more often than the illogical distractors. The test was given to 39 Arabic speaking learners of English studying in a pre-university course of intensive English.

5.1 Introduction

In the previous chapter I presented the design of a new measure of L2 knowledge of polysemous meanings, referred to as the Test of Polysemous Meanings (TPM). The TPM was developed in response to the results of a replication of Verspoor and Lowie's research (2003). The replication study did not corroborate the findings of the original study and this result was primarily attributed to the lower proficiency level of the participants in the replication. While Verspoor and Lowie's research was based on a teaching intervention for L2 polysemous words, the TPM was designed as a measure of L2 knowledge of polysemous meanings. The change from a teaching intervention to a measurement of knowledge was made in part because of the constraints imposed by the lower proficiency level of the participants. However, the TPM was also developed because there is little known about how L2 learners develop their knowledge of polysemous senses outside of explicit teaching.

This change to a measurement of L2 knowledge of polysemous meanings is in line with the aims of this thesis as articulated in Chapter 1, to investigate whether L2 learners use semantic relatedness to develop their knowledge polysemous senses. Two crucial hypotheses have emerged from the investigations in the previous three chapters: an intra-word frequency hypothesis and a semantic similarity hypothesis. The intra-word frequency hypothesis proposes that learners are more likely to learn meanings according to the frequency of the different meanings of a word relative to other senses of that word. The semantic similarity hypothesis proposes that learners

extrapolate their knowledge of words from one sense to another. Though grounded in previous research, these represent new conceptual approaches to the problem of how polysemous knowledge is acquired, and will shape the development and application of the TPM.

To investigate these hypotheses, the TPM was designed with two types of test items. The first type were the acceptable items (see 4.3.2 above); these were designed based on the intra-word frequency of meanings. For each of the 20 target words there is an item for a high-frequency sense, a middle-frequency sense and low-frequency sense. The second type were the distractor items (see 4.3.3 and 4.3.4); these were designed to be independent of the effect of frequency. Each target word is assessed by a logical distractor and an illogical distractor. The logical distractor was designed to be a semantically understandable but incorrect use of the target word, while the illogical distractor was designed to be semantically unacceptable. The design of the TPM forms the basis for the following two research questions:

RQ1   *Do the learners respond to the frequent senses with more 'Hits' than to the infrequent senses?*

RQ2   *Do the learners respond to the logical distractors with more 'False Alarms' than to the illogical distractors?*

I decided to focus on 'Hits' and 'False Alarms' because these responses indicate that the learner has made the confident decision that the target word is used meaningfully and acceptably in the context of the sentence. This method of marking is based on signal detection theory (see Shillaw, 2009 for a discussion of its history). The participant's decision can be said to be confident because the option of the 'Don't Know' response was available to indicate a lack of confidence. The decision can also be said to be meaningful and acceptable because in both cases the participant has responded that the use of the target word in the sentence is 'Correct'. One might object to RQ2 because it focuses on 'False Alarms' rather than on 'Correct Rejections'. 'Correct Rejections' may be seen as more valid because this response would be in agreement with the native speakers' response. However, there are two reasons why a participant might say that a logical distractor is incorrect. On the one hand, the

participant might have construed the target word meaningfully, but rejected the use, while on the other hand, the participant might not have seen any meaningful use of the target word at all. To focus on the 'Correct Rejections' would be unreliable because this response does not indicate whether the participant has construed the target word meaningfully or not. In contrast, the 'False Alarm' response indicates unambiguously that the participant finds the use of the target word both meaningful and acceptable.

5.2 Study

5.2.1 Participants

The participants were Arabic learners of English studying in an English medium university in the Arabian Gulf. They were all female, aged 18-20 years old, and had studied English through their K-12 Qatari curriculum. At the present time, the participants were studying 20 hours a week in an intensive L2 English programme. They were in the final level of the programme before entering their university courses, conditional upon an IELTS score of 5.0, which places them at the proficiency level B1 (Independent User) in the Common European Framework of Reference. The participants were from three intact classes and a total of 39 participants completed the test. This was considered a sufficient sample size; however, given the homogeneity of the participants, there is a concern that the results would not extend to populations of different educational backgrounds.

5.2.2 Procedure

There were 100 items in total: five sentences for 20 target words. The items were presented in blocks of five, and the order of the items was mixed so that no target word was repeated within any five-sentence block. Thus, the participants would not be expected to deduce the number of acceptable items or distractors for the target words. In each sentence the target word was capitalised to highlight it from the rest of the sentence. Instructions directed the participants to decide whether the use of the target word in the sentence was "Correct" or "Incorrect". A third option of "Don't Know" was included as an option for participants who couldn't decide.

131

The study was presented to the participants through the Internet application "Survey Monkey" (2011). Five items were presented at one time on the monitor screen, as shown in the screen shot (Figure 5.1).

The application randomised the presentation of the five-item pages so that each participant had item-blocks presented in a different order. The participants were required to make a decision on all the sentences. The study was presented over two classes with one week apart. Half the study (50 sentences) was presented in each class. This was done because participants may have lost focus when performing 100 decisions in a single sitting.

The participants were given as much time as needed to complete the study. All completed each set of 50 sentences within 30 minutes. No dictionaries or other aids were allowed.

Figure 5.1
Example of the online survey application.

5.3 Results

The mean number and standard deviation of 'Don't Know' responses was 17.54 (*SD* = 16.98). Two participants were removed from the analysis because the number of times they responded with 'Don't Know' was more than two standard deviations from the mean: 56 of 100 and 86 of 100. The responses of these two participants would have lacked discriminatory detail (between acceptable and unacceptable responses). This left 37 participants. The revised mean number and standard deviation of 'Don't Know' responses was 14.73 (*SD* = 11.68).

The results of the Test of Polysemous Meaning were scored according to a system used in detection theory (Shillaw, 2009). If an acceptable item was answered as 'Correct' it was marked as a 'Hit', but if it was answered as 'Incorrect' then it was marked as a 'Miss'. If a distractor was answered as 'Correct' it was marked as a 'False Alarm', but if a distractor was answered as 'Incorrect' it was marked as a 'Correct Rejection'.

The results for the acceptable items on the TPM are presented in Table 5.1 and Figure 5.2 . They present the mean response rate for the three acceptable item-types across the 20 target words for the 37 participants. The standard deviation is presented in brackets.

Table 5.1
Results of the acceptable items on the TPM by sense frequency (n = 37)

Sense	Hits	Misses	Don't Know
High	15.22 (3.49)	3.03 (2.03)	1.76 (2.20)
Middle	12.54 (3.14)	4.41 (2.17)	3.05 (2.91)
Low	10.92 (2.78)	5.70 (2.97)	3.38 (2.72)
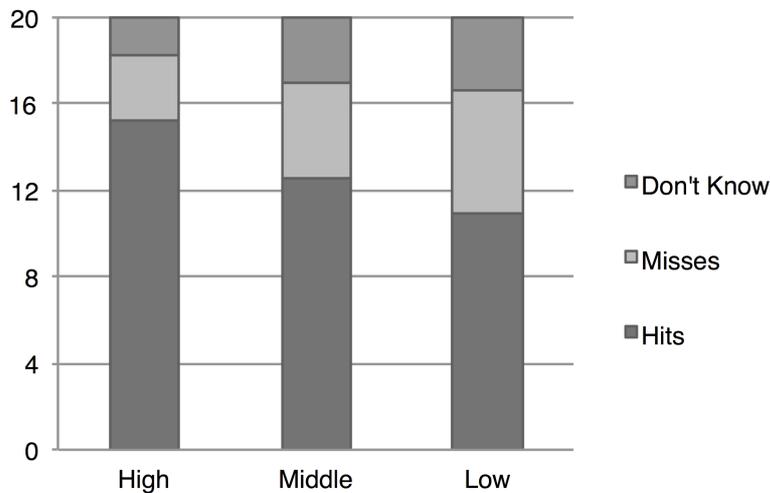
Figure 5.2
Mean number of responses by participant for the acceptable items.

The mean number of 'Hits' for the acceptable items was 15.2 (*SD* = 3.49) for the high-frequency sense, 12.5 (*SD* = 3.14) for the middle-frequency sense, and 10.9 (*SD* = 2.78) for the low-frequency sense.

The results for the distractor items on the TPM are presented in Table 5.2 and Figure 5.3. They present the mean response rate for the two distractor item-types across the 20 target words for the 37 participants. The standard deviation is presented in brackets.

Table 5.2
Results of the Distractor Items on the TPM by sense frequency (n = 37)

Sense	False Alarms	Correct Rejections	Don't Know
Logical	9.73 (3.93)	7.54 (3.48)	2.73 (2.69)
Illogical	6.65 (3.41)	9.54 (3.11)	3.81 (3.26)
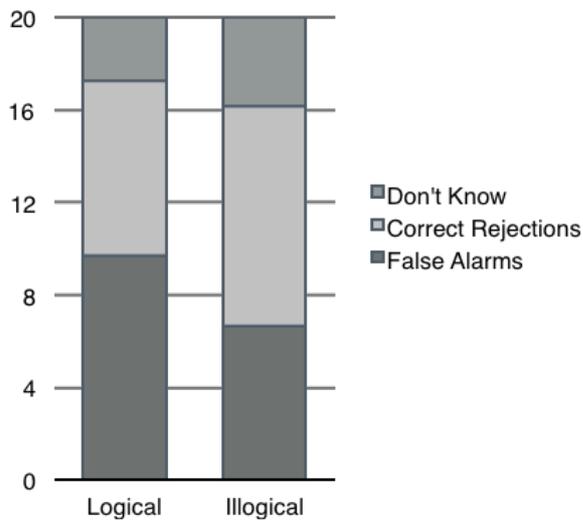
Figure 5.3
Mean number of responses by participant for the distractor items.

The mean number of 'False Alarms' for the logical distractors was 9.73 ($SD = 3.93$), compared to 6.65 ($SD = 3.41$) for the illogical distractors.

5.4 Analysis of the results

The results were analysed by a comparison of means for the 'Hits' and 'False Alarms', respectively for the acceptable items and distractor items. The 'Hits' correspond to the participants responding that acceptable items were correct. The 'False Alarms' correspond with the participants responding that distractors were correct.

5.4.1 Test for Normality of the Data

The objective of the analysis was to determine whether the mean number of 'Hits' or 'False Alarms' of one item type were significantly different from the mean number of the other item types. As a precursor to this analysis, the data were first analysed for normality to determine whether parametric or non-parametric tests should be used. The Shapiro-Wilk test of normality was conducted for the results of the Pilot TPM study for item type by participant. This test tests the null hypothesis that the results came from a normally distributed population. Importantly, a significant result less than $a = .05$ indicates that the data is not normally distributed. The results

135

are presented in Table 5.3 for the normality of 'Hits' to the acceptable items and in Table 5.4 for the normality of 'False Alarms' to the distractors.

Table 5.3
Shapiro-Wilk Test of Normality for the 'Hits' on the Pilot TPM

Item type	Statistic	df	Sig.
High	0.924	37	0.015
Middle	0.947	37	0.078
Low	0.963	37	0.258

Table 5.4
Shapiro-Wilk Test of Normality for the 'False Alarms' on the Pilot TPM

Item type	Statistic	df	Sig.
Logical	0.955	37	0.144
Illogical	0.971	37	0.447

For the acceptable items, the distribution of 'Hits' by participant for the high-frequency sense was significantly different from the null hypothesis of normality at $p > .05$. Furthermore, the skewness value for the high-frequency sense and middle-frequency sense were significantly skewed (*skew(high)* = -.789, z = -2.03, $p < .05$; *skew(middle)* = -.794, z = -2.05, $p < .05$). The skewness indicates some of the of the participants correctly identified most of the high and middle-frequency senses. This result was not unexpected for the high-frequency senses because the test was designed with the expectation that the participants would know at least one sense of the target words. The results to the middle-frequency senses were negatively skewed due to a few participants who scored with very few 'Hits' to these items. Non-parametric tests were used for comparisons with the acceptable items. For the distractor items, the distribution of 'False Alarms' by participant was not significantly different from the null hypothesis of normality. Parametric tests were used when comparing the results for the distractor items.

5.4.2 Comparison of the acceptable items

RQ1   *Do the learners respond to the frequent senses with more 'Hits' than to the infrequent senses?*

With reference to Figure 5.2 above, the results show that more higher frequency items were answered correctly than lower frequency items. A non-parametric Friedman test of differences among repeated measures found that there was a significant difference between the three levels for the acceptable items, Friedman χ^2 $(2, N = 37) = 29.84$, $p < .001$. Post-hoc tests with a Bonferroni correction ($\alpha = 0.017$) were conducted to compare the means of the separate item types. The number of 'Hits' for the high-frequency sense ($M = 15.22$, $Mdn = 15.22$, $SD = 3.49$) was significantly higher than both the middle-frequency sense ($M = 12.54$, $Mdn = 13.00$, $SD = 3.14$), $z = 2.92$, $p = .004$, and the low-frequency sense ($M = 10.92$, $SD = 2.78$), $z = 4.73$, $p < .001$. The number of 'Hits' for the middle-frequency sense was significantly higher than the low-frequency sense, $z = 2.28$, $p = .002$. This result confirms the hypothesis of intra-word frequency. The participants identified the use of the target words as both meaningful and acceptable based on their intra-word frequency. However, the strength of this result can be questioned. First of all, in nearly 25% of the cases for the high-frequency sense, the participants did not identify the target word as correct. The target words were very common words in the language and it is unlikely that the participants did not know the most common meaning of the target word for one quarter of the words. For some of the cases, the participant likely knew the meaning of the high-frequency sense but could not construe that meaning in the given sentence context. Also, the difference in 'Hits' between the high and low-frequency senses is not that great. If the results were solely due to intra-word frequency, then one would expect that the low frequency senses would have had far fewer 'Hits'. There is the possibility that for some of the low-frequency sense items, the participant was able to make a meaningful sentence based on the similarity sense to the core sense of the target word, despite having not encountered the sense before.

5.4.3 Comparison of the distractor items

Figure 5.3 above shows that the participants responded with more 'False Alarms' to the logical distractors than the illogical distractors. A paired-samples T-test was conducted to compare the number of 'False Alarms' between the two distractor items for the pilot of the TPM. The mean number of 'False Alarms' for the logical distractors ($M = 9.81$, $SD = 3.96$) was significantly higher than the mean number for the illogical distractors ($M = 6.83$, $SD = 3.27$), $t(35) = 5.26$, $p < .001$. This result lends support to the semantic similarity hypothesis, because the logical distractors, which were designed to be semantically related to the core sense of the target word, were identified as correct more often than the illogical distractors. Furthermore, the mean number of 'False Alarms' to the logical distractor (9.81) was comparable to the mean number 'Hits' for the low-frequency sense (10.92). This similarity supports the proposition made above that semantic similarity might have been a strong factor in the size of the results for the acceptable items as well. One other point to note here is the number of 'False Alarms' for the illogical distractors. These were items in which the target word was used incorrectly and with no semantic relationship to acceptable senses. The fact that these items were identified as correct one third of the time raises a reliability issue. This result indicates the participants might have responded due to creative factors that the test design had not controlled for. The exact nature of these creative factors would be difficult to predict and might have varied from participant to participant. However, just as the participants made creative interpretations for the illogical distractors, they may also have made similar creative decisions for the other item types. In future, an adjustment will be made to the results to account for uncontrolled, creative decisions. This consideration is discussed more thoroughly in

5.5 Discussion

The purpose of this chapter was to see whether the TPM could measure L2 learners' knowledge of polysemous words. Specifically, the study was designed to indicate whether the learners' understanding of polysemous senses could be explained simply by intra-word frequency, or whether the learners used their current knowledge of the senses of a word to interpret the meaning of an unknown sense through semantic similarity.

As predicted, intra-word frequency was a strong influence on whether learners knew a sense or not. For each word, three acceptable senses were tested, a high-frequency sense, a middle-frequency sense and a low-frequency sense. The acceptable items representing these three senses were more correctly identified in order of frequency, with the high-frequency sense being significantly different from the middle- and low-frequency senses. Thus, the results supported the hypothesis that intra-word frequency is a factor influencing which senses the learners know.

However, the results also provide evidence that learners extrapolate their knowledge of words from known meanings to unknown meanings. The participants identified the logical distractors as correct significantly more often than the illogical distractors. The logical distractors were designed to be incorrect but meaningful based on semantic similarity to the core sense of the polysemous target word, whereas the illogical distractors were designed to bear no semantic similarity to the polysemous target word. These results indicate that the participants were more likely to construe the sense of the target word meaningfully in the context of the logical distractors than in the context of the illogical distractors. Since both distractors are incorrect uses of the word, the participants could not have encountered these uses of the word elsewhere. The semantic similarity to other senses of the target word can explain why the logical distractors were identified as correct more often than the illogical distractors.

Semantic similarity is not the only factor affecting the acceptability judgements for the distractors. The mean number of 'False Alarms' for the illogical distractors was

6.65 responses for the 20 items ($SD = 3.41$). The question must be asked why the participants should have found the meaning of the target words acceptable in sentences which were constructed to be nonsensical. Relevance theory can provide one possible explanation. This theory of pragmatics maintains that the speaker's intention of a given utterance is only partly explained by decoding its linguistic parts. Much of the speaker's intention is inferred by the hearer because "utterances automatically create expectations which guide the hearer toward the speaker's meaning," (Wilson & Sperber, 2005). In relation to the illogical distractors, Wilson and Sperber's proposition supports the idea that L2 learners will expect each sentence in the TPM to be meaningful. Unlike the native speakers who judged these sentences to be unacceptable, the L2 learners do not have enough linguistic knowledge to inhibit the expectation that sentences should be meaningful.

The illogical distractors were designed to express no semantic similarity to the core sense of the target word. Since the participants responded with 'False Alarms' to these items, they did so despite the lack of semantic similarity. For this reason, the results for the other items cannot be attributed solely to the similarity to the core sense or intra-word frequency. In some cases, the participants might have responded with a 'False Alarm' to the logical distractors for the same reasons that they did to the illogical distractors. The same is true for the acceptable items, which also may have been responded to with a 'Hit' for uncontrolled reasons. Meara and Buxton (Meara & Buxton, 1987) developed a Yes/No test of word knowledge which also measures the number of 'False Alarms' to distractors. In the case of their test, the distractors were non-words that followed spelling conventions. In their test, a 'Hit' was a target word identified correctly as a real word and a 'False Alarm' was a non-word identified incorrectly as a real word. Meara and Buxton proposed the possibility that the participants responded with a 'Hit' to some target words which they did not know. To address this issue, they used the number of 'False Alarms' to the non-words to adjust downward the number of 'Hits' to the target words. The result is meant to be a more reliable measure of what words the test-taker knows. The same method can be used when calculating the results to the acceptable items

for the TPM, by adjusting the number of 'Hits' downwards according to the number of 'False Alarms' to the illogical distractors. For the TPM, the adjustment would reduce the number of 'Hits' that could be attributed to uncontrolled creative factors. The resulting, adjusted number of 'Hits' would be a more reliable measure of knowledge due to either intra-word frequency or to semantic-similarity compared to other known senses.

It would not be appropriate to adjust the results for the distractor items downwards. Obviously, the results to the illogical distractors should not be adjusted downwards because the 'False Alarms' to these items are used as the measurement of uncontrolled, creative judgements. Likewise, results to the logical distractors should not be adjusted downwards because the results to these items are already compared to those of the illogical distractors. It would be more appropriate to measure the effect size when comparing the logical distractors to the illogical distractors. The effect size would indicate how much the semantic similarity present in the logical distractors affected the participants' responses in comparison to uncontrolled, creative factors as measured by their responses to the illogical distractors[11].

While the results to the distractor items indicate that semantic similarity can explain the difference in the number of 'False Alarms', the greater interest is for the influence of semantic similarity on the participants' responses to the acceptable items. The difficulty in measuring the influence on the responses to acceptable items is to separate the factor of semantic similarity from other factors such as intra-word frequency. One possible way to investigate the effect of semantic similarity on the acceptable items is to present the test items so that the participants have the opportunity to compare one use of a target word to the other uses. If there is a greater number of 'Hits' for the acceptable items in a grouped presentation compared to a mixed presentation, then this can be taken as evidence that the participants used semantic similarity to help them decide if target word was used

---

11. A measurement of effect size will also be reported for the comparisons among the acceptable items.

correctly or not. In the pilot study discussed in this chapter, the test items were presented to the participants in a mixed order. This meant that the participants did not have the opportunity to compare the use of one target word across different items. In Chapter 7, I will discuss the effect of a grouped presentation compared to a mixed presentation and whether semantic similarity is an influence when the participants have the opportunity to compare items to one another.

Finally, even though the results to the pilot study supported the hypotheses of intra-word frequency and semantic similarity, we cannot rule out the possibility that some of the items were problematic themselves. Problematic items could account for both a lower number of 'Hits' to the high-frequency sense items and to the higher number of 'False Alarms' to the illogical distractors. For example, if some high-frequency senses had been used in different contexts more participants might have responded that the the meaning was 'Correct'. Alternatively, the participants might have responded that an illogical distractor was 'Correct' if the sentence context led them to a creative interpretation which had not been considered when the item was created. The TPM was created based on the acceptability judgements of three native speakers (see 4.3.5 above). This is a small sample of informants, and even though there was agreement in their judgements, there might be differences if a larger sample of informants was consulted. For this reason, the next chapter will address the semantic acceptability of the test items using a larger sample of native speaker informants. The results to the revised test can then be analysed using the downward adjustment that Meara and Buxton used based on 'False Alarms' to illogical distractors.

5.6 Conclusion

This study investigated a method for measuring polysemous knowledge using the Test of Polysemous Meanings (TPM), the design of which was reported in Chapter 4. The method presented sentences containing target words in different senses. The senses varied in their frequency of use and in their acceptability. Participants were asked to decide if the target words were used correctly or not in the sentences. The method was shown to be effective for two main reasons. First, of the three acceptable senses for the target words, the participants correctly identified the most frequent

senses most often and correctly identified the least frequent senses least often. Secondly, the participants identified the distractors which bore a semantic similarity to the core sense as correct significantly more often than the distractors which bore no semantic similarity. This was taken to indicate that rather than learning each individual sense discretely, the learners extrapolated their current knowledge of known senses to unknown senses. However, the strength of these results can be questioned because when designing the instrument, the items were normed with only 3 native speakers. It is important to norm the items with a larger number of native speakers, and to replace any unreliable items with new ones. Using the revised test items, we can then confirm how robust the results to the TPM are.

Chapter 6: Establishing a native-speaker acceptability judgement

In this chapter, the items in the Test of Polysemous Meanings are reviewed for their acceptability with native speakers of English. In a three-stage revision process, groups of 12 to 13 native speakers judged whether the acceptable items were unanimously acceptable and the distractor items were unanimously unacceptable. Through this process, some items had to be discarded and new items created. Possible reasons for variation among the native speakers is discussed.

6.1 Introduction

The Test of Polysemous Meanings (TPM) was designed to investigate L2 English learners' knowledge of the different senses of common polysemous words. The test included 20 target words with five test items for each target word: three acceptable items and two distractor items. Sentences (1) (2) and (3) present an example of an acceptable item, a logical distractor and an illogical distractor respectively.

(1)   Let's go outside and get some fresh AIR.
(2)   My suitcase was so full I didn't have AIR for anything else.
(3)   Blackbirds lay their AIR in March.

The design of TPM, presented in Chapter 4, used native-speaker judgements to confirm the semantic acceptability for the 60 acceptable items and 40 distractor items. These judgements confirmed that native speakers would find the meaning of target words semantically acceptable in the acceptable items, but semantically unacceptable in the distractor items. The results of the pilot TPM provided results worthy of further investigation; however, one criticism of the design of the pilot was that the sample of native-speaker judges was very small ($n = 3$) and such a small sample is unlikely to be representative of the larger population. It would be worthwhile to establish the acceptability of the items with a larger sample of native speakers. This would lend more confidence to any analysis based on the convergence or divergence of non-native speaker responses to the native-speaker norm.

6.2 Objective of the revision process

The objective of this study was to review the TPM items from the perspective of native speaker judgements, and to revise any items where acceptability was disputed. To develop an optimal set of items, a new group of native speaker judges was recruited to assist with identifying which items might not be ideal representatives of their category as 'acceptable', 'logical distractor' and 'illogical distractor'. An iterative process of revisions was used to optimize the set, as outlined below. In the first stage, 13 native-speaker informants provided judgements on the semantic appropriateness of the target words in all 100 items, and problematic items were identified for revision. In the second stage, a different group of 12 native-speaker informants provided judgements for the revised items, and again problematic items were identified for a second revision. In the final stage, the same group of 12 native-speaker informants provided judgements for the twice-revised items.

6.3 First revision

6.3.1 Native speaker responses

In the first stage 13 native speakers were presented with the 100 items of the pilot-TPM. An item was a complete sentence with the target word capitalised as in the following example: 'She waved her HAND at the crowd'. There were 20 target words in total and each target word was represented by three items intended to be viewed as 'acceptable', and two distractor items, where the target word was intended to be viewed as unacceptable. To ensure that the results were informed by a variety of native speaker backgrounds, five British native-speakers and eight North American native-speakers were asked to participate. This measure was taken especially because the logical distractors tend to encourage creative interpretation. It is difficult to say why a sense might be understandable and yet unacceptable. Having a wider variety of native speaker informants lends confidence that logical distractors break acceptable use of the target word. In the results, however, there

were no items where the informants of one variety answered differently as a group from those of the other variety.

The items were randomized into three different orders using Excel. They were presented to the native-speaker informants through Google's online survey service. The informants were asked to consider the meaning of the target words and to decide whether they were used correctly or not at the sentence-level context of the items. They were given space after each item to write an optional comment on their decision. These were the directions presented to the participants:

> In the sentences below you'll find one WORD in capital letters. Please decide if the meaning of that WORD is used correctly or not. If you find the decision[12] is problematic, please add a comment in the 'Other' text box.

The informants made a decision by clicking either a button labeled "Correct" or one labeled "Incorrect". The responses were noted when they differed from the expected response, i.e. if an acceptable item was identified as "Incorrect" or a distractor item was identified as "Correct". For these unexpected responses, I looked for more information for the informant's decision. First the comment box was consulted and then the informant was emailed. The email presented the item and the response and the informant was asked to comment on his or her decision: for example, "In your survey, you felt that *branch* was used incorrectly in the following sentence: … Can you comment on your decision?" In some cases, the respondents changed their response. For example, one respondent had said that *hand* was used incorrectly in the item "She waved her *hand* to the crowd." In his comment, he explained that *hand* wasn't used incorrectly, but that he wouldn't use it in the sentence himself because the meaning was evident without the word. As the sentences were originally sourced from the Longman Dictionary of Contemporary English (LDOCE) (2010) and the Cambridge Advanced Learner's Dictionary (2011), in cases such as this,

---

12. "Decision" referred to the informant's decision of acceptability, not the researcher's decision when constructing the item. This ambiguity didn't seem to generate any confusion.

where the respondent reversed his decision, the item was deemed acceptable. For other items, the respondents confirmed their initial decision. For example, one respondent had said that the distractor item, "I'll tell you when to get off the face," was acceptable. In her comment, she explained that she had interpreted the sentence as "You have to get off the mountain face." Thus, this item was confirmed as problematic and needed to be replaced by a new item.

The following items were judged differently from expected; the number of dissenters is presented in parentheses:

Illogical Distractors

- I'll tell you when to get off the *face*. (1 dissenter)

Logical Distractors

- I can't buy a new car because my bank *position* is too low. (1 dissenter)
- A new car is in the price *order* between 30 000 and 150 000 riyals. (2 dissenters)
- Carl began in the fashion *branch* by running a clothing shop. (3 dissenters)
- The sales woman put the money into the cash *case*. (4 dissenters)

Acceptable Items

- Toyota-America is the American marketing *arm* of a Japanese company. (3 dissenters)
- We've got to take *heart* from the fact that we played well. (1 dissenter)
- There were six houses arranged in the *form* of a square. (3 dissenters)
- He has a happy but quiet *character*. (3 dissenters)

6.3.2 Revision of the distractor Items

In five cases, as listed above, the native-speaker informants confirmed that they judged the meaning of the target words in the distractor items as acceptable. New items were created for these items. The following discussion explains the revision process by item.

Four logical distractors were revised. One native speaker informant felt that *position* was used acceptably in the sentence, "I can't buy a new car because my bank position is too low." He felt that "one's position at the bank in this context would refer to a financial position, i.e. how much money held," or "it could also refer to one's status at a bank in the context of being overdrawn." In this case, the informant clearly thought that the sense was acceptable and so a new item was constructed. The logical distractors were created in the same procedure described in Section 4.3.3 by identifying 'false friends' to the target in languages related to English: French, German, Spanish, and Italian. The procedure for creating the logical distractors can be explained with the target word 'cover'. First a direct equivalent of the English target word is looked up in the bilingual dictionary. In French *cover* translates as *coverture*. Then the meanings of the translation equivalent are surveyed to identify a case where that word is translated into a different word in English and couldn't be translated using the English target word. The word *coverture* can mean ROOF in English. From there the different English translation word is looked up in a learner's dictionary to find a suitable example sentence. For example, the LDOCE has the sentence "The house was old and its roof needed to be repaired." The target word is substituted into the example sentence, as in "The house was old and its cover needed to be repaired."

In consultation with foreign language dictionaries, it was was identified that *position* is translated with the German word *Platz*. *Platz* is also used in German to describe contexts where in English the world *place* would be used but *position* could not: "das beste Hotel am Platz = the best hotel in the place," (Oxford-Duden German Dictionary, 2011). The item constructed for trial was "It was the best hotel in the position."

The second logical distractor to be revised was, "A new car is in the price *order* between 30 000 and 150 000 riyals." Two informants found this use of *order* to be acceptable. Again a new item for the illogical distractor was constructed. *Order* is translated into Italian as *ordine*. In Italian, *ordine* can be used to refer to CLASS OF TRAVEL, "di prim'ordine = first-class, first-rate, high-class," (Oxford-Paravia Italian

Dictionary, 2011). The sentence "We prefer to travel in first class," was taken from the LDOCE (2010) and used to create the revised item, "We prefer to travel in first order."

The third logical distractor to be revised was, "Carl began in the fashion branch by running a clothing shop." Three informants felt that this item was acceptable. One of them explained that the use of *branch* sounded like industry jargon. To construct the new illogical distractor, *branch* was translated as *branche* in French, where it could be used to describe the arms of a pair of GLASSES, or the blades of SCISSORS, (Oxford-Hachette French Dictionary, 2011). A use of *arm* couldn't be found in reference to GLASSES in the learner dictionaries consulted, nor *blade* in reference to SCISSORS. As alternatives, authentic uses of these senses were found on the internet: "I broke the arm of my glasses!"(Yahoo Answers!, n.d.), was revised as, "I broke the branch of my glasses!"; and "Scissors have blades less than 15 cm long," (eNotes Study smarter, 2012), was revised as "Scissors have branches less than 15 cm long." In this case, two distractor items were created. It is often difficult to find an appropriate 'false friend' for the target words in general and since these two looked comparable, I decided to solicit native-speaker judgements on both and then later discard one in favour of the other.

The final logical distractor to be revised was, "The sales woman put the money into the cash case." Four respondents felt that this item was acceptable, indicating that it was clearly problematic. The item was revised with a 'false friend' from Italian. In Italian *case* translates as *caso*, but *per caso* can mean *by chance* in English (Oxford-Paravia Italian Dictionary, 2011). The sentence, "I bumped into her quite by chance in Oxford Street," was taken from the LDOCE (2010) and used to create the revised item, "I bumped into her quite by case in Oxford Street."

One illogical distractor was found to be acceptable to one informant, "I'll tell you when to get off the face." As mentioned above, the informant felt that the item could be used in the context of mountain climbing, and so the item was revised. The objective of the illogical distractor was to have no semantic relationship between the

150

target word and the context (Read, 1993). The sentence, "You must maintain a minimum balance of $1,000 in your bank account," was taken from the Merriam-Webster Learners' Dictionary (Merriam-Webster's Advanced Learner's English Dictionary, 2008). It was felt that FACE and BANK BALANCE did not express a semantic relationship, nor FACE and MINIMUM. The sentence was used to create the item, "You must maintain a minimum face of $1,000 in your bank account."

6.3.3 Revision of the acceptable items

The low frequency items for *arm* and *heart* were found to be problematic. The sense of *arm* in "Toyota-America is the American marketing arm of a Japanese company," was unacceptable to two informants who felt that marketing wasn't a separate enough division to warrant the use of *arm*. The sense of *arm* was preserved and only the context was changed in the revised item, "She works in the research arm of the company," (Oxford Advanced Learner's Dictionary, 2011). One informant felt that *heart* sounded awkward in the the sentence, "We've got to take heart from the fact that we played well." It might have been that the general sense was unacceptable to the participant. For this reason, the item was changed with the new sense, "I have a house in the heart of London," (Longman Dictionary of Contemporary English, 2010). According to the LDOCE, the new sense of *heart* was less frequent than the sense used for middle frequency sense .

One middle frequency item was problematic. Two informants felt that *form* was used incorrectly in the sentence, "There were six houses arranged in the form of a square." One felt that "shape" was better suited to the configuration of objects, another felt that *form* should describe a more "organic, natural shape," while the third preferred a different morphosyntactic formation, "they formed a square". Because two of these opinions focussed on details of the context, two items were constructed for trial: "The glass bottle was made in the form of a fish," which was taken from a phrase found on the Internet (British Museum, n.d.) and "The stadium was in the form of a circle," (Cambridge Advanced Learner's Dictionary, 2011). The

circle sense might have been better than "form of square" because circles contain no sharp angles, and are perhaps 'more organic'.

The high frequency use of *character* was problematic for three people. Two alternative items were created for trial. The first one maintained the sense of "all somebody's qualities," "It's not in her character to be jealous," (Cambridge Advanced Learner's Dictionary, 2011). Two of the informants, however, felt that the item, "He has a happy but quiet character," was problematic because the verb *have* was used. They suggested that it was more appropriate to use the verb 'be' with *character*. For example, one informant said, "If using character, one would say 'He was an interesting character', I don't think you 'have' a character." This criticism questions the entry in the LDOCE. The problematic item was taken from the first sense listed in the dictionary, i.e. the most frequent sense, which was listed as ALL SOMEBODY'S QUALITIES. The informants were indicating a preference for the second listed sense, that of PERSON. For this reason, a second item was trialled using the second most frequent sense in the LDOCE, with the example sentence of, "Linda was something of a character,".

The revised items were analysed for the frequency of their constituent words using the lexical profiler tool on the Lextutor website (Cobb, n.d.). The following words from the revised items were identified as outside the 2000 most frequent words according to the General Service List (West, 1953): *stadium, London, research, bumped, Oxford, Linda,* and *jealous*. The proper nouns, *Oxford Street* and *Linda*, were preserved, but *London* was removed to create the item, "I have a house in the heart of the city," which was felt to be clearer. Several items were preserved because they were considered to be known to the participants for the following reasons: *stadium* and *flu* have been borrowed into Arabic, the participants' L1; *research* and *jealous* are known to have been a focus of the curriculum through the students' course book. The word *bumped* was used in the idiom, "bumped into someone', which might have been too obscure for some of the participants. The item in question was revised as, "I met her quite by case in Oxford Street."

6.4 Second revision

6.4.1 Native speaker responses

Twelve new native speaker informants were recruited to assess the revised items[13]. In the first trial British English speakers had been included, but there had been no difference between the judgements of the North American English speakers compared to the British English speakers. For this reason, the lack of variety among the informants wasn't considered a strong limitation. In this trial, all of the informants were North American.

This time the items were presented on paper instead of through online software. The online software had resulted in access problems for some informants in the previous trial. Again, the items were randomised across three different versions to ensure against an order effect. Two lines were placed below each item preceded by the word "Comment?" Comments were also solicited with the following sentence in the directions, "If you find the decision is problematic, please add a comment."

Four of the revised items were found to be problematic because one or more of the native speaker informants responded against expectation. One informant found the high frequency item, "Linda was something of a character," to be problematic, but the alternative high-frequency item, "It's not in her character to be jealous," was accepted by all of the informants and so it was selected for the final instrument. The two middle-frequency items for *form* were both problematic: "The glass bottle was made in the form of a fish," was unacceptable for three participants, and "The stadium was in the form of a circle," was unacceptable to six participants. Finally, "She works in the research arm of the company," was unacceptable to four of the informants. The rest of the acceptable items were acceptable to the informants and all of the distractors were unacceptable, as expected. Two items had been trialled for the logical distractor for *branch*. The item "I broke the branch of my glasses!" was

---

13. This group of native speakers was also used in the third revision presented below. The assumption was that these twelve and the first thirteen informants were sufficiently representative for it not to matter whether the same people were consulted more than once.

chosen for its brevity over the sentence "Scissors have branches less than 15 cm long."

6.4.2 Item revisions

The middle-frequency sense of *form* to mean SHAPE was problematic in each of the previous two trials. For this reason, a new sense was tested. According to the LDOCE, *form* to mean SHAPE was the third most frequent sense of the word, while the second most frequent sense of *form* was THE WAY SOMETHING IS/APPEARS. The example sentences provided by LDOCE used several low frequency words and so other dictionaries were consulted for example sentences. Three sentences were selected to be reviewed: "The disease can take several different forms," (Oxford Advanced Learner's Dictionary, 2011); "Help in the form of money will be very welcome," (Oxford Advanced Learner's Dictionary, 2011); and "The medicine comes in liquid form," (Cambridge Advanced Learner's Dictionary, 2011).

 The low-frequency sense of *arm* to mean PART OF A GROUP was also problematic in each of the previous two trials. This was the fifth-listed sense in LDOCE and it was replaced with the third-listed sense of *arm*, meaning 'the part of a chair, sofa etc that you rest your arms on'. No example sentence was offered for this sense in the LDOCE and so the following example was found on the internet, "Please don't sit on the arm of the chair," (Mumsnet Talk, n.d.).

One final item was revised. The low frequency sense of *case* was "Polly carried her cases upstairs to the bedroom." There was a potential that another item could influence the learners' decision because it included the word *case* in *suitcase*: "My suitcase was so full I didn't have air for anything else." This item was substituted with one which was connected to the most frequent sense and of similar frequency, "He has a bad case of the flu," (Merriam-Webster's Advanced Learner's English Dictionary, 2008). A frequency analysis of these items found that only *flu* was outside of the first 2000 most frequent words in English (Cobb, n.d.). As stated above, *flu* is borrowed into Arabic, the learners' L1.

6.5 Third revision

The revised items were presented on paper to the same 12 native-speaker participants who were asked to judge the meaning of the acceptable items according to the methodology used above. Again, all were native speakers of North American English.

The low-frequency senses of *arm* and *case* were acceptable to all 12 informants and were used in the final instrument. Three items were reviewed for the middle-frequency sense of *form*. The item, "The disease can take several different forms," was problematic for one participant; however, the use of *form* in the two other senses was acceptable to all 12 informants. The sentence, "The medicine comes in liquid form," was chosen over "Help in the form of money will be very welcome," because the syntax was simpler.

6.6 Discussion

The native speakers' acceptability judgements of the items were not straightforward. There were quite a number of cases where native speakers made judgements against expectation. Initially 24 of the 100 items received an unexpected response and in the end nine of the items were revised. The participants often reversed their initial decision when they were asked to comment. At times they said their answer was just a mistake. At other times, even though they still reversed their decision, their comments explained why they made their initial judgements. Below, I will discuss certain judgements where the comments of the native speakers provide insight into the nature of the task.

One native speaker initially accepted the following logical distractor, "The centre part of an apple is called its heart." In his comment he said, "Looks like I slipped up there! Should be 'core' of course, but 'heart' sounded reasonable at the time!" Another speaker initially accepted the following illogical distractor, "He turned on the line to listen to music." In her comment she said, I guess I was visualising working with a soundboard with a band, there are different lines connecting to amps. Not really "correct" native-speaker usage, I agree, but for me personally, I

would understand the context." I believe this illustrates how there is a strong tendency to construe meaning of the whole sentence and create a context above the meaning of the individual words.

The native speaker comments raise two points for discussion. The first concerns subtle semantic distinctions and the second concerns the importance of syntax and collocation. For the native speakers, the distinction between what was semantically acceptable and unacceptable illustrates the fuzzy boundaries of the target words' semantic concepts (Rosch, 1978). One informant commented, "To clarify, it's not that I felt the words were used correctly so much as not incorrectly." The context of the word was important and even minor changes could throw off the meaning. One informant initially said *class* was used incorrectly in, "English classes start at 5:15." "I'm not sure now why I put this as 'wrong'. Maybe it is the time (5:15) and it makes no sense that classes would start that early in the morning." The informant's use of a word can also be quite nuanced, as in relation to this item, "This branch of the river eventually empties into the Atlantic Ocean." The informant said, "I guess I was thinking that when I use this reference, I prefer to use 'arm' instead, but 'branch' would not be incorrect." In cases such as these, there is possibly an effect from making many quick semantic judgements one after another.

The second point of discussion concerns the comments that deal with the influence of syntax and collocation on the meaning of the target words. As mentioned earlier, one informant felt that the target word *hand* was redundant in the sentence, "She waved her hand to the crowd." At times the target word was said to be incorrect because of its part of speech. One informant initially had a problem with the item, "There were six houses arranged in the form of a square." She said that she would say, "they formed a square." As discussed above, there is also a problem of collocation, which led to the following item being rejected, "He has a happy but quiet character." One participant felt that *character* collocated better with 'be' than with 'have': "If using 'character', one would say 'he was an interesting character'. I

don't think you 'have' a character." Another said that he would accept the collocation "'quiet character' more readily than 'he has a happy character'."

The objective of the native-speaker revision was to create robust items that reflected native-speaker judgements about acceptable and unacceptable uses of the twenty target words. As this discussion of the native-speaker comments shows, however, factors such as collocation, part of speech, personal usage, and sentence construal can greatly affect the informant's judgement of semantic acceptability. Furthermore, the number of items was a possible influence on the informants' judgements, because with so many items, they tended to make decisions quickly. However, when they were asked to comment on their decisions, they probably spent more time because they were only asked about a few items ($max = 5$). Also, the first survey was presented online through survey software while the second and final surveys were on paper. There could be implications where certain informants prefer one format over the other. While acknowledging that these factors should be considered when interpreting the learners' results of the test, it was considered that after this process of item revision, the TPM was robust enough to use in further analysis of learner data.

Chapter 7: Comparison of the two TPM presentations: (Grouped and Mixed)

In what follows, the revised Test of Polysemous Meanings is given to two groups of English learners in two different presentations. For one group, the presentation was the same as in the pilot study, where 10 items were presented on a page with no target word being repeated among any of the 10 items. The second group received a different presentation where the five items sharing the same target word were presented together on the same page. The reason for including the grouped presentation was to see if the participants benefitted from the opportunity to compare one use of the target word with other uses. In addition to the comparison of presentations, the results for the acceptable items were compared for the effect of intra-word frequency and the results of the distractor items were compared for the effect of semantic similarity. In the analysis of the results for the acceptable items, a new calculation was introduced. This calculation reduced the total number 'Hits' according to the number of 'False Alarms' a participant made in judgements for the illogical distractors. The rationale for this calculation and the subsequent analysis of the results are discussed.

7.1 Introduction

In the previous chapter I described how the Test of Polysemous Meanings (TPM) was revised from the original design which had been used in the pilot study. In the revision, the acceptability judgements were solicited from a larger sample of native speakers to revise nine of the 100 items. In this chapter, the revised version of the TPM will be tested with a new sample of participants. The results of the revised version can be compared to the results of the pilot study to see if the findings are replicated. The test of the revised version of the TPM will expand on the method of investigation used in the pilot. The pilot study was designed to limit the participants' opportunity to compare the target word across items. In the pilot study, the TPM items were presented in a mixed order like this:

(1)   Then they call out our names in ORDER and we answer yes or no.
            He looked at the BODY to check the time.
            Be careful with that needle - it has a very sharp POINT.

Our staff combine efficient service with a personal HAND.

Let's go outside and get some fresh AIR.

However, it is not known exactly how the participants would be affected if the items were presented when grouped by target word. In the grouped presentation, the items are organised like this:

(2)   The house was old and its COVER needed to be repaired.

        The cloud COVER in the morning should clear later.

        There is a plastic COVER over the meal.

        The COVERS had slipped off the bed in the night.

        The people have a sense of COVER in their neighbourhood.

With the opportunity to compare one item to another, would the participants note the similarity between the senses? There is some suggestion from the literature that this might be the case. Verspoor and Lowie (2003) found that the learners in their study benefitted from the opportunity to compare the sentence-contexts using the same polysemous word to guess at the meaning of an unknown polysemous sense. In the psycholinguistic research, Klepousniotou (2002) and Klepousniotou and Baum (2007) found that there was an advantage for polysemous words over homonyms in semantic priming tasks. The argument here is that polysemes share in the same mental representation, so the construal of the prime sense can facilitate the construal of the target sense. Based on the above research, the hypothesis is that when items are grouped by target word, the participants will respond with more 'Hits' to acceptable items and more 'False Alarms' to logical distractors on the TPM because the construal of one sense can facilitate the construal of another similar sense. The following research questions are proposed for this chapter:

RQ1   *Do the learners respond with more 'Hits' to acceptable items and more 'False Alarms' to logical distractors when the test items are grouped by target word?*

RQ2   *Do the learners respond to the frequent senses with more 'Hits' than to the infrequent senses?*

RQ3   *Do the learners respond to the logical distractors with more 'False Alarms'*
               than to the illogical distractors?

It is important to consider whether grouping the items by target word poses a
reliability problem. There are three acceptable items and two distractor items for
each target word. If the participants figured out this system, it could influence how
they respond. They may come to see that there is a pattern of three acceptable and
two distractor items. While the items can be randomised so that the acceptable and
distractor items appear in no regular order, it remains a question whether the
participants are influenced by this aspect of the design. To establish whether there is
an inherent design problem with grouped presentations the following question will
be considered first: Do the learners respond with a pattern of three 'Correct' and two
'Incorrect' responses more often in the grouped presentation than the mixed
presentation?

7.2 Study

Two presentations of the TPM were created for this study: a grouped presentation
and a mixed presentation. The TPM tested knowledge of 20 target words across 100
items, with five items for each target word. The grouped presentation presented the
five items per target word together on the page. The order of the five items for each
target word was randomised, as in example (2) above on page 160. There were three
versions of the grouped presentation, whereby the order of the target words
themselves was randomised across each version of the test to control for an order
effect[14].

In the mixed presentation, the order of the 100 items was arranged so that no item
sharing the same target word was repeated on one page, and then each page set was
randomised. Three versions of the mixed presentation were also created to control

---

14. A version of the grouped presentation is presented in Appendix 3.1 on page 343.

for an order effect[15]. The participants also wrote a test of L2 vocabulary size along with the TPM. This test is discussed in Chapter 8.

7.2.1 Participants

There were two data collection sessions, one for the mixed presentation and one for the grouped presentation, with different groups of participants for each session. Each participant saw the stimuli only once. In both sessions the data was collected from students enrolled in the final course of an intensive English programme at a technical college in Qatar. After completing this course the students were expected to achieve a 5.0 on the IELTS proficiency test in order to enter their English-medium, college-level programmes in business or applied technology. This score placed them at the proficiency level B1 (Independent User) in the Common European Framework of Reference. According to a questionnaire, the students were all L1 speakers of Arabic and none of them spoke a third language after their English L2.

Thirty-eight students completed the grouped presentation of the TPM ($n = 38$). These 38 participants wrote the TPM in a classroom during their lunch hour and they received a remuneration of QAR 150 (≈£25) for their participation. Their results were used in this and subsequent chapters. Data was also collected from 29 participants using the mixed presentation of the TPM ($n = 29$). These 29 participants wrote the TPM during their English classes, but they did not receive remuneration. Their results were also used in this and subsequent chapters.

7.2.2 Procedure

As a researcher, I explained to the participants that I was looking at L2 knowledge of different meanings of the same word and I presented example items using the example word *break* to illustrate the task. The front page of the TPM has directions and example items using the example word *break*. (See Appendix 3.1 for an example of the test). These directions were read aloud to the students. The directions explained that they were permitted to guess, but if they felt they couldn't guess, they

---

15. A version of the mixed presentation is presented in Appendix 3.2 on page 354.

could choose the option of 'Don't Know'. The participants took 20-35 minutes to write the test.

7.2.3 Reliability of the grouped presentation

There was a concern that the participants might recognise a pattern in the grouped presentation of three correct and two incorrect items. To establish whether this was the case, the frequency of response types (Correct, Incorrect and Don't Know) was compared between the two presentations (Grouped and Mixed).
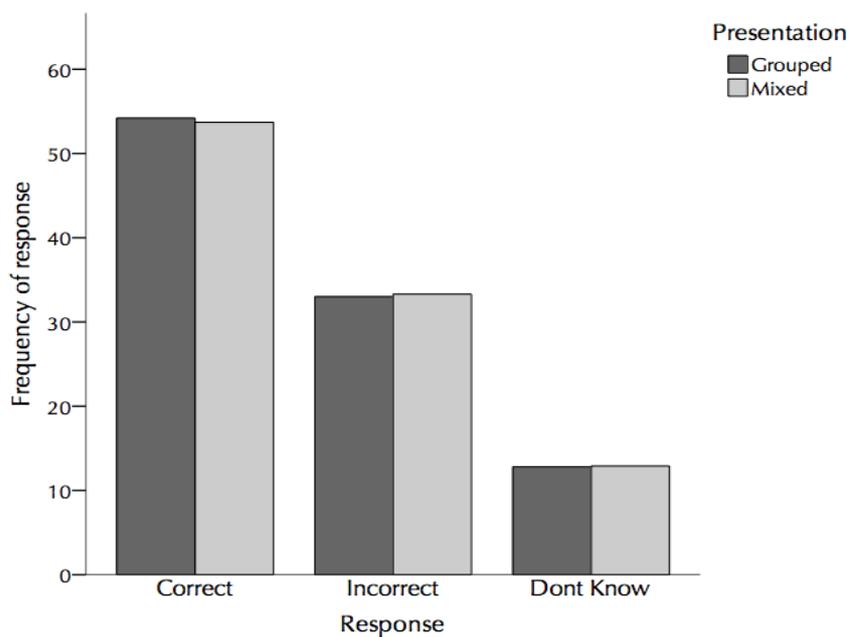
Figure 7.1
Frequency of response-type for grouped presentation of the TPM compared to the mixed presentation

Since the number of participants in the grouped presentation ($n = 38$) were not equal with the number in the mixed presentation ($n = 29$), the results are presented as proportions. The proportions of the response types for the grouped presentation were 54.2% Correct, 33.0% Incorrect, and 12.8% Don't Know; the proportions for the mixed presentation were 53.7% Correct, 33.3% Incorrect and 12.9% Don't Know. The frequencies are presented in Figure 7.1 as percentages.

The response rate was nearly identical across the two presentations, and as one would expect there was no significant relationship between response-type and presentation: Pearson χ^2 (2, $N = 6700$) = .125, $p = .94$.

While the above analysis shows that there was no significant difference in the frequency of response rate across the tests as a whole, there could be a concern that participants in the grouped condition guessed the pattern of three acceptable items and two distractor items across the five items by target word. In order to address this concern the following calculation and analysis were conducted. First, the pattern of three 'Correct' and two 'Incorrect' response-types for each target word was counted by participant for both the grouped and mixed presentations. This calculation gave the frequency of individual participants responding to the set pattern by presentation condition.

The distributions of these frequencies were analysed for normality to determine whether parametric or non-parametric tests should be used. The results to a Shapiro-Wilk test of normality found that the distribution of the grouped condition was significantly different from normal, $W(51) = .944$, $p = 0.02$. The means of the two frequencies were compared using a non-parametric test.

A one-tailed Mann-Whitney test indicated that the grouped condition ($Mdn = 5$) was not biased towards the pattern of three 'Correct' and two 'Incorrect' response-types, when compared with the mixed condition ($Mdn = 4$), $U = 629.5$, $p = .13$, $r = .12$.

Since it is unreasonable to propose that the participants could have guessed at the frequency of responses in the mixed presentation, there is little reason to conclude that the participants in the grouped presentation would have guessed the design of three acceptable and two distractor items for each target word.

7.2.4 Scoring

As discussed in Chapter 5 above, the results of the Test of Polysemous Meaning were scored according to a system used in detection theory (Shillaw, 2009). If an acceptable item was answered as 'Correct' it was marked as a 'Hit', but if it was

answered as 'Incorrect' then it was marked as a 'Miss'. If a distractor was answered as 'Correct' it was marked as a 'False Alarm', but if a distractor was answered as 'Incorrect' it was marked as a 'Correct Rejection'.

There is always the risk in an experiment of this sort of design that participants will misread an item and perhaps see a creative link that was not controlled for. The illogical distractors were included as a means of gauging the extent of this risk. Although each item carries its own individual risk level for each participant, one can nevertheless gauge a participant's propensity to misread items on the basis of how they respond to the illogical distractors. The premise is that there is no other reason to respond to an illogical distractor as 'correct' than on the basis of misreading or uncontrolled creative interpretation. The tendency to respond this way can be used to calculate the likelihood that other items judged 'Correct' were also caused this way, rather than on the bases under examination. In the pilot study of the TPM (Section 5.5), I discussed how the following formula, taken from Meara and Buxton (1987), could be used in order to accommodate for this chance of error. The formula was used to calculate the likelihood that a participant can be deemed to know that an acceptable item is correct.

$$P(k) = \frac{P(h) - P(fa)}{1 - P(fa)}$$

Meara and Buxton used this formula in their investigation of Yes/No tests which incorporate both real and imaginary words. In scoring the TPM, *P(h)* (i.e. the probability of a 'Hit') is the proportion of acceptable items the participant recognises; *P(fa)* (i.e. the probability of a 'False Alarm') is the proportion of illogical distractors that the participant identifies as 'Correct'. The formula adjusts the 'Hits' score downwards if the 'False Alarms' score for the illogical distractors is large. *P(k)* indicates how many senses represented in the acceptable items the participants can claim to know. The formula is applied by participant to the results for each of the acceptable-item categories: high, middle and low-frequency senses. There were 20

items for each item category (*max* = 20) as well as for the illogical-distractor category (*max* = 20).

The results for logical distractors were not adjusted downwards. These results were compared to the number of 'False Alarms' for the illogical distractors, unlike the acceptable items, which are compared to one another. Since the results to the logical distractors were already being compared to the results of the illogical distractors, to adjust them downwards would be to factor the effect of the illogical distractors twice.

7.3 Results for the two different presentations

Two presentations of the TPM were tested. In the mixed presentation no two items used the same target word on any one page of the test. In the grouped presentation all five items that tested the same target word were presented together; therefore, in the grouped presentation the learners had the opportunity to compare how the same target word was used in different contexts. A sample of the presentations were presented in the Introduction (Section 7.1) at (1) and (2).

Unlike in the pilot of the TPM, presented in Chapter 5, there were no cases where any participant responded with too many 'Don't Know' responses. However, the results of four participants were removed because of a high number of 'False Alarms' to illogical distractor items. The rationale for the removal of these results is presented below in Section 7.3.3, where the downward adjustment of the results is discussed. With these four participants removed from the analysis, there were a total of 37 participants who wrote the grouped presentation and 26 participants who wrote the mixed presentation of the TPM.

7.3.1 Results for the acceptable items

The results for the acceptable items on the TPM are presented in Tables 7.1 and 7.2. Table 7.1 presents the results for the 37 participants who wrote the grouped presentation and Table 7.2 for the 26 participants who wrote the mixed presentation. The results show the mean number of responses for the high, middle and low-

frequency item-types across the 20 target words. The standard deviations are presented in brackets. The same information is illustrated in Figure 7.2.

Table 7.1
Results of the acceptable items on the grouped presentation of the TPM by sense frequency (n = 37)

Item type	Response			
	Hit	Miss	Don't Know	Total
High	16.59 (2.40)	2.05 (1.88)	1.35 (1.38)	20
Middle	13.43 (2.77)	4.16 (2.41)	2.41 (2.45)	20
Low	11.78 (3.66)	5.46 (3.74)	2.76 (2.71)	20

Table 7.2
Results of the acceptable items on the mixed presentation of the TPM by sense frequency (n = 26)

Item type	Response			
	Hit	Miss	Don't Know	Total
High	16.58 (2.40)	2.12 (2.01)	1.31 (1.23)	20
Middle	12.73 (3.01)	4.15 (1.97)	3.12 (2.18)	20
Low	11.31 (3.37)	5.65 (3.10)	3.04 (2.11)	20
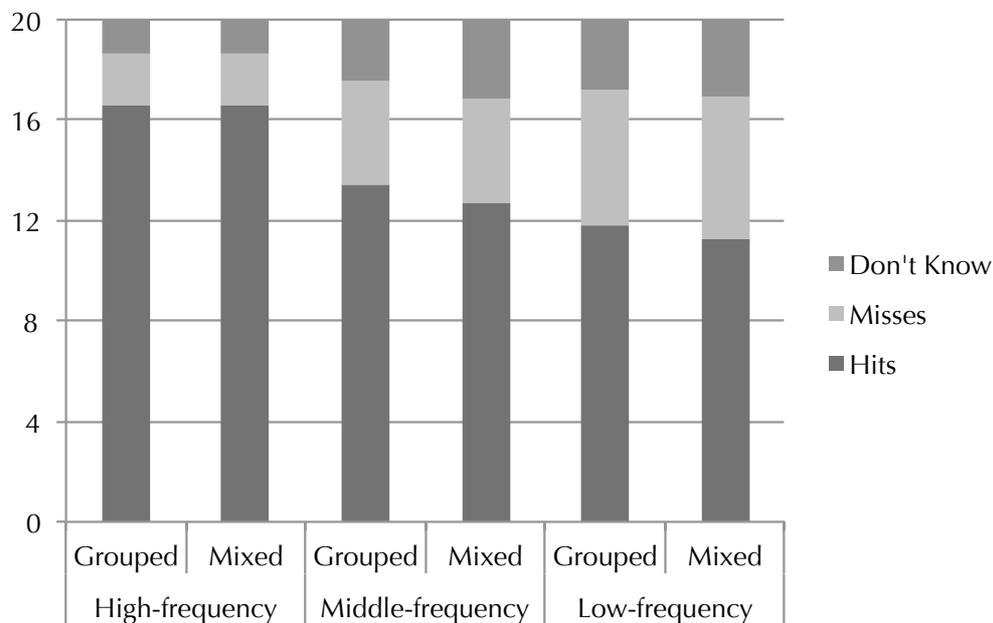
Figure 7.2
Mean number of responses for acceptable item-types for the grouped and mixed presentations

167

The frequency of response for the acceptable items was largely comparable between the two presentations. There was a slight advantage to the grouped presentation for the middle and low-frequency items; whether this advantage proved significant or not is analysed in Section 7.4.1 below. However, the results do appear to corroborate the hypothesis of frequency. The greatest number of 'Hits' in both presentations was for the items where the target word was used in its high-frequency sense and the fewest number of 'Hits' where the target word was used in its low-frequency sense[16].

7.3.2 Results for the distractor items

The results for the distractor items on the TPM are presented in Tables 7.3 and 7.4. As with the previous analysis, the results show the mean number of responses for the logical and illogical distractors across the 20 target words. The standard deviations are presented in brackets. The same information is illustrated in Figure 7.3.

Table 7.3
Results of the distractor items on the grouped presentation of the TPM by sense frequency (n = 37)

		Response		
Item type	False Alarms	Correct Rejections	Don't Know	Total
Logical	7.46 (3.73)	9.30 (4.14)	3.24 (2.81)	20
Illogical	4.73 (2.81)	12.11 (3.89)	3.16 (2.99)	20

Table 7.4
Results of the distractor items on the mixed presentation of the TPM by sense frequency (n = 26)

		Response		
Item type	False Alarms	Correct Rejections	Don't Know	Total
Logical	7.54 (3.58)	9.62 (2.94)	2.85 (2.29)	20
Illogical	5.00 (3.36)	12.15 (3.78)	2.85 (2.60)	20

---

16. In Section 7.4.2 I consider the possibility that there is no significant difference between the results to the middle and low-frequency items.

The results to the distractor items show no benefit to the grouped presentation over the mixed presentation[17]. This is congruent with the results to acceptable items presented in Figure 7.3. As predicted there were a larger number of 'False Alarms' for the logical-distractor items than there were for the illogical-distractor items. This indicates that as a whole the participants saw the logical distractors as more acceptable than the illogical distractors.
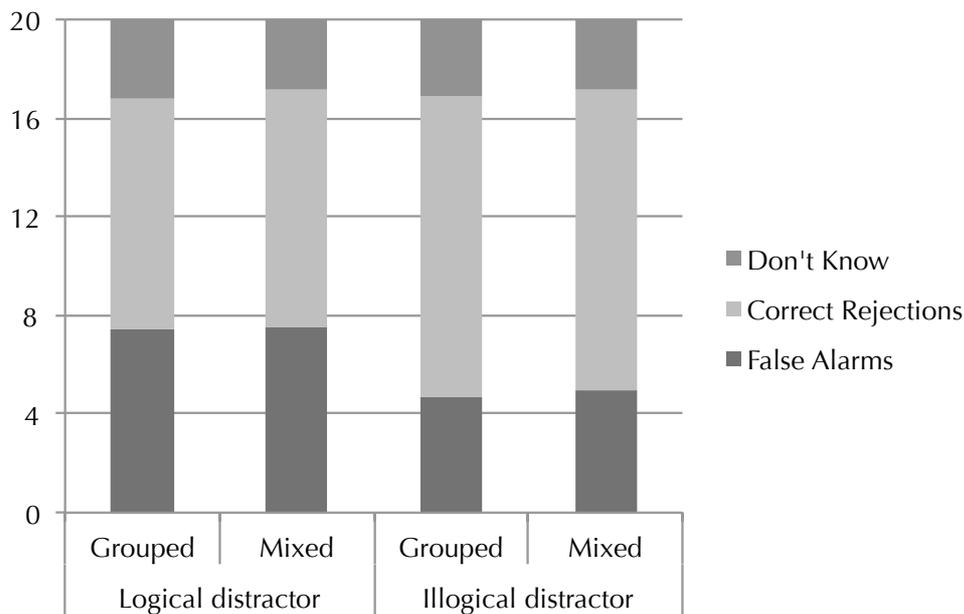
Figure 7.3
Frequency of response for distractor item-types for the grouped and mixed presentations

7.3.3 Adjustment of the results

As discussed in Section 7.2.4, the results for the acceptable items by participant were adjusted downwards according to the formula taken from Meara and Buxton (1987). The formula uses the number of times a participant responded with 'False Alarms' to the illogical distractors in order to adjust their number of 'Hits' to the high, middle and low-frequency items. In four cases, the participants responded with more 'False Alarms' to the illogical distractors than 'Hits' to one of the three acceptable item-types. For example, one participant had 14 'False Alarms' to the illogical distractors

---

17. The results are tested for statistical significance in the analysis Section 7.4.3 below.

but only only 10 'Hits' to the middle-frequency items. When the formula is calculated for these cases, the result is a negative number. It is worthwhile considering what the implications of this finding are before moving forward with the results.

Meara and Buxton (1987) originally used this formula on the Y/N test. This test asks test-takers whether they know the meaning of a specific word form by responding with yes or no. In their test the distractors are non-words: they follow the spelling conventions of the target language but do not have any meaning. The distractors in their test act as a check that the test-taker is being honest in his responses. Because it is unlikely that a test-taker would mistake non-words for real words more often than actual real words, the results of a test-taker with a high number of 'False Alarms' can be discounted as unreliable.

The task on the TPM is different from the Y/N test. A response that an illogical distractor in the TPM is correct is a different decision than that a non-word is correct in the Y/N test. Unlike distractors on the Y/N test, the illogical distractors contain only real words and are syntactically correct sentences. There are a number of reasons why four of the participants should have scored a higher number of 'False Alarms' to the illogical distractors than 'Hits' to one of the acceptable-item categories. First of all, the participants might have misunderstood the directions of the task and focussed on something other than the meaning and use of the target word. On the other hand, they may have followed the directions correctly and still have found the distractor to be correct. They may have misread the surrounding context and construed a meaningful sentence, or they might have been quite creative and applied an interpretation that hadn't been considered when the instrument was designed. This issue of creativity brings up a necessary limitation to the design of the test. The TPM requires that the participant have a sense of when a word is incorrect. The test could not be validly used with a participant who could creatively make sense of the target word in any context.

In sum, the illogical distractors have been designed to bear no semantic similarity to conventional uses of the target word and these items have been normed for unacceptability to native speakers. A high number of 'False Alarms' to the illogical distractors relative to 'Hits' indicates either a misunderstanding of the task or a strategy of creative interpretation beyond the design limitations of the test itself. Based on these considerations, the results of these four participants were removed from further analysis because their responses cannot be reliably attributed to their semantic understanding of the target words.

Table 7.5
A comparison of the unadjusted and adjusted mean number 'Hits' to the acceptable items by presentation

Item type	Grouped ($n = 37$)		Mixed ($n = 26$)	
	Unadjusted	Adjusted	Unadjusted	Adjusted
High-frequency items	16.59 (2.40)	15.40 (3.31)	16.58 (2.40)	14.85 (4.30)
Middle-frequency items	13.43 (2.77)	11.24 (3.64)	12.73 (3.01)	9.96 (4.45)
Low-frequency items	11.78 (3.66)	9.32 (4.39)	11.31 (3.37)	8.42 (3.72)

Table 7.5 presents the mean number of 'Hits' for both presentations when adjusted for the amount of error indicated by the 'False Alarms' for the illogical distractors. There were 20 items for each of the item types. The adjustment was made for each participant separately. The table shows the mean results by item type and presentation. Standard deviations are shown in parentheses. The downward adjustment between the two presentations is comparable, as can be seen more directly in Figure 7.4, where the unadjusted scores are presented on the left and the adjusted scores are presented on the right.
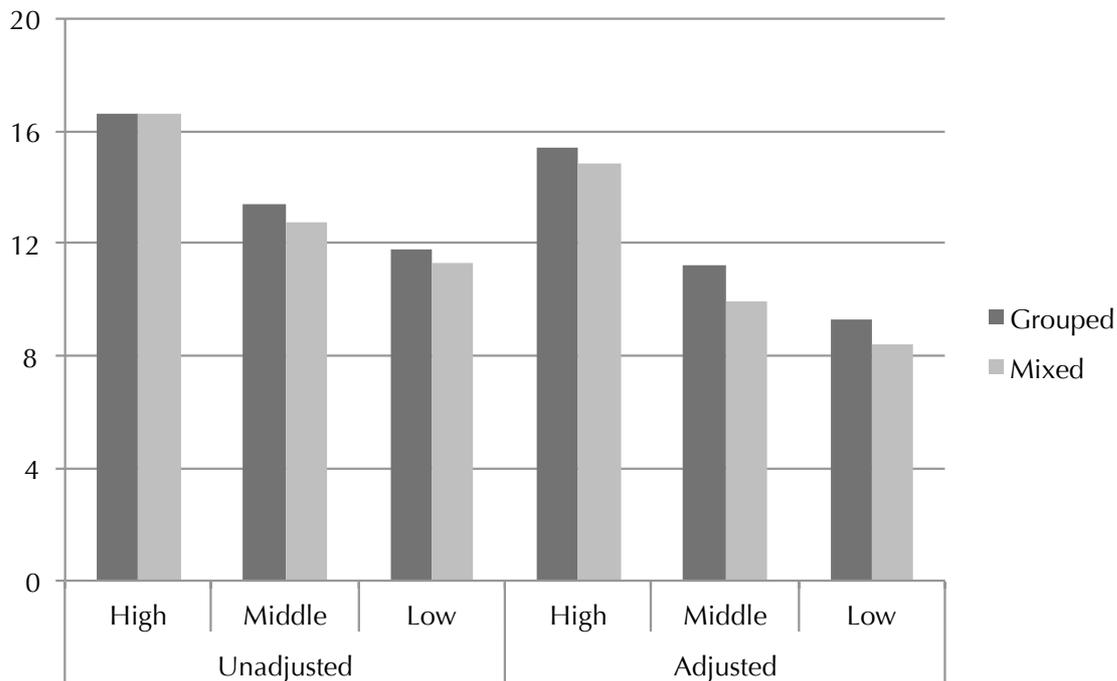
Figure 7.4

Comparison of mean number of 'Hits', both unadjusted and adjusted, across the two presentations

7.4 Analysis

7.4.1 Comparison of TPM presentations

RQ1 Do the learners respond with more 'Hits' or 'False Alarms' when the test items are grouped by target word?

It was hypothesised that the learners in the grouped version of the TPM would respond with more 'Hits' to the acceptable items because they would have the opportunity to guess that unfamiliar senses were correct based on their similarity to familiar senses. Likewise, it was also hypothesised that the learners in the grouped treatment would respond to the logical distractors with more 'False Alarms'. In this scenario, the participants would again have had the opportunity to note semantic similarity between the logical distractors and the acceptable items.

Data for the acceptable items were analysed using a mixed-design ANOVA. A mixed design was chosen because the analysis contained both a repeated measure

172

and an independent measure. The comparison of the acceptable items was a repeated measure because the same participants responded to each of the acceptable items being compared, making a within-subjects factor of acceptable-items (high, middle and low-frequency senses). The comparison between proficiency groups was an independent measure because the groups comprised of different participants, making a between-subject factor of presentation (grouped and mixed).

Because the analysis includes a repeated measure, Mauchly's test of sphericity was conducted to ensure that there was an equality of variance between the three levels of the acceptable items. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2 (2) = 6.3$, $p < .043$; therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.95$). This correction reduces the degrees of freedom to ensure a valid F-ratio. The main effect of acceptable items was significant $F(1.90, 116.0) = 79.6$, $p < .001$, $\eta p^2 = .57$; however, the main effect of presentation was not significant $F(1, 37.8) = 785.5$, $p = .27$.

Data for the distractor items were also analysed using a mixed-design ANOVA with a within-subjects factor of distractor items (logical and illogical distractors) and a between-subject factor of presentation (grouped and mixed). Mauchly's test of sphericity was not calculated because there were only two levels to either main effect ($df = 1$). The main effect of distractor items was significant, $F(1, 61) = 52.1$, $p < .001$, $\eta p^2 = .46$; again however, the main effect of presentation was not significant $F(1, 61) = .05$, $p = .82$.

The lack of significant difference for the main effect of presentation is understandable when we compare the proportions presented in Figures 7.2 and 7.3 again. There is very little difference between the proportion of responses by presentation for either the acceptable items (Figure 7.2) or the distractor items (Figure 7.3). For both ANOVA tests, however, the within-subjects factor of acceptable items and distractor items showed a significant difference for the number of 'Hits' and 'False Alarms' respectively. These differences are analysed below for each presentation type separately.

7.4.2 Comparison of the acceptable items

RQ2 Do the learners respond to the frequent senses with more 'Hits' than to the infrequent senses?

For the acceptable items, the responses were categorised as 'Hits' for items responded to as 'Correct'. Figure 7.5 shows the mean number of 'Hits' by item type for the acceptable items. The means are for the 37 participants in the grouped presentation and the 26 participants in the mixed presentation. The means have been adjusted downwards for the number of 'False Alarms' to the illogical distractors.
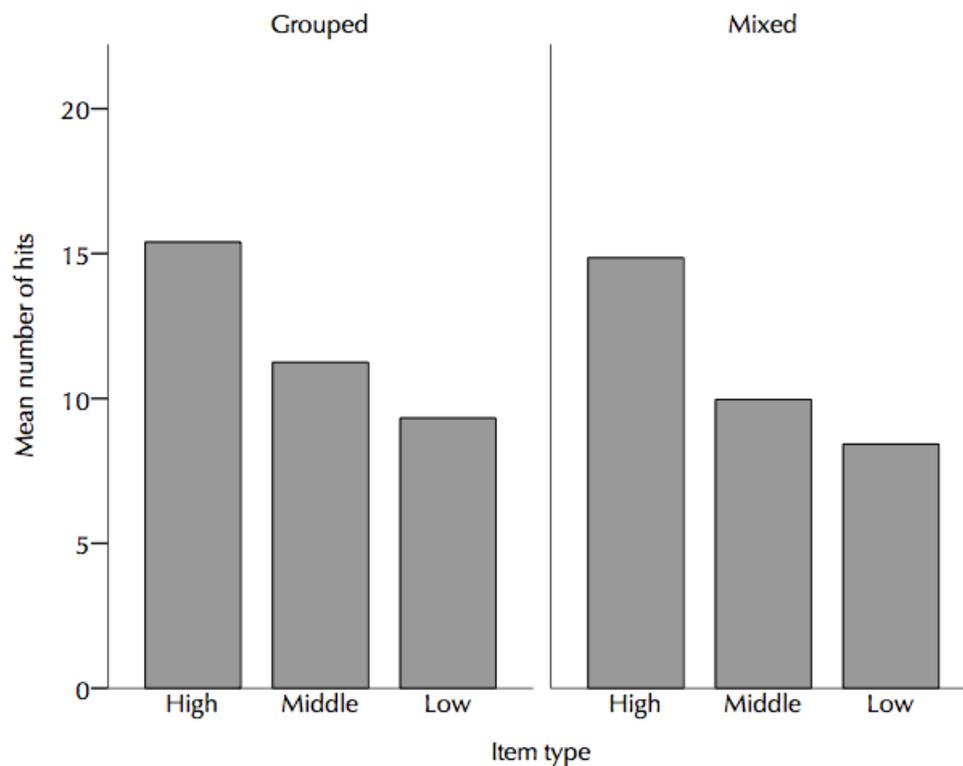
Figure 7.5
Adjusted mean number of 'Hits' on the TPM by acceptable item-type, organised by presentation.

The adjusted data for the acceptable items were analysed for normality to determine whether parametric or non-parametric tests should be used. Table 7.6 presents the results to a Shapiro-Wilk test of normality.

Table 7.6

Shapiro-Wilk Test of Normality for the Results of the Acceptable Items TPM

Presentation	Item type	Statistic	Sig.
Grouped (df = 37)	High	0.892	0.002
	Middle	0.969	0.388
	Low	0.982	0.814
Mixed (df = 26)	High	0.881	0.006
	Middle	0.972	0.683
	Low	0.957	0.344

The high-frequency sense for both presentations showed significant difference from the null hypothesis of normality at $p < .05$. The results for these items were also significantly skewed, *skew(grouped)* = -1.43, $z = -3.68$, $p < .05$ and *skew(mixed)* = -1.25, $z = -2.74$, $p < .05$. As discussed in the results for the pilot study (Chapter 5, page 136), this result is likely due to a ceiling effect because the participants were expected to know at least one sense of each of the target words. According to the hypothesis of intra-word frequency, the participants were most likely to know the high-frequency sense. Non-parametric tests were used for comparisons with these results because the distribution was skewed and did not conform to a normality.

A non-parametric Friedman test of differences among repeated measures found that there was a significant difference between the three levels for the acceptable items of both presentations: $\chi^2 grouped(2, N = 37) = 47.6$, $p < .001$ and $\chi^2 mixed(2, N = 26) = 28.9$, $p < .001$. Post-hoc tests using a Bonferroni correction were conducted to compare the means of the separate item types.

For the grouped presentation, the high-frequency sense ($M = 15.40$, $Mdn = 16.47$, $SD = 3.31$) was significantly different from both the middle-frequency sense ($M = 11.24$, $SD = 3.64$), $z = 4.97$, $p < .001$, $r = .82$, and the low-frequency sense ($M = 9.32$, $SD = 4.39$), $z = 5.48$, $p < .001$, $r = .91$. The middle-frequency sense was not significantly different from the low-frequency sense at $t(36) = 2.46$, $p = .019$, ($\alpha = .016$).

For the mixed presentation, the high-frequency sense ($M = 14.85$, $Mdn = 16.09$, $SD = 4.30$) was significantly different from both the middle-frequency sense ($M = 9.96$, $SD = 4.45$) at $z = 4.18$, $p < .001$, $r = .81$, and the low-frequency sense ($M = 8.42$, $SD = 3.72$),

$z = 3.85$, $p < .001$, $r = .74$. Again, the middle-frequency sense was not significantly different from the low-frequency sense at $t(25) = 1.83$, $p = .079$, $(\alpha = .016)$.

The results indicate that senses that express more intra-word frequency are more acceptable to the L2 learners. However, since there was no significant difference found between the results of the middle and low-frequency senses, it appears there may be a minimum number of occurrences for intra-word frequency to influence the acceptability judgements. These results are considered in more detail in the discussion (Section 7.5).

7.4.3 Comparison of the distractor items

RQ3   *Do the learners respond to the logical distractors with more 'False Alarms' than to the illogical distractors?*

For the distractor items, the responses were categorised as 'False Alarms' for items responded to as 'Correct'. Figure 7.6 shows the mean number of 'False Alarms' for the distractor items. The data are organised by the grouped presentation ($n = 37$) and mixed presentation ($n = 26$). The means were not adjusted for the distractor items.
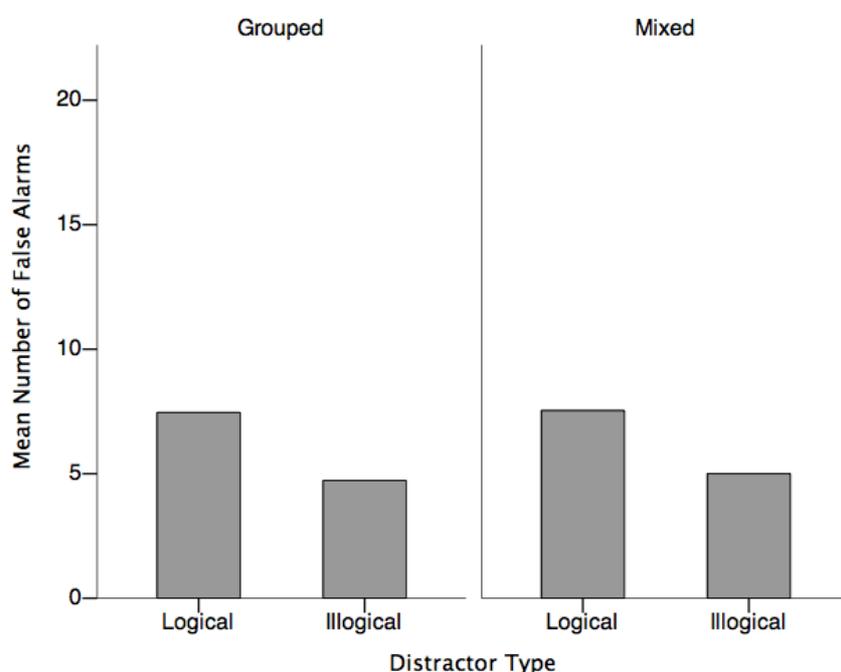
Figure 7.6
Mean number of 'False Alarms' on the TPM by distractor item-type, organised by presentation.

176

Table 7.7 presents the results to a Shapiro-Wilk test of normality to determine whether parametric or non-parametric tests should be used.

Table 7.7
Shapiro-Wilk Test of Normality for the Results of the Distractor Items TPM

Presentation	Distractor	Statistic	Sig.
Grouped (df = 37)	Logical	0.935	0.031
	Illogical	0.958	0.172
Mixed (df = 26)	Logical	0.943	0.160
	Illogical	0.938	0.117

The mean for the logical distractor in the grouped presentation showed significant difference from the null hypothesis of normality at $p < .05$. Non-parametric tests were used for comparisons within the results for the grouped presentation, but parametric tests were used for the mixed presentation.

For the grouped presentation, the mean number of 'False Alarms' to the logical distractor ($M = 7.46$, $SD = 3.73$) was significantly different from the illogical distractor ($M = 4.73$, $SD = 2.81$), $z = 4.26$, $p < .001$, $r = .70$. For the mixed presentation, the mean number of 'False Alarms' to the logical distractor ($M = 7.54$, $SD = 3.58$) was also significantly different from the illogical distractor ($M = 5.00$, $SD = 3.36$), $t(25) = 5.76$, $p < .001$, $d = 0.73$.

Both presentations showed a medium effect size for the difference between the logical and illogical distractors (*Grouped, r = .70*; *Mixed, d = .73*). This indicates that, for both presentations, semantic similarity was an influence on the L2 learners' semantic acceptability judgements, outside of the influence of intra-word frequency.

7.5 Discussion

The items of the revised version of the TPM were presented to the participants in one of two ways, grouped or mixed. In the grouped presentation, all the items using the same target word were presented together. In the mixed presentation, no items sharing the same target word were repeated on the same page of the test. The reason

for having the two presentations was to see if the participants would benefit from the opportunity to compare the use of the target word across items. At the outset, however, there was a concern that the participants might see a pattern of acceptable and distractor items in the grouped presentation. This turned out not to be the case. When the frequency of responses was analysed independent of item type, the two presentations were nearly identical in the pattern of three 'correct' responses for every two 'incorrect' responses. The participants cannot be said to have identified the pattern of item types in the grouped presentation any more than they would have for the mixed presentation.

The participants in the grouped presentation did not appear to benefit from the opportunity to compare items. It was hypothesised that these participants might have seen semantic similarities across the use of the target words. This semantic similarity might have made the use of the target word in some items more acceptable and prompted the participants to respond that the item was correct more often in the grouped presentation. While the mean number of 'Hits' for the acceptable items was slightly higher in the grouped presentation, there was no significant difference for the main effect of presentation, and so there is no evidence that the participants benefitted from the grouped presentation.

This result is in contrast with findings from Klepousniotou et al. (2008), Klepousniotou and Baum (2007) (reviewed in Section 2.4.1) and Verspoor and Lowie (2003) (reviewed in Section 2.6 and replicated in Chapter 3). Although these researchers were not directly testing the L2 knowledge of polysemous words, there is some value in looking at whether there are points of contact. Using a priming methodology, Klepousniotou et al. (2008) and Klepousniotou and Baum (2007) found faster judgements for polysemous word pairs with highly overlapping senses compared to ambiguous words with non-overlapping senses. Drawing on this finding, one might propose that the grouped presentation would also 'prime' the meaning of the target word, making the more semantically similar senses more acceptable. However, the priming methodology is an online task, which indicates how the senses are stored and represented. In contrast, the TPM gives the participant

more time to deliberate, and the results are more indicative of how meaning is processed. The two methodologies are not comparable and any differences between the findings of the two studies can be attributed to the different cognitive processes they engage.

Unlike the priming methodology, the methodology Verspoor and Lowie used is more comparable to the TPM. In Verspoor and Lowie's study, the participants were shown two sentences using the same polysemous word in different senses. They were given the meaning of one sense and asked to guess the meaning of the other sense. The researchers showed that the participants were more successful in guessing if they were given a core sense as a cue rather than an extended sense. While the participants in Verspoor and Lowie's study were shown to benefit from the opportunity to compare a known sense to an unknown sense, the grouped presentation of the TPM did not show the same benefit. One possible reason for the difference is that the participants were not explicitly required to compare the senses on the TPM. It might be that participants would focus solely on sentence level context to construe the sense of a target word, and they would ignore the other uses of the word unless required to do so. Furthermore, the TPM presented both acceptable and unacceptable items. If all the items were acceptable and all the senses shared in semantic similarity, the grouped presentation might have benefitted the participants. However, the TPM used the target words in senses with varying degrees of semantic similarity, including illogical distractors designed to have no relation to any acceptable sense. The inconsistency in semantic similarity across the items might have inhibited the participants from comparing a known sense to an unknown sense.

As it stands, there isn't any evidence that the participants responded to the test items any differently on the grouped presentation than they did on the mixed presentation. I will now turn to a discussion of the results for the acceptable items and distractor items, but I won't give special attention to the different presentations. In subsequent chapters the data from the two presentations will be combined.

Across both presentations, the mean number of 'Hits' for the high-frequency items was significantly greater than the number of 'Hits' for either the middle or low-frequency items. This result supports the intra-word frequency hypothesis, which proposes that learners will learn polysemous senses based on how frequently the senses are encountered. As the senses become less frequent, the effect becomes less pronounced. The number of 'Hits' for items using the middle-frequency sense were greater than the those using the low-frequency sense; however, this difference was not significant in either presentation when adjusted using a Bonferroni correction. This indicates that a sense might have to be highly frequent for intra-word frequency to be a reliable factor in deciding whether a learner knows the sense or not. As senses become less frequent, other factors might be more important. Intra-word frequency is the only factor considered in this analysis, and so it remains ambiguous why less frequent senses are acceptable or unacceptable to the participants.

In addition to intra-word frequency, another factor affecting the participants' responses is the relation of the target sense to the core sense of the target word. From the perspective of linguistic description, the core sense is often identified as the central sense to the category of the polysemous word. The core sense can be understood as the sense which lends the greatest semantic coherence to the different polysemous senses (Geeraerts, 2007; Verspoor & Lowie, 2003). Researchers have proposed that L2 learners can learn a new sense of a polysemous word by its similarity to the core sense (Boers, 2013; Boers & Lindstromberg, 2007; Boers & Lindstromberg, 2008; Csábi, 2004; Khodadady & Khaghaninizhad, 2012; MacLennan, 1994; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). This research was reviewed in chapters 1 and 2.

The organisation of the acceptable items by intra-word frequency does not necessarily correspond to how the items would be organised in relation to the core sense. Often the most frequent sense would also be the core sense of the target word, but there are cases where the core sense might also be less frequent. For example, the

high-frequency sense for *course* refers to EDUCATION; however, the core sense[18] for *course* is the middle-frequency sense which refers to PLANNED DIRECTION. Furthermore, some items can be considered semantically closer to the core sense than others. For example, the core sense of *arm* refers to the BODY. The low-frequency sense of *arm* refers to FURNITURE, which might be considered to be semantically closer to the core sense than the middle-frequency sense of WEAPONS. The point is that if L2 learners develop their knowledge of a polysemous word as a semantically coherent category, then the relationship of a sense to the core sense might counter or reinforce the influence of intra-word frequency.

The distractor items were designed as a way of measuring the influence of semantic similarity to the core sense independent of the influence of intra-word frequency. This independence was achieved because neither sense was listed in the dictionary (OED Online, 2011); moreover, native speakers rejected the distractor items as unacceptable uses of the target words. For these reasons, I am confident that the participants would not have encountered the distractors in their L2 language experience. The reason why the distractors can be used to measure the influence of semantic similarity is because the logical distractors were designed to share semantic similarity to the core sense, while the illogical distractors were designed to share no semantic relationship. The participants responded with significantly more 'False Alarms' for the logical distractors than the illogical distractors in both presentations. These results lend support that the learners will extend their existing knowledge of the polysemous words to novel senses. The analysis here was conducted using all the learners and all the distractors. Going forward, I will also consider the extent to which individuals, groups, or stimuli did not adhere to this pattern.

Since the logical distractors provide support for the influence of semantic similarity to the core sense of the target word on the participants' responses, it is reasonable to propose that semantic similarity is an influence in their responses to the acceptable

---

18. The identification of the core sense is based on insights from cognitive linguistics. The method of identification is explained in detail in Chapter 9.

items as well. However, the current analysis was not designed to investigate the effect of semantic similarity for the acceptable items. The acceptable items were originally categorised by intra-word frequency (high, middle and low-frequency senses) in order to investigate the influence of intra-word frequency on L2 learner knowledge of polysemous words. In order to reanalyse the responses for the influence of semantic similarity, the acceptable items need to be re-categorised by their similarity to a proposed central sense. This analysis is reported in Chapter 9. The expectation is that learners will have better knowledge of senses that are more similar to the central sense than words that are semantically distant.

Besides semantic similarity to the core sense, three other factors may also be an influence on L2 knowledge of polysemous words. First, learners' may find it easier to learn concrete words over more abstract words. If this is so, then the senses associated with concrete words like *arm* or *board* might be easier to learn than senses the associated with more abstract words like *position* or *form.* An analysis was conducted where the mean number of 'Hits' for more concrete words was compared to less concrete words. Concreteness ratings were taken from Brysbaert, Warriner and Kuperman's list of norms of concreteness ratings (2014). There was no significant difference for the mean number of 'Hits' between the two groups, $p > .05$. This result corroborates the finding of Kellerman (1986), who reported that subjective measures of concreteness did not correlate with judgements on whether a polysemous sense could be transferred from the L1 to the L2. However, this analysis does not entirely disqualify the influence of concreteness. One issue is that concreteness ratings were made based on the word form and not on specific senses. For example, the *arm* of a body would likely be judged as more concrete than the retail *arm* of a large corporation. However, to the best of my knowledge, concreteness norms at this level of nuance have not been conducted. Until such norms exist, the effect of concreteness will remain a question.

Another influence on L2 knowledge of polysemous word is the L2 learners' first language (their L1). If a polysemous word in the L1 expresses the same senses as the equivalent word in the L2, then it is reasonable to propose that the L1 might be a

positive influence on the L2 learner's knowledge of these senses. In Chapter 10 the factor of L1 influence will be investigated in addition to the factor of semantic similarity to the core sense.

L2 proficiency is a third factor that may have influenced the learners' knowledge of polysemous words. The participants have been treated as a single group and individual differences haven't been considered. This is warranted because they represent a largely homogenous group studying in the same L2 English course and speaking Arabic as their L1. However, there are differences in their L2 proficiency which can be investigated. The participants all wrote the Levels Test (Nation, 1990), which is a test of breadth of vocabulary knowledge. This test can be used as an independent measure of their proficiency in L2 knowledge to see whether this factor can explain differences between participants. In Chapter 8, the factor of proficiency will be investigated using the participants' scores on the Levels Test.

7.6 Conclusion

The main goals of this study were to investigate the effect of three factors on learner responses to a revised version of the TPM: intra-word frequency, semantic similarity and item presentation. First, intra-word frequency was investigated by comparing the number of 'Hits' for the high, middle and low-frequency sense items. The participants responded with more 'Hits' for the high-frequency items than the middle-frequency items, and the low-frequency items were responded to with the fewest 'Hits'. These results were taken as evidence that L2 learners develop their knowledge of polysemous senses according to the intra-word frequency of the senses' occurrence in the language. Second, semantic similarity was investigated by comparing the number of 'False Alarms' between the distractor items. The logical distractors were designed to be meaningful but incorrect uses of the target word, while the illogical distractors were designed to have no semantic relationship to the core sense of the target words. The logical distractors were responded to with more 'False Alarms' than the illogical distractors. Since the logical distractors were incorrect, the participants could not have encountered the use of the word before. This indicates that the participants found the logical distractors more acceptable

based on similarity to the core sense of the polysemous word. Finally, the comparison of a mixed and grouped presentation of the items found that there was no significant difference in the responses of the participants according to any of the analyses. The frequency of participants responding with Correct, Incorrect or Don't Know was nearly identical across presentations, indicating that the participants did not guess the pattern of correct and incorrect items in the grouped presentation. Also, the mean number of 'Hits' for the acceptable items or 'False Alarms' for the distractor items was not significantly different between presentations, indicating that participants did not benefit from having items grouped by target word. Since no measurable difference has been found between the two presentations, I will combine the data from the two presentations into a single data set for further analysis. In the following chapters, the TPM data will be analysed for proficiency, semantic similarity, and L1 influence.

Chapter 8: The effect of proficiency

The results gathered in Chapter 7 are reanalysed for the effect of proficiency. Proficiency is here determined according to the learners' L2 vocabulary size using the Vocabulary Levels Test to identify a high-proficiency group and a low-proficiency group. The prediction is that the learners in the high-proficiency group will respond to the acceptable items with more 'Hits' and to the logical distractors with fewer 'False Alarms'. The results are compared both between proficiency groups and within the same group between the different item types.

8.1 Introduction

In the previous chapter, I presented the results of the revised version of the Test of Polysemous Meanings (TPM), taken by Arabic learners of English. The results of the acceptable items confirmed that the learners could identify more senses of English polysemous words if those senses were more frequent in the language. This result supports the intra-word frequency hypothesis which proposed that learners are more likely to learn meanings according to their intra-word frequency relative to other meanings of that word. The results for the distractor items showed that semantic similarity was an influence on the learners' responses. The learners would more frequently identify a distractor item as acceptable when used in that sentence context if the sense of the target word bore semantic similarity to the core sense of the target word. This can't be attributed to intra-word frequency because the distractor sentences used the target word in a sense that did not appear in the dictionary and that native speakers deemed were unacceptable. This result supported the hypothesis of semantic similarity that learners extrapolate their knowledge of words from one sense to another.

However, there is evidence that the results were not consistent across all the learners. Table 8.1 shows that while the mean number of 'Hits' decreased as the senses expressed less intra-word frequency, the standard deviation of the number of 'Hits' increased. This increase in the standard deviation indicates that there was

greater variation in the number of 'Hits' between learners for the lower frequency senses.

Table 8.1
Mean number of Hits and standard deviations for the acceptable items on the TPM[19]

Item type	Hits	Standard Deviation
High-frequency	15.17	3.73
Middle-frequency	10.71	4.01
Low-frequency	8.95	4.12

Likewise, Table 8.2 shows that the standard deviations for the mean number of 'False Alarms' to the distractor items were also high.

Table 8.2
Mean number of False Alarms and standard deviations for the distractor items on the TPM

Item type	False Alarms	Standard Deviation
Logical distractor	7.49	3.64
Illogical distractor	4.84	3.02

Thus, while there was evidence to support the hypotheses of intra-word frequency and semantic similarity, the standard deviations to the 'Hits' and 'False Alarms' indicate that there was a fair amount of variation in how the the learners responded.

One possible reason for the variation is that there may have been differences in L2 proficiency among the group of learners who took the TPM. As mentioned in Section 7.2.1, the learners were all studying in the same course of intensive English, with the objective of achieving a 5.0 on the IELTS English Language Test. The IELTS measures English language proficiency through a suite of four examinations for

---

19. As discussed in Chapter 7, the results of the grouped and mixed presentations were combined. Also, the results presented here were adjusted downwards based on the formula presented in Section 7.2.4.

186

reading, writing, listening and speaking. It's reasonable to assume that the participants were within the range of 4.0 to 5.0 on the IELTS. This variation in their general language proficiency might explain the variation in the results. One would expect that as the learners' proficiency increased, their responses on the TPM would approach the native speaker norm. This proposition can be presented as a hypothesis of proficiency for L2 knowledge of polysemous words: as learners improve in their L2 proficiency, they will respond with more 'Hits' to the acceptable items and fewer 'False Alarms' to the distractor items.

To investigate the influence of proficiency on the learners' knowledge of polysemous meanings, it is necessary to use a measure of proficiency that is independent of the TPM. The Vocabulary Levels Test (VLT) (originally written by Nation, 1990; revised by Schmitt, Schmitt, & Clapham, 2001) is a widely used measure of vocabulary size. It measures whether learners can demonstrate competent knowledge of words according to frequency bands: the bands are the first 2000, 3000, 5000 and 10,000 words in English. The test focuses on a single feature of language knowledge, breadth of vocabulary knowledge. Despite this narrow focus, Milton notes that tests of vocabulary size, such as the VLT, are good indicators of overall language competence (Milton, 2009, pp. 170-171).

There is some evidence that the Levels Test, specifically, correlates well with other measures of language competence. Stæhr (2008) found that the Levels Test correlated with a test of reading at $r = 0.83$, with a test of writing at $r = 0.73$ and with a test of listening at $r = 0.69$. Milton (2009) argues that these results are indicative of the greater importance that vocabulary knowledge holds for the skills of reading and writing where text coverage is important for comprehension. In contrast, he comments that speaking skills can rely more on context and gesture and lower vocabulary knowledge is less of a detriment.

It is understood that other measurements of proficiency might also be useful, such as a test of grammar knowledge or a writing sample. However, the fact that the Levels Test correlates well with a measurement of reading makes it an appropriate

187

comparison to the TPM: the test's mode of presentation is written and the participants' performance on the test is directly related to their English reading skills. By using the Levels Test, proficiency is understood to be primarily a construct of vocabulary size and reading comprehension.

Two research questions are proposed for the investigation of influence of proficiency on polysemous word knowledge:

RQ1   *Do the learners with a higher L2 proficiency respond with more 'Hits' to acceptable items than learners with a lower proficiency?*

RQ2   *Do the learners with a higher L2 proficiency respond with fewer 'False Alarms' to distractor items than learners with a lower proficiency?*

8.2 Method

8.2.1 Instruments

8.2.1.1 Vocabulary Levels Test

The Vocabulary Levels Test (VLT) was used to measure the participants' vocabulary size and to organise the participants into high and low proficiency groups. The VLT is one of the best-known instruments for measuring a learner's vocabulary size, it correlates well with tests of reading, and it has a history of use for research purposes (Schmitt et al., 2001). This supports the decision to use the results of this test for a comparison with the results of the TPM. The entire set of four tests covers the 10,000 most frequent words in English in four bands. The 10,000 level test was excluded from this study because it was considered far above the level of the students. There are 30 target words to each test. On each test, a set of three target words is grouped together with three distractor words. In each set, the learner is required to match the three target words to the correct three definitions. An example of one set of words is presented below:

1. original
2. private   _ complete
3. royal   _ first
4. slow   _ not public

5. sorry

6. total

There are six such sets of target words to each test. The test is designed to indicate whether or not the learner has competent knowledge for the frequency band. Competency is set at a minimum result of 90% or 27 of the 30 items correct. The test is not designed to index proficiency on a scale below this minimum. The VLT was originally written by Nation (1990) and then revised for content validity by Schmitt, Schmitt and Clapham (2001) in two versions. The students took Version 1 of the revised test as a diagnostic at the beginning of their course. (The test is presented in Appendix 4 on page 365.)

8.2.1.2 TPM

The Test of Polysemous Meanings (TPM) and its results were presented in Chapter 7. An example of the test is presented in Appendix 3 on page 343. To briefly summarise what was said there about the design, the TPM tests the participants' knowledge of the meanings of 20 polysemous words as nouns. The 20 target words were presented in sentences as five different item types. Three were correct uses of the word in different senses and categorised by intra-word frequency: high-frequency, middle-frequency and low-frequency. The two others were distractors, one logical and the other illogical. There were 100 items in total. The participants had to select whether the use of the target word was 'Correct' or 'Incorrect', and there was also a 'Don't Know' option.

8.2.2 Participants

As presented in Chapter 7, data was collected from 67 Arabic speakers enrolled in the final course of an intensive English at a technical college in Qatar. After completing this course the students were expected to enter their English-language college-level programs in business or applied technology. Two participants were

189

absent on the day the VLT was taken. This left 65 participants who took the 2000-band test of the VLT.[20]

8.3 Results and analysis

8.3.1 Results

8.3.1.1 Vocabulary Levels Test

The participants who took the TPM represented a single treatment group. The purpose of the results to the VLT was to identify which of these participants represented learners with a high proficiency of English and which participants represented learners with a low proficiency. Once the two proficiency groups were identified using the results of the VLT, the two groups could be compared using their results on the TPM. The two proficiency groups needed to be of equal number because they were to be compared using a two-way ANOVA. Of the total 65 participants who wrote the VLT, 13 achieved a result of 90% or higher on the 2000-band test of the VLT. These 13 participants were selected to comprise the high-proficiency group. To comprise the low-proficiency group, the 13 participants with the lowest VLT scores were initially selected; however the variances were unequal for the low and high proficiency groups, $F(1, 24) = 11.13$, $p = .03$. The three participants with the lowest VLT scores (4, 8, and 9 out of 30) were removed as outliers and replaced with three participants whose VLT scores were closer to the median of the group ($Md = 16$). With this revision the variances were equal for the low and high proficiency groups, $F(1, 24) = 7.44$, $p = .397$. The results of the remaining participants were excluded from the analysis.

The descriptive statistics for the high-proficiency group are as follows: $max = 30/30$, $min = 27/30$, $Mdn = 29/30$, $M = 28.46/30$. These are the descriptive statistics for the low-proficiency group: $max = 19/30$, $min = 15/30$, $Mdn = 17/30$, $M = 17.15/30$. The differences between the two groups are quite substantial. This means that

---

20. Thirty-seven of the participants also wrote the 3000 and 5000 bands of the VLT. However, only two of these participants showed competency at the 3000-band test, indicating that the 2000-band test was appropriate for measuring their vocabulary proficiency.

differences in the results on the TPM can confidently be attributed to L2 proficiency and not to other factors of individual difference.

8.3.1.2 Test of Polysemous Meanings

The results from the TPM were presented in Chapter 7. In Tables 8.3 and 8.4 below, the results for the participants in the two proficiency groups have been selected from the total data set.

Table 8.3
Mean number of 'Hits'* to the acceptable items by proficiency group.

		Acceptable item type						
		High frequency		Middle frequency		Low frequency		
		Mean	SD	Mean	SD	Mean	SD	N
Proficiency group	High	17.91	(1.57)	14.69	(2.96)	10.09	(3.65)	13
	Low	12.19	(4.96)	7.68	(3.73)	7.29	(3.57)	13

**Note:* The number of 'Hits' was adjusted downwards by participant based on the number of 'False Alarms' to the logical distractors.

Table 8.4
Mean number of 'False Alarms' to the distractor items by proficiency group.

		Distractor item type				
		Logical distractor		Illogical distractor		
		Mean	SD	Mean	SD	N
Proficiency group	High	5.23	(3.22)	2.85	(2.27)	13
	Low	9.23	(3.49)	6.08	(3.35)	13

8.3.2 Analysis

8.3.2.1 Acceptable items

The results for the number of 'Hits' to the acceptable items were analysed by a 3 x 2 analysis of variance using a mixed-design. Acceptable items (high, middle and low frequency) were set as the within-subjects factor and proficiency (high and low

proficiency) was set as the between-subjects factor. Levene's test of homogeneity based on the median indicated that variances were homogeneous (i.e. more or less the same for both high and low-proficiency groups) for all levels of the acceptable items repeated: high frequency $F(1,24) = 3.02$, $p = .095$, middle frequency $F(1,24) = 0.42$, $p = .53$, and low frequency $F(1,24) = .02$, $p = .904$. Mauchly's test indicated that the assumption of sphericity had not been violated for the within-subjects comparison of acceptable items, $\chi^2 (2) = 2.23$, $p = .33$, indicating an equality of variance between the three levels of the acceptable items.

The main effects were significant for both item type $F(2,48) = 37.45$, $p < .001$, $\eta p^2 = .609$, and proficiency $F(1,24) = 15.88$, $p = .001$, $\eta p^2 = .398$. However, their significance was qualified by an interaction between acceptable items and proficiency $F(2,48) = 6.77$, $p = .003$, $\eta p^2 = .220$. To identify the source of the interaction, pairwise comparisons were made between the results of the two proficiency groups for same item types (i.e. the results of the high-proficiency group for the middle-frequency sense compared to those of the low-proficiency group). In order to conduct pairwise comparisons, a Shapiro-Wilk test of normality was conducted on the interaction between item type and proficiency group. The test showed that the distribution of the results for the low proficiency group on the high frequency item-type were not normally distributed, $W(13) = .83$, $p = .017$. Comparisons with this group of data used non-parametric tests. Three pairwise comparisons were made. With multiple comparisons there is an increased chance that a rare event will cause an incorrect rejection of the null hypothesis (a Type I error); therefore, a Bonferroni correction was used to lower the desired *alpha* from $\alpha = .05$ to $\alpha = .016$ (i.e. $.05/3 = .016$). Post-hoc analyses using Bonferroni correction indicated that there was a significant difference between the high and low proficiency groups for the high frequency items, $U = 13.5$, $p < .001$, $r = .71$, and the middle frequency items, $t(24) = 5.13$, $p < .001$, $d = 2.01$, item types, but there was no significant difference between the proficiency groups for the low frequency items, $t(24) = 1.11$, $p = .28$. This analysis generally answers the first research question, that the learners with a higher L2 proficiency do respond with more 'Hits' to acceptable items than learners with a

192

lower proficiency. The lack of a significant difference between the proficiency groups for the low-frequency items can be explained by the hypothesis of intra-word frequency. Since the low-frequency senses of the target words are so infrequent, it is likely that neither proficiency group had encountered these senses very often. That being the case, neither group would have been familiar with many of these senses. In contrast, the high-proficiency group performed better on the middle-frequency items than the low-proficiency group. This might mean that the high proficiency group had encountered these senses more frequently than the low proficiency group, which would support the hypothesis of intra-word frequency. However, this is not the only possible explanation for the results. The participants in the low-proficiency group may have been less successful at learning the middle-frequency senses despite encountering the senses as frequently as the high-proficiency group.

8.3.2.2 Distractor items

The results for the number of 'False Alarms' to the distractor items were analysed by a 2 x 2 analysis of variance using a mixed design, with proficiency (high and low) as the between subjects variable and item type (logical distractor and illogical distractor) as the within subjects variable. Levene's test homogeneity based on the median indicated that the condition of variance was satisfied for both item types: Logical distractor, $F(1,24) = .663$, $p = .42$, and illogical distractor, $F(1,24) = .832$, $p = .37$. The main effects of distractor items, $F(1,24) = 18.21$, $p < .001$, $\eta p^2 = .431$, and proficiency, $F(1,24) = 13.09$, $p = .001$, $\eta p^2 = .353$, both showed significant differences, but there was no interaction between the two main effects, $F(1,24) = .35$, $p = .56$. These results confirmed the second research question, that the learners with a higher L2 proficiency responded with fewer 'False Alarms' to distractor items than learners with a lower proficiency. This indicates that as the learners' proficiency increases, they not only develop an expanded understanding of a word's use, but also a stronger understanding of the restrictions on a word's use.

8.4 Discussion

At the outset of this discussion, it is worthwhile to consider the effect of the downward adjustment on the results to the acceptable items for each proficiency

group. After discussing the adjustment of the results, I will move on to see how the hypotheses of intra-word frequency and semantic similarity relate to the proficiency results.

8.4.1 Adjustment of the results

As discussed in Chapter 7, the results of the acceptable items were adjusted downwards based on the number of 'False Alarms' to the illogical distractors. This was done because a participant who responds with many 'False Alarms' to the illogical distractors can be expected to interpret all the item types creatively in ways which hadn't been controlled for in the instrument design. The participants in the low-proficiency group responded with a higher number of 'False Alarms' to the illogical distractors, thus there was a greater downwards adjustment to their results than to the results of the high-proficiency group. Figure 8.1 presents the downward adjustment for the acceptable-item types by proficiency group.

Figure 8.1

194
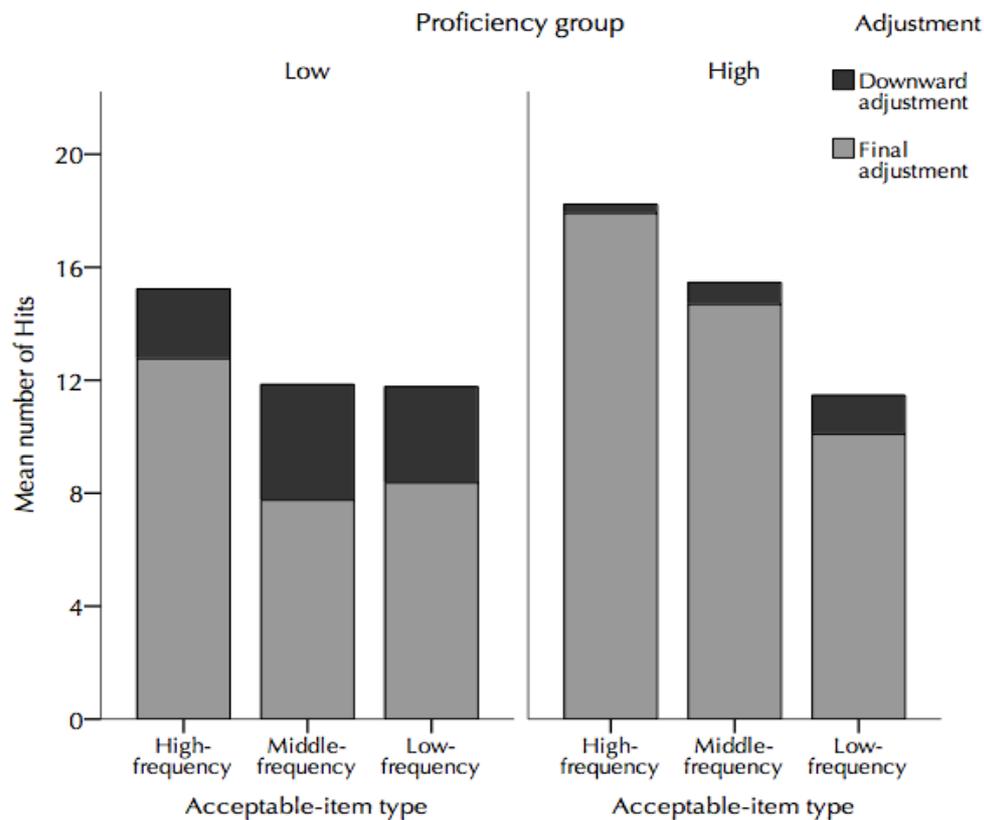
Comparison of the downward adjustment by item type between proficiency groups.

For the high-proficiency group, there is a small downward adjustment for each item type. This contrasts with the adjustment for the low-proficiency group, where the adjustment is quite sizeable. The differences in adjustments increases the relative distance between the high and low groups.

Although the amount of adjustment appears quite substantial, an analysis of the unadjusted results proved to be comparable to adjusted results. For the unadjusted results, the main effects of proficiency, item type and the interaction between the effects were also significant ($\alpha =. 01$). Perhaps the best way to address the consequence of the adjustment is to look at the effect size for the three main effects, as presented in Table 8.5.

Table 8.5
Comparison of the unadjusted and adjusted results by effect sizes from 2x3 analyses of variance.

	Effect size (ηp^2)		
	Item-type	Proficiency	Interaction
Unadjusted	.674	.247	.183
Adjusted	.626	.470	.158

While the effect sizes for item type and interaction are comparable between groups, there was a strong increase in the effect size for proficiency when the results were adjusted. The low-proficiency group was more liable to uncontrolled creative interpretation as indicated by their higher frequency of 'False Alarms' to the illogical distractors. For this reason, I believe that the adjusted results are a better reflection of the participants' understanding of the target words in relation to the native speaker norms.

Since the adjustment is based on the number of 'False Alarms' to the illogical distractor, it is fitting to ask why the lower-proficiency group should respond with more 'False Alarms' to the illogical distractors. The target words were selected from the first 1000 most frequent words in English, so the learners in the low-proficiency group were expected to know at least one sense of the target words. However, the

VLT results indicated that, unlike the high-proficiency group, the low-proficiency group did not have command of the 2000-word frequency band. Since the items were designed using the 2000-word frequency band, it could be that the lower-proficiency group had difficulty with the sentence context of the items.

This poses a reliability issue. The sentence context was designed to be unambiguous so that the learner could focus on the target word. The number of 'False Alarms' from the low-proficiency group, coupled with their low VLT scores, indicates that the sentence context might have been too difficult for them. However, the results of the low-proficiency group are much better than chance. As discussed in Chapter 7, the results of any participant were removed if the number of 'False Alarms' to the illogical distractors was higher than 'Hits' for any of the acceptable item-types. For example, a given participant might respond to the high, middle and low-frequency items with a mean number of 15, 10 and 8 'Hits' respectively. That participant would be excluded from the analysis if he responded to the illogical distractors with a mean number of 9 'False Alarms', because there would be more 'False Alarms' than 'Hits' for the low-frequency items. If the TPM was responded to according to chance, then there would be a 50% probability of more 'False Alarms' for the illogical distractors than 'Hits' for one of the acceptable item categories. Since there are three categories of acceptable item, there are three possible occasions for more 'False Alarms' than 'Hits'. Therefore, according to chance, there would be an 87.5%[21] probability that the participants would have responded with more 'False Alarms' than 'Hits' to at least one of the three acceptable item-types. In the actual results, only 3 of the 10 participants with the lowest VLT scores had more 'False Alarms' than 'Hits' for any of the acceptable item-types.[22] Thus, we can be fairly confident that the number

---

21. *P*(higher 'False Alarms' than 'Hits' for 1 of 3 categories)
    $= 1 - P$(higher 'Hits' than 'False Alarms') ^ 3 categories
    $= 1 - (1/2)$ ^ 3
    $= .875$

22. The results of these participants were replaced with the results of participants who responded with more 'Hits' than 'False Alarms'.

'False Alarms' for the low-proficiency group was less due to chance and more due to other strategies of responding.

What response strategies might a low proficiency learner employ? A creative learner might take risks on making meaning, as indicated by the 'False Alarms', and in the end actually learn effectively and leave the low-proficiency group. In Chapter 9, I consider how certain types of metaphor and metonymy might facilitate an acceptable judgement on the part of a more creative learner. In contrast, a less creative learner might respond only based on collocation. For example, in one item the target word *class* appears as the collocation "English class". This association might prompt the learner to respond that the target word is used acceptably. Lacking such an association, the learner might respond haphazardly and thus create more 'False Alarms'. One would expect this learner to develop his deep word knowledge more slowly and may never get out the low-proficiency group.

8.4.2 The hypothesis of intra-word frequency

The hypothesis of intra-word frequency is better supported by the results of the high proficiency group than the low proficiency group. The hypothesis of intra-word frequency is that L2 learners will have better knowledge of a polysemous sense if it occurs more frequently in the language. For the high proficiency group, as the frequency of the item type increased, so did the number of 'Hits'; however, for the low proficiency group, there was no significant difference between the number of 'Hits' to the middle-frequency items and the low-frequency items. Figure 8.2 shows the interaction between between proficiency group and acceptable-item type.
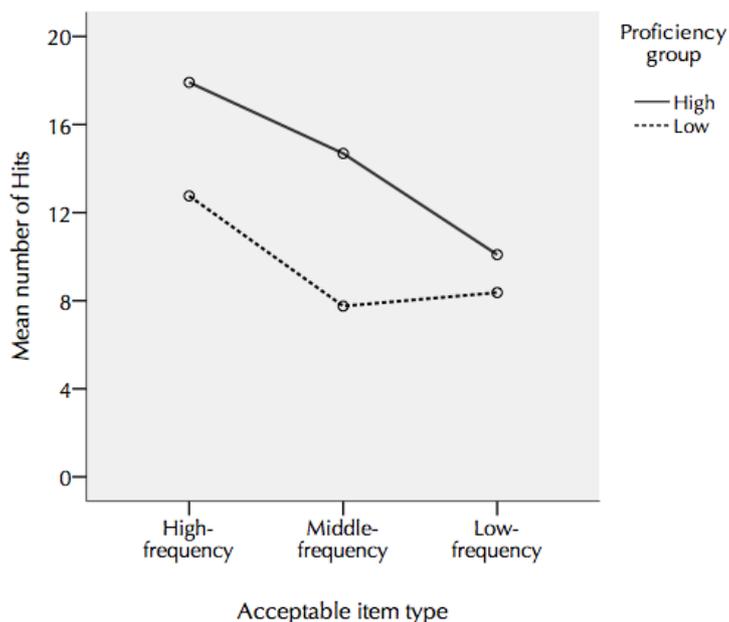
Figure 8.2
Mean number of 'Hits' to the acceptable items by proficiency group.

This difference between the two proficiency groups offers a way to quantify the effect of intra-word frequency. There was no significant difference between the two groups for the results to the low-frequency items. This result might indicate a lower limit to the effect of frequency. It is unlikely that the learners from either proficiency group had encountered the low-frequency senses many times, if at all. It is likely that both groups were equally unfamiliar with the low-frequency senses, and so the results to these items for the two groups were largely comparable. However, there was a significant difference between the two groups in their results to the middle-frequency items. The high-proficiency group had better knowledge of these senses than the low-frequency senses, which likely means these learners had encountered middle-frequency senses often enough to find them acceptable. In contrast, the low-proficiency group knew the middle-frequency items no better than they knew the low-frequency items, even though the middle-frequency senses were more frequent in the language. According to the hypothesis of intra-word frequency, this result indicates that the low-proficiency group had likely not encountered these senses

198

frequently enough. In other words, the effect of intra-word frequency was minimal on the results to the middle and low-frequency items for the low-proficiency group.

It should be pointed out that the measurement of intra-word frequency is only relative between the senses of a single word. It could be that a middle-frequency sense occurs less frequently in a corpus than a low-frequency sense. However, the results represent 20 middle-frequency senses compared to 20 low-frequency senses of word forms with comparable corpus frequency. Based on this number of senses, it is likely that the middle-frequency senses generally occur more frequently in a corpus than the low-frequency senses. However, this conclusion is only tentative without stronger corpus data.

There is another reason to explain the responses of the low-proficiency group. As mentioned above, the results may also be because the participants in the low-proficiency group were simply weaker language learners. However, there is little reliable evidence supporting either the position of lower corpus frequency or the position of weaker language learners. Neither do we know how many times a participant encountered a given word, nor do we know how many times a specific participant would need to encounter a new sense to learn its meaning.

8.4.3 The hypothesis of semantic similarity

The other factor that was controlled for in the experimental design is semantic similarity to the core sense. The hypothesis of semantic similarity is that an L2 learner will identify a polysemous sense as an acceptable use of the word if it bears close similarity to the core sense. The logical distractors were designed to measure the influence of semantic similarity. The results for the logical distractors can be taken as an indication of the effect of semantic similarity on the participants' responses. Since the senses of logical distractors were not attested in the corpora I used, they were considered to be independent of the effect of intra-word frequency. Using the results of the logical distractors, it might be possible to measure the effect of semantic similarity on the results of the acceptable items.

As proficiency gets higher, one would expect that the learners' responses would move towards native speaker norms. This turned out to be the case, not only for acceptable items but also for the distractors. It seems that as their proficiency (based on vocabulary size) increased, the learners were better able to recognise when words are used incorrectly. This was true for both types of distractors: the difference in the mean number of 'False Alarms' between the two proficiency groups was significant, and the interaction between the two groups was consistent. Figure 8.3 compares the means of the two proficiency groups by the mean number of 'False Alarms' to the distractor items.
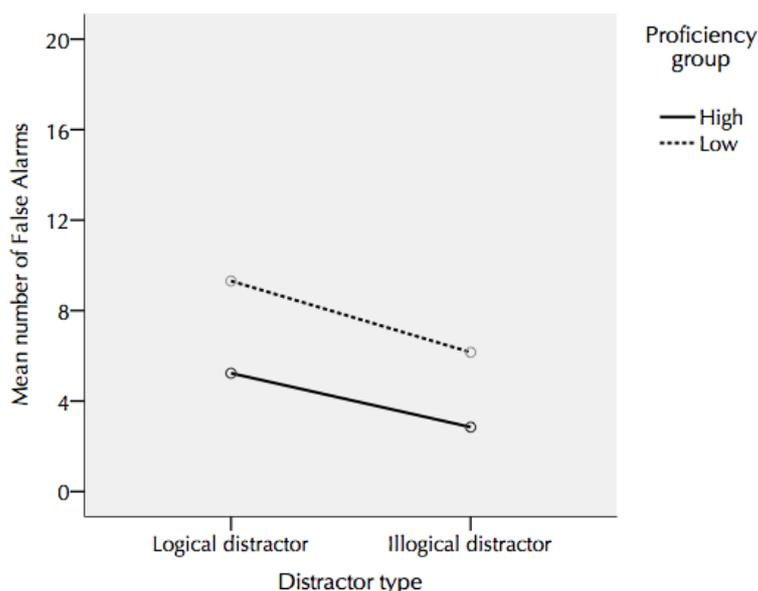
Figure 8.3
Mean number of 'False Alarms' to the distractor items by proficiency group.

When the two proficiency groups are compared, we see that the high-proficiency group committed fewer 'False Alarms' to both types of distractors. When the two groups are considered jointly and the distractor items are compared, we see that both groups responded with more 'False Alarms' to the logical distractors than the illogical distractors. It is more challenging to explain why participants correctly rejected a distractor item than it is to explain why they responded with a 'False

Alarm'. One can argue that the participants responded with more 'False Alarms' to the logical distractors because these items were designed to bear semantic similarity to the core sense. Thus, semantic similarity acts as positive evidence for the results of logical distractors in general. It is not so straightforward to explain why the high-proficiency participants correctly rejected the distractor items more often. The specific reasons might be different for each item. If we were to generalise, it might be best to say that greater breadth of vocabulary knowledge, as measured by the higher VLT scores, also extends to greater depth of vocabulary knowledge, as indicated by the fewer number of 'False Alarms'. This conclusion is supported by Milton's claim that tests of breadth of knowledge, such as the VLT, are good indicators of general proficiency despite only measuring one aspect of language knowledge, breadth of vocabulary (Milton, 2009).

8.4.4 Comparison of the acceptable items

Up to this point the results from the acceptable items and the distractor items have been analysed separately. This was done because the item types are qualitatively different. The senses of the acceptable items were attested in the corpora I consulted, while the senses of the distractors were not. However, a case can be made for comparing the results of the logical distractors to those of the acceptable items. It is likely that the participants had not encountered many of the low-frequency senses. This means that the low-frequency items and the distractor items could be measuring the same factor. To address this possibility, a sub-analysis was carried out, whereby the results of the four item-types were compared to one another. The variable of proficiency was also factored into the analysis, since it has been proposed that the high-proficiency group might have had more L2 exposure than the low-proficiency group.

In order to conduct this analysis, the results of the logical distractors needed to be adjusted downwards. The adjustment equation presented in Chapter 7 was adapted by substituting the 'Hits' for the acceptable items with the 'False Alarms' for the logical distractors:

$$P(k) = \frac{P(log.fa) - P(il.fa)}{1 - P(il.fa)}$$

The results of logical distractors were adjusted downward by participant. In several cases the adjusted results were less than or equal to 0. These participants were removed from the analysis. After the adjustment, there remained 9 participants in the low-proficiency group and 9 participants in the high-proficiency group.

The results of the 'Hits' for the acceptable items and the 'False Alarms' for the logical distractors were analysed by a 4 x 2 analysis of variance for mixed design. Item type (logical distractors, and high, middle and low-frequency items) was set as the within-subjects factor and proficiency (high and low proficiency) was set as the between-subjects factor. Levene's test homogeneity based on the median indicated that the condition of variance was satisfied for all four item types: high frequency $F(1,16) = 2.16$, $p = .16$, middle frequency $F(1,16) = 0.26$, $p = .61$, low frequency $F(1,16) = 1.55$, $p = .23$ and logical distractor $F(1,16) = 0.64$, $p = .44$. Mauchly's test indicated that the assumption of sphericity had not been violated for the within-subjects comparison of acceptable items, $\chi^2(5) = 4.30$, $p = .51$. The main effects of item type $F(3,48) = 62.35$, $p < .001$, $\eta p^2 = .796$, and proficiency $F(1,16) = 7.67$, $p = .014$, $\eta p^2 = .324$, were qualified by an interaction between item type and proficiency $F(3,48) = 11.61$, $p < .001$, $\eta p^2 = .421$.

In order to conduct pairwise comparisons, a Shapiro-Wilk test of normality was conducted on the interaction between item type and proficiency group. The test showed that the distribution of the results for each level in the 4 x 2 design were normally distributed. The results are presented in Table 8.6 below.

Table 8.6
Results to Shapiro-Wilk test of normality for mean 'Hits' or 'False Alarms' to each item type by proficiency group (n = 9).

Item type	Proficiency group	W	p
High frequency	Low	.906	.292
	High	.912	.328
Middle frequency	Low	.964	.836
	High	.876	.142
Low frequency	Low	.975	.932
	High	.924	.431
Logical distractor	Low	.930	.483
	High	.927	.457

In this analysis, the focus of interest was on whether participants responded differently to the item types within the proficiency groups. For this reason, pairwise comparisons using a Bonferroni correction between the item types were conducted separately for each proficiency group. For the high-proficiency group, each item type was significantly different from the others at $p < .05$. This indicates that this group of learners was responding to each item type differently. For the low-proficiency group, only the mean number of 'Hits' to the high-frequency items were significantly different from the 'Hits'/'False Alarms' to the other item-types, $p < .05$. For the low-proficiency group, there was no significant difference between the mean number of 'Hits'/'False Alarms' for the middle-frequency, low-frequency or logical-distractor item types, $p > .05$. Figure 8.4 illustrates the results of the responses to the four item types by proficiency group.
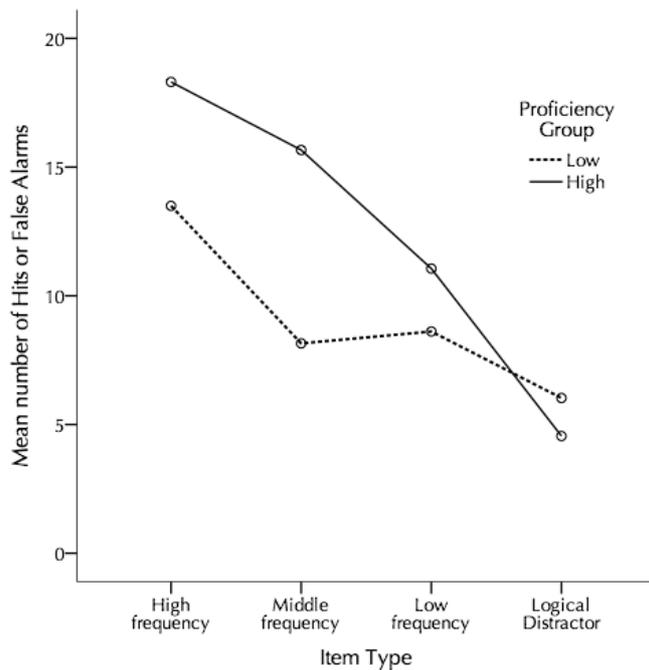
Figure 8.4
Mean number of 'Hits' or 'False Alarms' to each item type by proficiency group (n = 9)

The analysis indicates that the low-proficiency group not only responded to the low-frequency items as they did to the logical distractors, but they also responded to the middle-frequency items in the same way. The logical distractors were designed to be understood using only creative interpretation strategies. The results indicate that the participants in the low-proficiency group are using these same creative strategies when responding to the low and middle-frequency items. In contrast, the participants in the high-proficiency group responded to each item type differently, with a significantly higher number of 'Hits' depending on the frequency of the item type. However, participants in the high-proficiency group responded to the logical distractors in much the same way as participants in the low-proficiency group did. For this reason, creative interpretation was an influence for the high-proficiency group as well. Their stronger results to the other item types indicate that they had increasingly better knowledge of the acceptable items as the target word senses became more frequent in the corpus.

The suggestion has been put forward in this discussion that creative interpretation is one factor influencing how the participants judge the acceptability of the target

204

words. The results suggest that this influence plays a stronger role for participants in the low-proficiency group than those in the high-proficiency group. However, regardless of the strength of this influence, creative interpretation is probably an influence for all participants to differing degrees. In the analysis so far, creative interpretation has been operationalised through the logical distractors, which were designed to bear some similarity to the core sense of the target word. While semantic similarity is not the only factor involved in creative interpretation, it is a prevalent factor when investigating polysemy. For this reason, it is important to look at semantic similarity more carefully, because the strength of its influence remains largely unquantified.

It is likely that the influence of semantic similarity varies between stronger and weaker similarity to the core sense. For example, in the TPM, the target word *air* is used acceptably in three different ways: 'fresh air', 'air travel' and 'putting on airs'. It is apparent that each sense of *air* is quite different from the others with stronger or weaker similarity between them. However, to discuss the influence of semantic similarity to the core sense, we would first need a method to identify what the core sense is and then another method for categorising the similarity to the core sense. We could expect that such a classification would identify 'fresh air' as being consistent with the core sense, 'air travel' as being a figurative use of the core sense, and 'putting on airs' as being inconsistent with the core sense. With such classification conducted systematically for all items on the TPM, the results can then be reanalysed according to the influence of semantic similarity. This will be the focus of attention in the next chapter.

Chapter 9: The effect of similarity to the core sense

One of the overarching research questions in this thesis asked whether semantic similarity was a factor in how L2 learners develop their knowledge of polysemous senses. Up to this point, the effect of semantic similarity has been addressed by analyzing the responses to the logical distractors. In this chapter, the effect of semantic similarity will be addressed by analyzing the responses to the acceptable items. To achieve this goal the acceptable items are recategorized according to their similarity to a proposed core sense. This analysis is informed by methods used in the field of cognitive linguistics. It is predicted that the senses bearing greater similarity to the core sense will be judged as acceptable more often. One issue in conducting this analysis is that the polysemous senses need to be controlled for the effect of intra-word frequency. With that controlled analysis in place, the chapter concludes with a discussion of the relationship between frequency and semantic similarity for learning L2 polysemous senses.

9.1 Introduction

In the previous chapters I presented the results of Arabic learners of English for the Test of Polysemous Meanings (TPM). A main part of the experimental design of the test was based on testing participants' capacity to recognise meanings according to their relative frequency. For example, the target word *air* was used in the following items:

(1)   Let's go outside and get some fresh *air*. (highest frequency)

(2)   *Air* travel was growing rapidly. (middle frequency)

(3)   Trudy is always putting on *airs*. (lowest frequency)

In Chapter 7, I investigated whether the participants would accept the high frequency 'fresh air′ sense (1) more readily than the middle frequency 'air travel' sense (2), with the lowest frequency 'putting on airs′ sense (3) least likely to be accepted. Generally, the Arabic learners found the senses more acceptable as the senses expressed a higher intra-word frequency. In Chapter 8, this tendency was

most clearly expressed by those participants who scored the highest on the Vocabulary Levels Test, a test of vocabulary size.

Another variable interferes with these judgements: the distance of the sense from some semantic point of reference, which is specifically identified as the core sense. It is not clear from the analysis so far how semantic similarity to the core sense has affected the learners' responses to the items categorised by intra-word frequency. The 'fresh air' sense in (1) seems like the most central meaning of the word *air*, while the 'air travel' sense in (2) seems more semantically distant from this meaning, and the 'putting on airs' sense in (3) seems unrelated to either meaning. It is unclear to what degree the learners' responses were affected by intra-word frequency in contrast to semantic similarity.

In the design of the TPM, the effect of semantic similarity was addressed through the learners' responses to the logical and illogical distractors. However, these distractor items were designed to be unacceptable to native speakers in order to be independent of the effect of intra-word frequency. For this reason, the results for the distractor items cannot account for the interplay between semantic similarity to the core sense and intra-word frequency. The question remains about how semantic similarity to a core sense affects the learners' judgements to the acceptable items. Can intra-word frequency alone account entirely for their judgements or does semantic similarity play a substantial role?

In this chapter, the different senses of each stimulus item used in the experiments are categorised for this distance, to see if it is this variable of semantic similarity that most convincingly accounts for the patterns of responses. The first part of this chapter will explain the categorisation procedure used to identify the semantic relationship of the TPM senses to the core sense. The second part of the chapter uses this categorisation to investigate the effect of semantic similarity on the learners' responses to the TPM. In order to compare senses by semantic similarity, the senses should be balanced for frequency. This means that senses expressing different semantic similarity to the core sense can only be compared if they are from the same

intra-word frequency category. The following research question is investigated in this chapter:

RQ   *When the results are controlled for intra-word frequency, do the L2 learners respond with more 'Hits' to polysemous senses that bear more semantic similarity to the core sense than they do to senses that are less similar?*

9.1.1 Categorisation by semantic similarity

In Section 2.3 (on page 18), I examined the contributions that previous research has made to a basic framework used to define the senses of a polysemous word as a coherent semantic category. That framework will be used to organise the acceptable items into four different semantic categories based on their similarity to a proposed core sense: consistent with the core sense, a metaphor extension, and a metonym extension and a meaning that is inconsistent with the core sense.

9.1.1.1 The core sense

In Section 2.3.3, I discussed the guidelines for identifying the core sense. Following Stevenson (2010, p. xi), the core sense was defined as the typical, central use of a word for modern speakers. While the core sense will often be the most frequent sense in a corpus and the oldest sense etymologically, these characteristics are not considered necessary to their identification. What is essential is that the core sense provides coherence to the whole category of senses in that there are clear semantic relations between the senses and the central use (Geeraerts, 2007; Verspoor & Lowie, 2003). These clear semantic relations are what define the polysemous word as a radial category, rather than as a collection of discrete unrelated meanings with a shared word form.

The Oxford Dictionary of English (ODoE) is a suitable reference for identifying the core sense for each target word because it organises its entries by core senses[23]. One concern with this dictionary is that the ODoE takes a narrower view of the core sense

---

23. Not all dictionaries organise their entries according to the core sense. The Longman Dictionary of Contemporary English lists senses according to intra-word frequency and the Oxford English Dictionary lists senses according to historical etymology.

than the one taken in this thesis. This dictionary provides multiple core senses for a polysemous word, for which only one would be identified according to the principles of the radial category structure. In the introduction to the ODoE, the word *belt* is used as an example of a word with two 'core' senses:

(4)   a strip of leather or other material worn, typically round the waist, to support or hold in clothes or to carry weapons

(5)   a strip of material used in various technical applications

The primary difference between these two senses is the domain where the sense is applied. As discussed in Section 2.3.1, "domain" is understood here as the context in which the sense is experienced. The domain of the first sense (4) can be understood as CLOTHING, while the domain of the second (5) can be understood as MACHINES. The fact that the two senses express different domains indicates that they are clearly distinct senses. However, the meanings two senses are not inconsistent with each other because they are linked by "the strip of material". This linking provides a clear semantic relation between the two senses and unifies them as a single semantic category. Each sense shares equally in the common semantic relationship; however, it is reasonable to identify the CLOTHING sense as the core sense and the MACHINE sense as an extension of the core sense. The CLOTHING sense is common to more people since 'clothing belts' are more familiar than 'machine belts'. There are a number of other reasons why this sense should be considered the central use to modern speakers, for example, age of acquisition, corpus frequency, etymology, meaning dominance and association with the human body. However, beyond providing semantic coherence, no other property provides a necessary condition for the core sense.

While the editors do not state this explicitly, it appears that the first entry in the ODoE is usually the most central sense. For example, the CLOTHING sense is listed first in the *belt* entry. As a working principle, the first sense in the ODoE will be initially taken as the proposed core sense when classifying the target words on the TPM. Then the first sense will be confirmed as the core sense according to the criteria outlined in Geeraerts' definition: that it is the 'logical' central use of the word, that it

provides coherence to the whole category of senses, and that there are clear semantic relations between the senses and the central use (Geeraerts, 2007; Verspoor & Lowie, 2003).

9.1.1.2 Metaphor

Two types of semantic extensions from the core sense, metaphor and metonymy, will be used to classify the senses of the TPM. I'll address metaphor first and then return to metonymy.

In the discussion of *belt*, I noted that the MACHINE sense, while distinct, was not inconsistent with the CLOTHING sense. In this analysis, the MACHINE sense is understood as a metaphor extension of the CLOTHING sense. The two senses share the feature of a "strip of material," but in the second sense that feature was construed against a different domain. According to Croft and Cruse, a metaphor "involves a relationship between a source domain, the source of the literal meaning of the metaphorical expression, and a target domain, the domain of the experience actually being described by the metaphor," (Croft & Cruse, 2004, p. 195).

In the *belt* example, the two senses share the feature of a "strip of material". However, the shared similarity does not have to be as tangible as that. The third sense of *belt* listed in the ODoE is more abstract but still related:

(6)   a strip or encircling area that is different in nature or composition from its surroundings: *the belt of trees*

In the third sense, the similarity is now limited to only a "strip" of something. According to the cognitive linguistic theory, these three senses share an image schema. Conventionally, the shared "strip" could be understood as the image schema of a PATH[24]. As discussed in Section 2.3, image schemas are understood as basic concepts that arise out of our fundamental bodily experience of the world (Lakoff, 1987). As such, image schemas are considered to be more language

---

24. This particular image schema is commonplace in the literature (Clausner & Croft, 1999).

independent than other aspects of semantic knowledge[25]. A list of image schemas taken from Clausner and Croft (1999, p. 15) was presented in Section 2.3. It is presented again here for reference to explain how some of the TPM senses were classified as metaphors:

Table 9.1
An inventory of image schemas, collected by Clausner and Croft (1999)[26]

SPACE	UP-DOWN, FRONT-BACK, LEFT-RIGHT, NEAR-FAR, CENTER-PERIPHERY, CONTACT
SCALE	SOURCE-PATH-GOAL
CONTAINER	CONTAINMENT, IN-OUT, SURFACE, FULL-EMPTY, CONTENT
FORCE	BALANCE, COUNTERFORCE, COMPULSION, RESTRAINT, ENABLEMENT, BLOCKAGE, DIVERSION, ATTRACTION
UNITY/ MULTIPLICITY	MERGING, COLLECTION, SPLITTING, ITERATION, PART-WHOLE, MASS-COUNT, LINK
IDENTITY	MATCHING, SUPERIMPOSITION
EXISTENCE	REMOVAL, BOUNDED SPACE, CYCLE, OBJECT, PROCESS

A *belt* [CLOTHING] and a *belt* [TREES] are related by a common abstract shape of a two dimensional length between two terminal points. This abstract shape forms the PATH image schema, which, because of its basic conceptual natural, is proposed to be a generally inferable correspondence between the two senses. The PATH schema is used to construe the meaning of *belt* commonly enough to become conventionalised and listed as an entry in the dictionary. Language users can also make unconventional use of image schemas to construe the meaning of a word. For example, a *belt* might be what separates two other phenomena, such as two fields

---

25. Image schemas are probably not entirely independent of language. Bowerman and Choi (2001) found some evidence that Korean L1 children use different image schemas from English L1 children in their description of certain simple physical phenomena. However, image schemas are based on basic experiences of the physical world, likely making their use inferable across languages.

26. Clausner and Croft compiled the list from the previous research of Johnson (1987) and Lakoff and Turner (1989).

divided by "a belt of trees", in which case the BLOCKAGE schema is used. The BLOCKAGE sense may or may not be included as a dictionary entry, depending on how commonly it is used. Nevertheless, the example shows how image schemas provide a tool for describing the different ways in which a word is construed.

In Lakoff's approach to polysemy, he proposed that extensions from the core sense to extended senses are motivated by two models of categorisation: propositional knowledge and knowledge relating to image schemas. Propositional knowledge is more encyclopaedic, while knowledge of image schemas is informed by bodily experience. These two types of knowledge provide the correspondence between the core sense and the metaphoric senses. For the purpose of categorising the TPM senses, I will place more emphasis on image schema models to make semantic connections and less emphasis on propositional models. The reason for this is because propositional knowledge can be influenced by cultural knowledge which may not be shared with the Arabic participants who took the test. In contrast, image schemas are rooted in basic bodily experience and are less influenced by one's cultural background. For this reason, categorisation using image schemas can be expected to be accessible to the language learners from other cultures. When propositional knowledge is used to define a metaphor, I will note this explicitly and I'll also explain why I think the extension might reasonably be inferable by L2 learners.

9.1.1.3 Metonymy

Metonymy is a different type of semantic extension from metaphor. While in metaphor, the core sense and the extended sense are construed in different domains, in metonymy, the core and extended sense are associated within the same domain of experience. For example, the core sense of *face* refers to the domain of a BODY. A metonymic extension of this core sense would be:

(7)   I noticed several new faces tonight.

Here the domain is still a BODY, but *face* now refers to the whole person instead of just one part. This relationship is considered to be an intrinsic extension because the

core sense and the extended sense refer to different features of the same object. Croft and Cruse provide a list of possible intrinsic associations that support metonymy (2004, pp. 216-217). Two of these associations were relevant to the analysis of the TPM senses.

Part-whole
>
> Part for whole: *I noticed several new faces tonight.*
> Whole for part: *Do you need to use the bathroom?*

Entity-attribute
>
> Entity for attribute: *Shares took a tumble yesterday.*
> Attribute for entity: *He's a size ten.*

A metonym can also express an extrinsic association with the core sense. Extrinsic associations are where "A and B are associated contingently and non-inherently," (2004, p. 217). An example of an extrinsic association would be the use of *arm* in the following sentence:

(8)   Please don't sit on the arm of the chair.

The ARM of the chair and the ARM of a person occupy the same physical space and indeed the one is made for the other. In this way, the two concepts share the same domain of experience but refer to distinct construals.

9.1.1.4 The inconsistent meaning

According to the research from cognitive linguistics, a polysemous word can be described as a radial category because its senses are either consistent with the core sense or they are a metonym or metaphor extension. The fourth category deployed here is the 'inconsistent meaning'. The main justification for this category is that if the sense of the target word doesn't bear similarity to the core sense as a metonym, a metaphor or as a consistent sense, then it lies outside of the normal understanding of the polysemous word as a radial category. The meaning of a word is inconsistent with the core sense because none of the classification criteria is satisfied. First, the domain of the inconsistent meaning is distinct from the domain of the core sense. Second, there is no correspondence that provides consistency between this sense and

the core sense. That correspondence could either be an image schema or propositional content. The inconsistent meaning is comparable to the homonym category that Klein and Murphy (2001) used in their study. However, Klein and Murphy identified homonyms based on diachronic criteria. In their study, two senses were said to be homonyms if they were listed as separate entries in the OED, signifying that the two senses have different etymological origins. The weakness with the diachronic approach is that there might be an etymological connection between senses that would not be readily apparent to someone learning the language.

In contrast to homonyms, the inconsistent meaning is classified according to a more synchronic approach. The rationale behind this classification is that if a learner only knew the core sense, she would not be able to infer the meaning of the inconsistent meaning because no similarities are provided by a shared domain or a consistent image schema. For example, the sense of *board* to mean "room and board" refers to the food served at an inn. This sense is etymologically related to the core sense of *board* [BUILDING] and so it is technically not a homonym. However, the two senses are construed against different domains and there is no shared image schema between them. For this reason the sense of *board* [FOOD] would be classified as inconsistent with the core sense.

9.1.2 Order of acceptability

The semantic similarity hypothesis, introduced in Chapter 5, proposes that learners extrapolate their knowledge of polysemous words from one sense to another. One implication of this hypothesis is that the more central senses of the polysemous word will be more acceptable to the L2 learners and the more peripheral senses will be less acceptable. This implication is important because it offers a way to test whether or not there is support for the semantic similarity hypothesis. The polysemous senses can be organised into semantic categories from most central to most peripheral. The most central senses are proposed to be the most acceptable to L2 learners and the most peripheral senses are proposed to be the least acceptable. The TPM asks learners to judge whether a polysemous word is used acceptably or not. If the results

of the TPM support this order of acceptability, then these results can also be taken as support for the semantic similarity hypothesis. The proposed order of acceptability is as follows: first, senses that are consistent with the core sense; second, senses that express a metonym extension of the core sense; third, senses that express a metaphor extension; and finally, the meanings that are inconsistent with the core sense. The rationale behind this order is explained below.

The senses which are consistent with the core sense are expected to be the most acceptable to the L2 learners. While there isn't general research on what makes a word acceptable or not, there is research on learnability. Learnability can act as a proxy for acceptability for L2 learners, based on the assumption that the easily learned sense will be judged more acceptable. Laufer (1997) discussed how certain semantic properties of a word can affect word learnability. Several of these properties apply to the specific senses of the word, namely abstractness, specificity and idiomaticity. First, more concrete meanings are easier to learn. This property supports the learnability of the consistent sense on a case by case basis. The core senses of some words like *hand* are highly concrete but others like *order* or *position* are less so. Second, specificity refers to how well a sense can be generalised across different contexts. More easily generalised senses are considered easier to learn. The core sense is more easily generalised because it is often the default construal. For example, the non-sequitur, "Put a cover on it," can be applied to many contexts using the core sense of *cover*. In contrast, "bed covers" refers to a very specific context of BEDS and if the context isn't clear then the modifier "bed" has to be used to indicate the sense. Finally, idiomatic uses of language are considered more difficult to learn than non-idiomatic uses. Kellerman (1986) found that Dutch learners of English were more likely to transfer an L1 idiom to their L2 if the idiom used a core meaning. This indicates that while idioms are difficult to learn, the core sense facilitates their learning. Based on their semantic properties, it is likely that the core senses are easier to learn for L2 learners. This in turn indicates that those TPM items which are consistent with the core sense are likely to be more acceptable to the L2 learners.

Of the two extended senses, the metonym extension is proposed to be more similar to the core sense than the metaphor extension. This proposition is based on the psycholinguistic research of Klepousniotou (2002) and Klepousniotou and Baum (2007). They found that when L1 participants were primed with the core sense, they responded faster to metonyms than metaphors in lexical decision tasks. The results indicated that the core sense shares greater representation with metonym extensions than with metaphor extensions. This research is relevant to the analysis of the TPM results because the design of the study identified both metonyms and metaphors according to a similar framework used to classify the TPM senses. There are some notable differences however. These studies used metonyms and metaphors that follow restricted patterns. For example, the metonyms follow specific relationships, such as the object/substance relationship (*chicken*:ANIMAL vs *chicken*:MEAT); and the metaphors follow conventionalised patterns, such as HUMANS ARE PHYSICAL OBJECTS ("She's a star"). In contrast, the TPM senses show a great deal more variety, having been drawn from the example sentences used in a dictionary and not designed to follow specific patterns or relationships. Furthermore, Klepousniotou's studies investigate how quickly native speakers respond to senses. In contrast, the TPM was designed to measure how acceptable the different polysemous senses were to intermediate learners of English.

The question of whether metonyms are easier to learn than metaphors has also been addressed in developmental research of children learning their L1. Rundblad and Annaz (2010) compared children's understanding of novel metaphors and metonyms. In their study, 24 children, aged 5;3 to 18, were presented with stories where the critical task was to identify what the target metaphor or metonym referred to. For example, in one metaphor story, a *flood* referred to people, whereas in a metonym story, *Robbie Williams* referred to a CD. They found that the children correctly identified the metonym referents significantly more often than the metaphor referents across the developmental trend by chronological age. Rundblad and Annaz argue that metaphors are more cognitively complex due to the mappings between domains and thus more difficult for children to interpret correctly.

However, another study found results which qualify those of Rundblad and Annaz. Van Herwegen, Dimitriou and Rundblad (2013) also conducted a study with L1 children on the comprehension of novel metaphors and metonyms. In their study, there were two respective categories for the metaphors and metonyms. The metaphors were either sensory (e.g., a *marshmallow* referred to a soft pillow) or non-sensory (e.g., a *turtle* referred to a slow car), and the metonyms were either object/user (e.g. the *apron* instead of the cook) or part/whole (e.g., the *moustache* instead of the man with the moustache). The researchers found that there was no significant difference between identification of metaphor compared to that of metonyms. However, while the two categories of metaphors were comparable, the researchers found that object-user metonyms were identified correctly significantly less often than part-whole metonyms. This result indicates, at least for L1 learners, some metonyms are more challenging to process than others. While Van Herwegen et. al. only looked at very restricted types of metonyms, it is notable that extrinsic metonyms were more difficult to the learners than intrinsic metaphors.

In sum, the evidence is contradictory for whether metaphors or metonyms are easier for L2 learners to learn. Three studies provide evidence that metonyms should be easier for L2 learners, and in the one study which found evidence in favour of metaphors, the participants were children learning their L1. Adult L2 learners are likely more cognitively mature than the L1 learners and it is a question how applicable these L1 results are for L2 adults. Thus, L2 learners are proposed to find metonyms more acceptable than metaphors because a metonym shares the same domain of experience as the core sense. In contrast, metaphors express widely different contexts. Klepousniotou proposed that it was according to this difference that native speakers responded faster to metonyms. Given the absence of L2 learner research, this difference between metonyms and metaphors will also serve to predict that metonyms will be more acceptable to L2 learners. As such, the L2 participants are expected to respond to the metonym items with more 'Hits' than the metaphor items.

Finally, the L2 participants are proposed to respond to the inconsistent meanings with the fewest 'Hits'. These senses do not share semantic correspondence with the core sense. For example, the inconsistent meaning for *air* is, "Trudy is always putting on airs." If a learner knows another sense of a *air*, it is unlikely to benefit her when guessing the inconsistent meaning. Unless the learners already know the inconsistent meaning, they are unlikely to find it acceptable.

9.2 Classification

9.2.1 Identification of semantic categories

The acceptable items on the TPM were originally classified according to intra-word frequency. This section will describe the method I used to reclassify the TPM according to their semantic relationship to the core sense. The final product of the classification process is presented in Table 9.2 on page 224.

There are two steps to the classification. In the first step, the core sense is identified for the target word. In the second step, the sense of the target word in the TPM item is classified according to its relationship to the core sense. The categories are a consistent sense, a metonym extension, a metaphor extension, or an inconsistent meaning.

9.2.1.1 Identification of the core sense

The core senses of the target words were identified as the first listed sense in the ODoE. For example, the first listed sense for *air* was, "[mass noun] the invisible gaseous substance surrounding the earth, a mixture mainly of oxygen and nitrogen." This sense was considered to be the most central sense and it provided the most coherence when grouped with the other senses of *air*.

Next, I identified the domain of experience for each core sense. As discussed in Section 2.3.1, the domain is the contextual background against which the concept of the sense is understood and experienced. The identification process can be illustrated by considering the context of *air* in the following sentence.

(9)   These laws are meant to produce cleaner air.[27]

In this sentence, *air* is used in its core sense. It is part of the physical world with both chemical and physical properties. In this case, I identified the domain for *air* by the keyword [ATMOSPHERE], which I felt was general enough to take into account both "fresh air" and "air pressure", but distinct enough to differentiate it from "musical air" or "air of superiority". In Table 9.2, the domain is presented in capital letters and square brackets.

9.2.1.2 The consistent sense

The sense of an acceptable item was classified as a consistent sense if the construal of the target word was adequately described by the core sense. For example, the high-frequency item for *air* is,

(10)   Let's go outside and get some fresh air.

The target word was classified as consistent with the core sense of *air*, "the mixture of gases around the Earth, that we breathe", and the domain was identified as [ATMOSPHERE].

9.2.1.3 The metonym extension

A sense was classified as a metonym extension if it expressed a distinct construal from the core sense but the the domain of experience remained the same. For example, the middle-frequency item for *air* was,

(11)   Air travel was growing rapidly.

The LDOCE definition for this sense is "relating to or involving planes." As in the core sense, the domain of the sense is still [ATMOSPHERE] because airplanes use the same physical material as we do when we breathe "fresh air". However, *air* now refers to the mode of travel using airplanes. The core sense and the metonym

---

27. Taken from Merriam-Webster Dictionaries (2008).

extension use the same domain through an extrinsic association because we associate airplanes with air, in the same way as we associate boats with water.

Along with extrinsic associations, two other types of associations are used to classify metonyms, PART/WHOLE and ENTITY/ATTRIBUTE. Examples of these extensions were presented in Section 9.1.13. In Table 9.2 below, the type of association used to classify a metonym is indicated by an arrow →.

9.2.1.4 The metaphor extension

A sense was classified as a metaphor extension if it expressed a different domain of experience from the core sense but shared a correspondence. Image schemas were used to establish the correspondence between the core sense and the metaphor extension. For example, the core sense of *branch* is "a part of a tree which grows out from the trunk or from a bough" and the domain for the core sense of *branch* was identified as [TREES]. The middle-frequency sense of *branch* is a metaphor extension of the core sense.

(12)   All branches of government are having to reduce spending.

The domain for this sense of *branch* was identified as [GOVERNMENT], which is different from the core sense. There is a correspondence between the core sense and the extension. In the core sense, *branches* have a periphery relationship to the central trunk of the tree. Likewise, a government *branch* has a periphery relationship with the central government body. According to Clausner and Croft's (1999, p. 15) image schema table (see Table 9.1), CENTER-PERIPHERY, is one of the basic image schemas used to understand SPACE. The PERIPHERY component to the image schema allows *branch* to be used metaphorically in the domain of GOVERNMENT. This relationship forms a conceptual metaphor which is expressed as A BRANCH IS A PERIPHERY.

As with the metonym associations, the conceptual metaphors are indicated in Table 9.2 by an arrow →.

221

9.2.1.5 The inconsistent meaning

The final classification was that the meaning of the target word was inconsistent with the core sense. This meant that the target word expressed a different domain of experience from the core sense and there was no correspondence linking the two uses of the word across the different domains. For example, the low-frequency item for *air* was,

(13)   Trudy is always putting on *airs*.

The LDOCE definition for this meaning is "a way of behaving that shows someone thinks they are more important than they really are." The domain of experience for the meaning was identified as [SOCIETY]. When compared to the core sense of *air* [ATMOSPHERE], the domain of experience is different and there is no apparent correspondence between the two senses.

9.2.2 Classification table

Table 9.2 (beginning on page 224) presents the identification of the core sense for each target word and the re-categorisation of each acceptable-item sense according to its semantic similarity. The definitions for the different senses aren't included because they are quite long and it would be difficult to include the 63 definitions comfortably in a single table. Keywords are used in place of the definition. The keywords are taken from the LDOCE, where these terms are referred to as "signposts". These are terms of 1-3 words that the dictionary uses to distinguish one sense from another in longer entries. For example, the definition for the high-frequency sense of *air* is "the mixture of gases around the Earth, that we breathe." The keyword for this sense is GAS. This keyword is presented in the table, in capital letters, in place of the definition.

In the table, the first column presents the target word. Next to the first column, across the top of the other columns, is the proposed core sense. I've presented the core sense according to a keyword followed by the domain in square brackets. For example, the core sense for *air* is presented as follows: "The core sense of air is GAS [ATMOSPHERE]." The second column presents the original categorisation by intra-

word frequency, referred to as High, Middle and Low-frequency senses. The third column presents how the item was re-categorised in relation to the proposed core sense: Consistent, Metonym, Metaphor or Inconsistent. The fourth column also presents the sense as a keyword, followed by its domain in square brackets. The metaphor and metonym classifications include extra information in the fourth column, preceded by an arrow →. The metaphor classifications include the image schema that was identified as the part of the concept which transfers from the source domain of the core sense to the target domain of the metaphoric sense. For example, the middle-frequency sense of *branch* is "OF GOVERNMENT [GOVERNMENT] → A BRANCH IS A PERIPHERY". The metonym classifications include a note to what type of association I made between the core sense and the extended sense. For example, the middle-frequency sense of *air* is "PLANES [ATMOSPHERE] → Extrinsic association". At the end of each classification, I've included the TPM item after an integral symbol ∫.

Table 9.2

Classification of the target words into the semantic categories: consistent with the core sense, a metonym extension, a metaphor extension and inconsistent with the core sense.

air	The core sense of *air* is GAS [ATMOSPHERE]		
	High	Consistent	GAS [ATMOSPHERE] ∫ Let's go outside and get some fresh *air*.
	Middle	Metonym	PLANES [ATMOSPHERE] → Extrinsic association ∫ *Air* travel was growing rapidly.
	Low	Inconsistent	BEHAVIOUR [SOCIETY] ∫ Trudy is always putting on *air*s.

arm	The core sense of *arm* is BODY [BODY]		
	High	Consistent	BODY [BODY] ∫ Tim's mother put her *arms* around him.
	Middle	Inconsistent	WEAPONS [MILITARY] ∫ The growing *arms* trade is a problem for the country.
	Low	Metonym	FURNITURE [FURNITURE] → Extrinsic association ∫ Please don't sit on the *arm* of the chair.

board	The core sense of *board* is IN BUILDING [BUILDING]		
	High	Metaphor	INFORMATION [INFORMATION] → A BOARD IS A RECTANGLE[28]. ∫ I've put a list of names up on the *board*.
	Middle	Inconsistent	GROUP OF PEOPLE [GOVERNANCE] ∫ The *board* of Directors met yesterday.
	Low	Inconsistent	MEALS [LODGING] ∫ In the old-age home, she will have to pay for room and *board*.

28. Basic shapes are used by Lakoff as prototypical examples of image schemas; however, Clausner and Croft's (1999) did not include them in their inventory of image schemas.

body	The core sense of *body* is PEOPLE/ANIMALS [PEOPLE/ANIMALS]		
	High	Consistent	PEOPLE/ANIMALS [PEOPLE/ANIMALS] ∫ For their *body* size, these birds lay very small eggs.
	Middle	Metonym	GROUP [PEOPLE] ⟶ Part for whole ∫ The student *body* numbers 5000.
	Low	Metaphor	'body of something' [KNOWLEDGE] ⟶ A BODY IS A CONTAINER ∫ There is now a large *body* of knowledge about childhood.

branch	The core sense of *branch* is OF A TREE [TREES]		
	High	Consistent	OF A TREE [TREES] ∫ The top *branches* were full of birds.
	Middle	Metaphor	OF GOVERNMENT [GOVERNMENT] ⟶ A BRANCH IS A PERIPHERY ∫ All *branches* of government are having to reduce spending.
	Low	Metaphor	OF A ROAD, RIVER, ETC. [ROADS,RIVERS, ETC] ⟶ A BRANCH IS A PERIPHERY ∫ This branch of the river eventually empties into the Atlantic Ocean.

case	The core sense of *case* is SITUATION [EVENTS]		
	High	Consistent	SITUATION [EVENTS] ∫ In this *case*, several solutions could be tried.
	Middle	Metaphor	LAW/CRIME [LEGAL] ⟶ A CASE IS PART OF A WHOLE. ∫ He told the lawyer that he didn't want a court *case*.
	Low	Metaphor	DISEASE [MEDICAL] ⟶ A CASE IS PART OF A WHOLE. ∫ He had a bad *case* of the flu.

225

character	The core sense of *character* is ALL SOMEBODY'S QUALITIES [PERSON]		
	High	Consistent	ALL SOMEBODY'S QUALITIES [PERSON] ʃ It's not in her *character* to be jealous.
	Middle	Metonym	MORAL STRENGTH [PERSON] → Whole for part ʃ It takes strength of *character* to admit you are wrong.
	Low	Inconsistent	LETTER/SIGN [ORTHOGRAPHY] ʃ I can't read the *characters* on that sign.

class	The core sense of *class* is SAME TYPE OF THING [THINGS]		
	High	Metaphor	SOCIAL GROUP [SOCIAL HIERARCHY] → A CLASS IS A CATEGORY. ʃ She is a member of the working *class*.
	Middle	Metaphor	TEACHING PERIOD [TEACHING] → A CLASS IS A CATEGORY. ʃ English *classes* start at 5:15.
	Low	Inconsistent	STYLE/SKILL [QUALITY] ʃ These flowers will give your garden a touch of *class*.

course	The core sense of *course* is PLANNED DIRECTION [TRAVEL]		
	High	Metaphor	EDUCATION [EDUCATION] → PURPOSES ARE DESTINATIONS. ʃ Andy is doing a one-year business *course*.
	Middle	Consistent	PLANNED DIRECTION [TRAVEL] ʃ The plane changed *course* to avoid the storm.
	Low	Metaphor	MEDICAL TREATMENT [MEDICINE] → PURPOSES ARE DESTINATIONS. ʃ The doctor directed her to take a new *course* of medicine.

226

cover	The core sense of *cover* is PROTECTION [PHYSICAL OBJECTS]		
	High	Consistent	PROTECTION [PHYSICAL OBJECTS] ∫ There is a plastic *cover* over the meal.
	Middle	Metaphor	BED [BED] → COVERS ARE SURFACE OF CONTAINERS ∫ The *covers* had slipped off the bed in the night.
	Low	Metaphor	WEATHER [WEATHER] → COVERS ARE SURFACES OF CONTAINERS. ∫ The cloud *cover* in the morning should clear later.

cut	The core sense of *cut* is HOLE/MARK [PHYSICAL OBJECTS]		
	High	Metaphor	REDUCTION [BUDGETS] → A CUT IS PART OF A WHOLE. ∫ Teachers are expecting further *cuts* next year.
	Middle	Consistent	HOLE/MARK [PHYSICAL OBJECTS] ∫ Make a small *cut* in the paper.
	Low	Metaphor	SHARE OF SOMETHING [MONEY] → A CUT IS PART OF A WHOLE. ∫ She was determined to claim her *cut* of the prize money.

face	The core sense of *face* is FRONT OF YOUR HEAD [BODY]		
	High	Consistent	FRONT OF YOUR HEAD [BODY] → The core sense of face is FRONT OF YOUR HEAD [BODY] ∫ She had a beautiful *face*.
	Middle	Metonym	PERSON [BODY] → Part for whole. ∫ Gordon is a familiar *face* at the local flower show.
	Low	Metaphor	CLOCK [TIME PIECES] → A FACE IS THE FRONT. ∫ He liked the *face* of the watch.

227

form	The core sense of *form* is SHAPE OR CONFIGURATION OF SOMETHING [PHYSICAL OBJECTS]		
	High	Metaphor	TYPE [CATEGORIES OF THINGS] → A SHAPE IS A CATEGORY. ∫ The bicycle is an environment-friendly *form* of transportation.
	Middle	Metonym	THE WAY SOMETHING APPEARS [PHYSICAL OBJECTS] → Attribute for entity. ∫ The medicine comes in a liquid *form*.
	Low	Metaphor	PERFORMANCE [PERFORMANCE] → CONFIGURATION OF THINGS IS CONFIGURATION OF ACTIONS[29]. ∫ He's been in good *form* all this season.

hand	The core sense of *hand* is PART OF BODY [PERSON]		
	High	Consistent	PART OF BODY [PERSON] ∫ She waved her *hand* to the crowd.
	Middle	Metonym	HELP [PERSON] → Entity for attribute. ∫ Can you give me a *hand* to lift this?
	Low	Metonym	WORKER [PERSON] → Part for whole. ∫ The farm *hands* wake up at 5:00 in the morning.

head	The core sense of *head* is TOP OF BODY [PERSON]		
	High	Consistent	TOP OF BODY [PERSON] ∫ He turned his *head* and looked at me.
	Middle	Metonym	CALM/SENSIBLE [PERSON] → Entity for attribute. ∫ I went to sleep early to have a clear *head* for the exam.
	Low	Metaphor	FRONT/LEADING POSITION [TABLE] → HEAD IS UP. ∫ The president sat at the *head* of the table.

---

29. This metaphor is taken from Lakoff (1987). It seems equal in simplicity to the image schemas.

heart	The core sense of heart is *BODY* ORGAN [BODY]			
	High	Consistent	BODY ORGAN [BODY] ∫ Regular exercise is good for the *heart*.	
	Middle	Metaphor	EMOTIONS/LOVE [EMOTIONS] → THE HEART IS A CONTAINER OF EMOTIONS[30]. ∫ Edith loved her boy with all her *heart* and soul.	
	Low	Metaphor	CENTRE OF AN AREA [GEOGRAPHIC AREA] → THE HEART IS IN THE CENTRE. ∫ I have a house in the *heart* of the city.	

line	The core sense of *line* is ON PAPER/ON THE GROUND [PHYSICAL OBJECTS]			
	High	Consistent	ON PAPER/ON THE GROUND [PHYSICAL OBJECTS] ∫ He raced towards the finishing *line*.	
	Middle	Metaphor	WAY OF DOING SOMETHING [THOUGHT] → A LINE IS A PATH. THOUGHT FOLLOWS A PATH. ∫ We were both thinking along the same *lines*.	
	Low	Metaphor	FAMILY [ANCESTRY] → A LINE IS A PATH. LINEAGE HAS A SOURCE, PATH AND GOAL. ∫ She comes from a long *line* of actors.	

order	The core sense of *order* is ARRANGEMENT [THINGS OR EVENTS]			
	High	Metaphor	FOR A PURPOSE [PLANS] → ORDER IS A PATH WITH A GOAL. ∫ Samuel trained every day in *order* to make his presentation better.	
	Middle	Consistent	ARRANGEMENT [THINGS OR EVENTS] ∫ Then they call out our names in *order* and we answer yes or no.	
	Low	Metaphor	SOCIAL/ECONOMIC SITUATION [SOCIETY] → ORDER IS A PATH WITH A GOAL. ∫ The people of South Africa wanted a new social *order*.	

---

30. This heart metaphor is based on propositional knowledge. It is quite common across cultures (Gutiérrez Pérez, 2008).

point	The core sense[31] of *point* is A SINGLE IDEA IN AN EXTENDED WHOLE [IDEAS]		
	High	Consistent	A SINGLE IDEA IN AN EXTENDED WHOLE [IDEAS] ∫ That's a very interesting *point*.
	Middle	Metaphor	PLACE [PLACES] → POINTS ARE PARTS ITERATED UPON A WHOLE. ∫ No cars are allowed beyond this *point*.
	Low	Metaphor	SHARP END [INSTRUMENTS/TOOLS] → A POINT IS PART OF A WHOLE. ∫ Be careful with that needle - it has a very sharp *point*.

position	The core sense of *position* is PLACE WHERE SOMEBODY/SOMETHING IS [GEOGRAPHIC AREA]		
	High	Metaphor	SITUATION [EVENTS] → A POSITION IS PART OF A WHOLE. ∫ Next week we will be in a much better *position* to talk about it.
	Middle	Consistent	PLACE WHERE SOMEBODY/SOMETHING IS [GEOGRAPHIC AREA] ∫ Our hotel was in a central *position* near St Mark's Square.
	Low	Metaphor	JOB [EMPLOYMENT] → A POSITION IS PART OF A WHOLE. ∫ Bill took up his new *position* as Works Director in October.

31. The SHARP END sense is the first listed sense in the ODoE. However, the SINGLE IDEA sense lends greater coherence to the category of point senses.

9.3 The organisation of TPM senses by semantic category and intra-word frequency.

In the original design of the TPM, there were 60 acceptable items classified by three intra-word frequency levels: 20 high frequency items, 20 middle frequency items and 20 low frequency items. These 60 items were subsequently reclassified in the preceding section according to four categories for semantic similarity to the core: consistent with the core sense, a metonym extension, a metaphor extension, and inconsistent with the core sense. Table 9.3 presents the count of the four sense categories grouped by the categories of intra-word frequency. In total, the items categorised into twelve semantic category x intra-word frequency groups.

Table 9.3
Number of occurrences of the acceptable items for each semantic category when grouped by intra-word frequency.

Intra-word frequency	Consistent	Metonym	Metaphor	Inconsistent	Total
High frequency	13	0	7	0	20
Middle frequency	4	7	7	2	20
Low frequency	0	2	14	4	20

The count shows that the items were not equally distributed across the four semantic categories. Thirteen of the items were grouped as both high-frequency items and consistent with the core sense. In contrast, none of the low-frequency items were grouped as consistent with the core sense. This distribution is likely a reflection of how polysemous senses are distributed across the language as a whole. However, the purpose of reclassifying the items by semantic similarity was to compare how L2 learners responded to senses across different semantic categories when controlled for intra-word frequency. Obviously, no comparison can be made for groups where there are no representative items. However, some of the groups contain only a few items. For example, there are only two items grouped as both low frequency and metonyms. It is worthwhile to consider what the minimum number of items per group should be for the purpose of comparing the participants' responses.

In the analysis below, the semantic categories are compared by the adjusted mean number of 'Hits' for each participant. For example, the comparison for the high-frequency senses will be between the mean number of 'Hits' for the thirteen items consistent with the core sense and the seven items categorised as metaphors. Since the comparison is between the mean number of 'Hits' by participant, the disproportionate number of items by group will not affect the sample size of the comparison. However, there is a qualitative concern about whether there are enough senses to be representative for each level of frequency by semantic category. This issue can be addressed with the example of the four low-frequency items categorised as inconsistent with the core sense:

> These flowers will give your garden a touch of *class*.
> I can't read the *characters* on that sign.
> In the old-age home, she will have to pay for room and *board*.
> Trudy is always putting on *airs*.

It is apparent that these are not common uses of the target words; however, there may be reasons which might prompt the learner's response that don't have to do with intra-word frequency or semantic similarity. The learner might be familiar with expression 'touch of class' through advertising, or she might reject 'airs' because it ends in a plural 's'. It is usually hoped that with enough items, the effect of individual differences such as these can be minimised. Unfortunately, there aren't many senses in the inconsistent category at the low frequency level. To reduce the effect of individual differences between senses, I decided to include the two inconsistent items from the middle frequency level:

> The growing *arms* trade is a problem for the country.
> The *Board* of Directors met yesterday.

It is hoped that by increasing the inconsistent group to six items, the items as a whole will be more representative of senses inconsistent with the core sense. The responses to these six items can be compared to the fourteen low-frequency items categorised as metaphors. The prediction still remains that L2 learners would

232

respond to the metaphor extensions with more 'Hits' than the inconsistent items, despite including the two items at the middle frequency level.

Two other comparisons will be made. At the high frequency level, the responses to the thirteen consistent items will be compared to those for the seven metaphor items. The prediction is that the L2 learners would have found the consistent items more acceptable. Finally, at the middle frequency level, the responses for the seven metonym items will be compared to the seven metaphor items. The prediction is that the participants would have found the metonyms more acceptable. The other groups with four or fewer items are not included in the analysis.

9.4 Analysis

The semantic categories were compared to one another using the data previously presented in Chapter 7. This data was collected from 67 Arabic learners of English studying in an intensive English program at a technical college in Qatar. The participants' responses to the TPM were analysed for the effect of intra-word frequency in Chapter 7 and for the effect of proficiency in Chapter 8.

In order to compare the participants' responses by semantic category, the following procedure was followed to calculate the mean number of 'Hits' by participant. With reference to Table 9.3 above, three comparisons were made. The first comparison was between the responses to the thirteen consistent items and seven metaphor items when controlled for the high frequency level[32]. The mean number of 'Hits' for these groups of items were adjusted downwards based on the number of 'False Alarms' the participant made to the illogical distractors. Participants were excluded from the analysis if their adjusted mean number of 'Hits' was zero or a negative number. As the comparisons were for repeated measures, the participants were excluded across both groups of items in the comparison. Of the original 67

---

32. One difficulty with this comparison is that the frequency measures are intra-word and do not take into account the differences in frequency between senses of different words. This means a low-frequency sense of one word might have the same occurrence in a corpus as the high-frequency sense of another word. This issue is discussed at more length in Section 9.5.1.

participants, three participants were removed from the comparison between the consistent items and the metaphor items for the high-frequency category.

The second comparison was between responses for the seven metonym items and the seven metaphor items when controlled for the middle frequency level. Again, the mean number of 'Hits' was adjusted downwards and four participants were removed from the comparison because their proportion of 'Hits', when adjusted downwards, was zero or less.

The third comparison was between responses for the fourteen metaphor items at the low frequency level and the six inconsistent items at the combined middle and low frequency levels. When the mean number of 'Hits' for these items was adjusted downwards, the proportion of 'Hits' for 20 participants was zero or less. Their results were removed from the comparison leaving behind the results of 47 participants. The adjusted results for all three comparisons are presented in Figure 9.1.
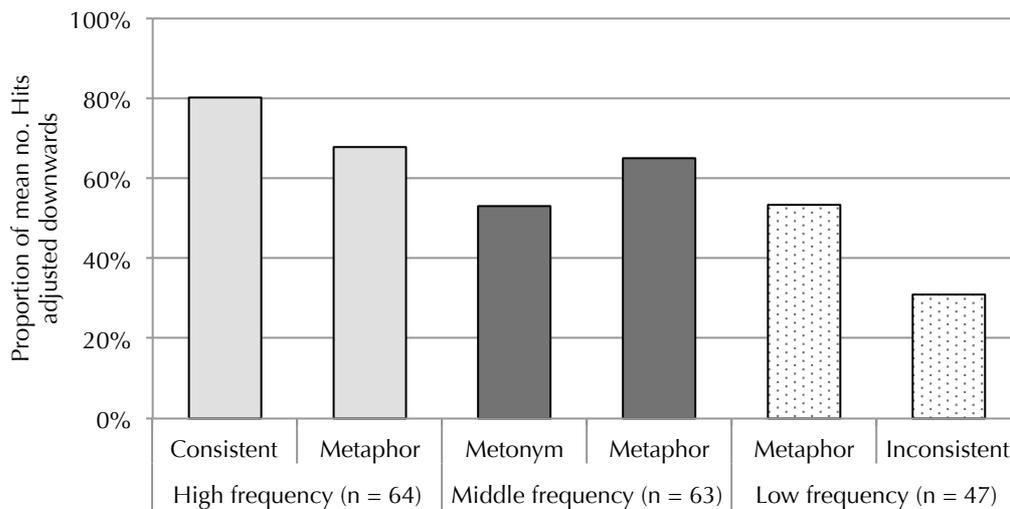
Figure 9.1
Mean number of 'Hits' to the semantic categories when controlled by intra-word frequency. The results were adjusted downwards and presented as proportions.[33]

---

33. The proportions for the metaphor categories look similar across frequency categories. This is partially an effect of the lower number of participants in the low-frequency category. A comparison of the metaphor items between frequency categories is presented in the discussion.

The Shapiro-Wilk test of normality was conducted for the mean number of 'Hits' for the four groups of items. The results are presented in Table 9.4 for the normality of 'Hits' for these four groups of items.

Table 9.4
Shapiro-Wilk Test of Normality for the mean number of 'Hits' to semantic categories when controlled by intra-word frequency

Intra-word frequency	Relation to the core sense	n	Statistic	df	Sig.
High-frequency items	Consistent	64	.82	64	< .001
	Metaphor	64	.95	64	.009
Middle-frequency items	Metonym	63	.97	63	.161
	Metaphor	63	.95	63	.019
Low-frequency items	Metaphor	47	.94	47	.025
	Inconsistent	47	.90	47	.001

For all the comparisons, there was at least one group of results where the distribution of 'Hits' by participant was significantly different from the null hypothesis of normality at $p < .05$. For this reason non-parametric tests will be used for all the comparisons. The results were significantly skewed for three of the groups: the high consistent items, *skew* = -1.60, z = -5.355, $p < .05$, the high metaphor items, s*kew* = -0.51, z = -1.71, $p < .05$, and the low inconsistent items, *skew* = -.90, z = -2.59, $p < .05$. A sign test was conducted with comparisons for these groups of items and a Wilcoxon signed-rank test was conducted with the comparison between the groups of middle frequency items.

In the first comparison, the differences between the consistent sense items and the metaphor items were controlled for by comparing only items within the high-frequency category. The mean proportion of 'Hits' for the items consistent with the core sense (*M* = .80, *Mdn* = .87, *SD* = .22) was significantly greater than the items expressing metaphor extension of the core sense (*M* = .68, *Mdn* = .71, *SD* = .21), z = 2.92, r = .37, p = .003. In the second comparison, the differences between the metonym items and the metaphor items were controlled by the middle-frequency

category. Against expectation, the mean proportion of 'Hits' for the items expressing a metonym extension of the core sense ($M = .53$, $Mdn = .50$, $SD = .25$) was significantly lower than the items expressing metaphor extension of the core sense ($M = .65$, $Mdn = .68$, $SD = .25$), $z = 2.86$, $r = .36$, $p = .004$. In the third comparison, the mean number of 'Hits' for the metaphor items at the low frequency level were compared to the mean number of 'Hits' for the inconsistent items at the low and middle-frequency levels. The mean proportion of 'Hits' for the items expressing a metaphor extension of the core sense ($M = .53$, $Mdn = .49$, $SD = .20$) was significantly greater than the items expressing a meaning inconsistent with the core sense ($M = .31$, $Mdn = .22$, $SD = .25$), $z = 3.28$, $r = .48$, $p = .001$.

The results supported the prediction that the learners would respond with the most 'Hits' to the items consistent with the core sense, and with the fewest 'Hits' to the items where the target word expressed an inconsistent meaning. However, the analysis also found that the learners responded with fewer 'Hits' to the metonym items than the metaphor items. This finding was the opposite of what was predicted based on previous research from Klepousniotou (2002) and Klepousniotou and Baum (2007). In what follows, the reasons for these results are explored.

9.5 Discussion

The analysis supported the prediction that L2 participants would judge the items most acceptable when the target words were consistent with the core sense and least acceptable when they were inconsistent with the core sense. However, against expectation, the participants found the items more acceptable when the target words were categorised as metaphors rather than as metonyms. It's worthwhile to consider how the analysis both qualifies and supports the semantic similarity hypothesis and also to consider what can be said about the relationship between semantic similarity and intra-word frequency.

9.5.1 Relationship between semantic similarity and intra-word frequency

In the analysis of semantic similarity, the acceptable items were grouped by frequency category in an attempt to control for the effect of intra-word frequency.

While it appears that semantic similarity is a definite factor in the learners' acceptability judgements, there are concerns with how well intra-word frequency can be controlled. The ideal situation would be where there was data on the frequency of occurrence for each polysemous sense in a large corpus. With this data it would be possible to plot a linear regression of the frequency of the polysemous sense against the number of 'Hits' for the item using that sense. Unfortunately, to the best of my knowledge, this frequency data isn't available. This is in part because the process cannot yet be automated but requires a lexicographer to identify the sense of each instance of the polysemous word. The Longman Dictionary of Contemporary English (LDOCE) was used in the creation of the TPM. This dictionary was created using the Longman Corpus Network of 330 million words. The LDOCE lists its senses by intra-word frequency, which indicates that there is some intra-word frequency count of senses to the Longman Corpus; however, that data is not publicly available. Since a linear regression by intra-word frequency was not accomplished, the TPM senses were categorised according to a procedure of relative frequency. One main limitation of this procedure is the variation in frequency counts of the different senses within the same frequency category.

The other option for dealing with frequency is to employ the approach taken in psycholinguistic studies with native speakers. In these studies, researchers use two measures to deal with frequency. First they match word forms for frequency in a corpus. Then they match senses based on meaning-dominance norms, which are the normative results of how often native speakers produce a certain meaning when asked, "what does this word mean?" This didn't seem like a viable approach for intermediate L2 learners. Native speaker norms of meaning dominance would not be applicable to L2 learners, and it would be contradictory to create L2 norms of meaning-dominance while also investigating their knowledge of those same senses. In sum there are certain limitations associated with the measure of intra-word frequency used on the TPM. Despite these limitations, the hypothesis of intra-word frequency was supported in Chapter 7, which lends some validity to the procedure taken to control for frequency.

9.5.2 Support for the semantic similarity hypothesis

Despite the limitations of controlling for intra-word frequency, the results support the proposition that the core sense is central to L2 development of semantic knowledge of polysemous words. The core sense is proposed to be central not because it is the most frequent but because it lends the most coherence to the category of different senses. With attempts to control for intra-word frequency, the participants judged the target words as more acceptable if they were consistent with the core sense rather than as a metaphoric extension. At the other end of the scale, the comparison between the metaphor items and the inconsistent items shows that there is a boundary to semantic similarity. This comparison can be reframed as one between senses that are part of the polysemous category and those outside the category. The comparison with inconsistent meanings was largely restricted to low-frequency items, which is actually a very broad category. It could include items many times more frequent than others in the same category as well as items the participants were likely to have never encountered before the test. The results indicate that when semantic similarity ends the participants' judgements become much more unpredictable. Of the 67 participants who took the TPM, 20 were excluded from this analysis, primarily because their responses to the inconsistent meanings were often indistinguishable from their responses to the illogical distractors. The results indicate that the participants are quite accepting in their interpretation of the target words so long as the words bear some semantic similarity to senses they already know.

9.5.3 Semantic variation among metonyms and metaphors

Contrary to the prediction that metonyms would be more acceptable to the participants, they judged the metaphoric use of the target words as more acceptable. This result contrasts with what previous research had found (Klepousniotou, 2002; Klepousniotou & Baum, 2007; Rundblad & Annaz, 2010). However, the results shouldn't be used critically of the previous research because this study was markedly different in design: these studies investigated L1 speakers or L1 learners and not L2 learners, and they used a priming methodology or narrative

238

comprehension task and not acceptability judgements. Aside from the differences in design, perhaps the most pertinent difference was in the selection of metonyms and metaphors. The previous research used metonymic relations expressing a very strong similarity, such as an object/content relationship where *book* [CONTENT] as a metonym of *book* [OBJECT]. In contrast, the metonyms in the TPM were far more distinct as senses: for example, 'air travel' was considered a metonym of 'fresh air', and 'student body' a metonym of 'human body'.

In Klepousniotou (2002), Klepousniotou and Baum (2007) and Rundblad and Annaz (2010), the metonyms are restricted to the same physical object in a different state or different perspective. In the TPM, however, the metonyms have a more extended range of reference, with the limitation that the two concepts are construed within the same domain of experience. The proposed difficulty of these metonyms finds support in the results of Van Herwegen et al. (2013). They observed that object/user metonyms were more difficult for L1 children than part/whole metonyms. The authors comment that children might have taken a more literal view than adults for such utterances as "I want to speak to the pen." It is possible that the L2 learners also took a literal understanding of the target words on the TPM and consequently found the metonyms unacceptable. However, what is notable is that many of the TPM metonyms are actually based on a part/whole relationship, which the L1 children found equally understandable to the metaphors. For example, the TPM item "Gordon is a familiar *face* at the flower show," involves a part/whole metonym. While cognitive maturity might be a factor, this difference between L1 children and L2 adults might also be due to the differences in the tasks the two studies employed. The L1 children were tested for comprehension whereas the L2 adults were tested for acceptability. Like the children, the L2 adults probably understood what was meant by the metonym, but to many of them the target word was used unacceptably when, for example, Gordon was referred to as a *face*.

Just as the metonyms in the TPM express less similarity than in Klepousniotou's research, alternatively the metaphors on the TPM might express more semantic similarity. In Klepousniotou and Baum (2007), two of their three types of metaphors

were based on propositional content: HUMANS ARE ANIMALS, "that man is a fox", and HUMANS ARE PHYSICAL OBJECTS, "that woman is a star." These types of metaphors were not used in the TPM because they appear to rely on cultural conventions. In contrast, the metaphors on the TPM were based on correspondences of image schemas, such as "branches of government". I would argue that the metaphors used in the TPM expressed more semantic similarity than the previous research, just as the metonyms on the TPM expressed less semantic similarity.

The results suggest that L2 learners prefer metaphors based on image schemas over distinct metonyms, but it is not entirely apparent why this might be so. There is a possibility that the results might have been due to uncontrolled factors and that a different sample of items might have produced different results. However, if L2 participants truly do find these type of metonyms less acceptable, then their rejection can be characterised in one of two ways. First, the learners might have rejected the metonym extension because they did not recognise the similarity with the core sense of the word. This is unlikely because the similarity seems no less apparent for the metonyms than for the metaphors. The second possibility is that the learners recognised the similarity expressed by both the metonym extension and the metaphor extension but found the metonyms unacceptable. The main difference between the metonym and metaphor extension is that metonyms express similarity within the same domain as the core sense while the metaphor extension expresses similarity across different domains of experience. The suggestion here is that L2 learners reserve stronger restrictions for the use of words in the source domain but find metaphoric extensions to other target domains more acceptable. While this type of judgement might be possible, it is beyond the scope of the TPM to provide evidence for it.

Two subsequent re-analyses of the results were conducted which indicate that as the learners' proficiency develops, they come to know the metonym extensions equally to the metaphor extensions. First, the results to the middle-frequency items for metonym and metaphor extensions were compared for the 13 participants who scored 90% or above on the VLT. The analysis showed that there was no significant

difference in the mean number of 'Hits' for this high-proficiency group, $t(12) = .475$, $p = .64$. Another consideration is the difference between the items themselves. There were only seven items to either the metonym or metaphor group. With so few items, it could be that a few outliers affected the results. In order to see if the difference between the results was due to the items, a second re-analysis compared the metonym and metaphor categories by item in an independent samples t-test. Since the high proficiency participants had already been shown to find the metonyms and metaphors equally acceptable, their responses were removed from the analysis, as were the responses for those participants already excluded in the downward adjustment procedure. The analysis found no significant difference between the number of responses for items classified as metonyms compared to items classified as metaphors $t(12) = 1.12$, $p = .29$. The second re-analysis indicates that difference between the metonym and metaphor categories lies with the participants and not with individual items.

I have proposed that there is semantic variation within the categories of metonyms and metaphors, and that this variation can explain why the results on the TPM did not follow the results found in Klepousniotou's studies. There is some research on the effect of variation within metonyms and metaphors. Murphy (1997) and Rodd, Berriman, Landau and Lee (2012) investigated how native speakers learned novel language. The novel language was designed to express different types of polysemous extensions. As discussed in Section 2.4.2, Murphy and Rodd et. al. were able to show that different formulations of metonyms and metaphors influenced the judgements of native speakers. Murphy created novel words and then extended the meaning as metonyms. The metonyms could be extended conventionally or unconventionally. For example, using a descriptive paragraph, a *deljid* was presented as a type of tree. Subsequently, in the conventional condition, *deljid* referred to the wood of the tree (a COUNT/MASS relationship); in the unconventional condition *deljid* referred to the forest of that tree (PART/WHOLE relationship). On a 7 point Likert scale for "appropriateness", the conventional extensions were rated 6.1 and the unconventional extensions 3.4. In Klepousniotou's studies, the metonyms

follow Murphy's conventional extensions. In contrast, the metonyms on the TPM are more similar to Murphy's unconventional senses: *face* referred to the whole person, *air* referred to the vehicle using air as a medium. Murphy's results are more consistent with those of the TPM. This indicates that we shouldn't talk about how metonyms are more semantically similar to the core sense as a category, but that certain conventional types of metonyms are more similar to the core sense than other types of metonyms.

Similar variation of meaning can be attributed to the metaphor category. Rodd et al. (2012) investigated native-speaker judgements of novel meanings for known words such as "an *ant* is a tiny device" (related) in contrast to "a *path* is a tiny device" (unrelated). In the related condition, a semantic feature of the conventional meaning is construed against a novel domain. The semantic feature for *ants* is that they are SMALL. The semantic feature SMALL is construed against the domain of DEVICES. In Lakoff's framework, this type of metaphor is based on propositional content because it involves knowledge of the world. This contrasts with the unrelated condition, where there is no similarity between the conventional meaning and the novel meaning. In Rodd et. al (2012), the words were placed in contextualising paragraphs. On a 7 point Likert scale for "plausibility", native speakers judged related meanings at 4.88 and unrelated meanings 3.16. I would argue that Klepousniotou's metaphors are comparable to the unrelated condition. To say that PEOPLE ARE ANIMALS (*He's a fox*) or PEOPLE ARE PHYSICAL OBJECTS (*She's a star*) is to employ conventional metaphors where the similarity between the core sense and the metaphor is no longer easily identified. In contrast, the connection in the related condition, such as between a 'small ant' and a 'small device', seems more inferable. The metaphors employed on the TPM are primarily based on shared image schemas between the core sense and the metaphor. The TPM metaphors are more similar to the related condition because both types of metaphor can be understood without the special cultural knowledge needed to understand the metaphors in Klepousniotou's studies.

It is tempting to combine the results from Murphy (1997) and Rodd et. al. (2012) and use them to explain the results on the TPM. Such a combination is complicated,

because the participants in these studies applied their judgements to instruments based on different research designs. Nevertheless, the measurement of 'appropriateness' and 'plausibility' do appear comparable. Despite these complications, it is interesting to note that the comparable categories from these studies corroborate the results on the TPM. The related metaphors in Rodd et. al. were rated at 4.88/7 for plausibility and the unconventional metonyms in Murphy were rated at 3.4/7 for appropriateness. In the TPM, the metaphor items were judged acceptable 66% on average, and the metonym items were judged acceptable 51% on average. The results on the TPM, as supported by other research, lead to the position that the categories of metonym and metaphor are much more complicated than presented in the design of Klepousniotou's studies.

9.5.4 The effect of intra-word frequency when controlled for semantic similarity

If we return to Table 9.3 (on page 231), we can see that the items on the TPM expressing a metaphor extension are represented across each of the three intra-word frequency levels. The metaphor extension was expressed in seven high-frequency items, seven middle-frequency items and fourteen low-frequency items. The distribution of the metaphor items across the frequency levels offers an opportunity to re-analyse the responses for the effect of intra-word frequency, while controlling for semantic similarity.

The intra-word frequency hypothesis (see Section 5.1) proposes that learners are more likely to learn meanings according to the frequency of the different meanings of a word relative to other senses of that word. In Chapter 7, the responses to the acceptable items were investigated for the effect of intra-word frequency but they were not controlled for semantic similarity. The metaphor category includes items across all three intra-word frequency categories. Thus according to the hypothesis of intra-word frequency, the participants are predicted to respond with the most 'Hits' to the metaphor items in the high-frequency category and fewest 'Hits' to the metaphor items in the low-frequency category. However, the results could be different from the previous analysis, because in Chapter 7 most of the items in the high-frequency category were consistent with the core sense. By controlling for items

in the metaphor category, a comparison across the frequency categories will further support or qualify the hypothesis of intra-word frequency.

The mean number of 'Hits' by participant was calculated for high, middle and low-frequency items while drawing only from the metaphor semantic category. The mean number of 'Hits' for these items was adjusted downwards by participant based on the number of 'False Alarms' to the illogical distractors. Participants were excluded from the analysis if their adjusted mean number of 'Hits' was zero or a negative number. Of the original 67 participants, six participants were excluded. The adjusted results are presented in Figure 9.2.
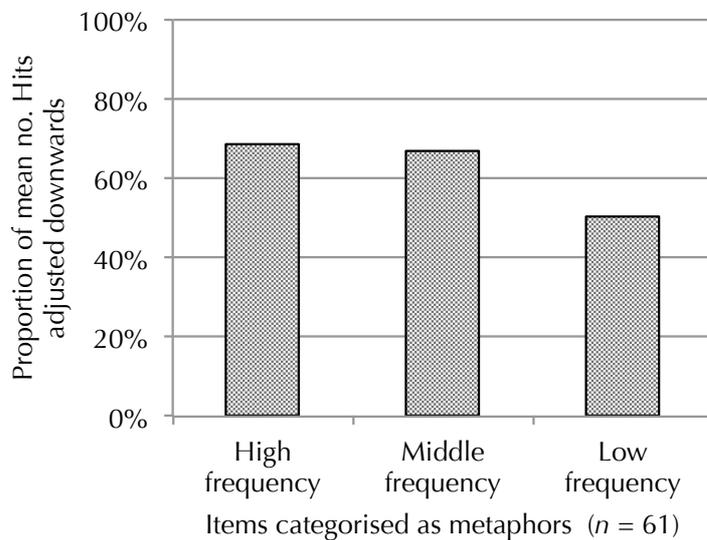
Figure 9.2
Mean number of 'Hits' to the intra-word frequency categories when controlled by the metaphor semantic category. The results were adjusted downwards and presented as proportions.

The Shapiro-Wilk test of normality was conducted for the mean number of 'Hits' for the three groups of items. The results are presented in Table 9.5 for the normality of 'Hits' for these three groups of items.

Table 9.5
Shapiro-Wilk Test of Normality for the mean number of 'Hits' to intra-word frequency for senses expressing a metaphor extension.

Intra-word frequency	Relation to the core sense	n	Statistic	df	Sig.

High-frequency	Metaphor	61	.95	61	.009
Middle-frequency	Metaphor	61	.96	61	.026
Low-frequency	Metaphor	61	.98	61	.279

For the high-frequency and middle-frequency items, the distribution of 'Hits' by participant was significantly different from the null hypothesis of normality at p < .05. Non-parametric tests will be used for the comparisons between all three groups.

The results for the items expressing a metaphor extension of the core sense were analysed using a Friedman one-way ANOVA for repeated measures with a within-subjects factor of intra-word frequency (high, middle and low-frequency senses). This non-parametric analysis was used because the data were not normally distributed for two of the levels and because comparisons were made between responses made by the same participants (repeated measures). The main effect of metaphor extension was significant χ^2 (2, $N = 61$) = 34.27, $p < .001$, *Cramér's V* = .53.

A post hoc analysis was conducted with the Wilcoxon signed-rank test. This test compares the means of two related samples when the distribution of at least one of the samples is not normally distributed. There was no significant difference between the mean number of 'Hits' for the high-frequency items ($M = .69$, $Mdn = .74$, $SD = .20$) and the middle-frequency items ($M = .67$, $Mdn = .68$, $SD = .23$), $z = .57$, $p = .57$. The mean number of 'Hits' to the low-frequency items ($M = .50$, $Mdn = .46$, $SD = .21$) was significantly lower than both the high-frequency items, $z = 4.69$, $r = .42$, $p < .001$, and the middle-frequency items, $z = 4.35$, $r = .39$, $p < .001$.

The results for this analysis were again somewhat against expectation. In earlier chapters, the analysis of the results confirmed the hypothesis of intra-word frequency, that learners are more likely to learn one meaning of a word according to the intra-word frequency of that meaning. In those analyses, the results were not controlled for semantic category. This analysis focussed only on those items where the target words expressed a metaphoric relationship to the core sense. The results

found that the learners responded equally well to the middle-frequency items as to the high-frequency items. This finding indicates that the acceptability of a word is based as much if not more upon semantic factors as upon its frequency of occurrence in the language.

9.6 Summary

The analysis of the TPM items by semantic category largely supported the semantic similarity hypothesis. The analysis provided evidence that the core sense was central to the L2 learners' semantic understanding of the target word because the participants found the target words were used most acceptably when they were consistent with the core sense. Likewise, senses that were inconsistent with the core sense were the least acceptable to the learners, indicating that the learners use semantic similarity to make sense of novel polysemes. Exactly what type of similarity is more or less acceptable to L2 learners isn't entirely apparent. Against expectation, the metaphor extension was found more acceptable than the metonym extension of the core sense. This difference was found primarily among learners at a lower proficiency as measured by a vocabulary size test. It might be that the L2 learners found it more acceptable to extend the core sense figuratively to other domains of experiences rather than extending the meaning within the same domain. However, the results could be due to individual differences among the items because the sample of items used in the comparison was small. Finally, when the senses expressing a metaphor extension were compared across frequency categories, there was no significant difference in the number of 'Hits' between the high and middle frequency items. This finding qualifies the early results supporting the hypothesis of intra-word frequency. The higher number of 'Hits' for the high-frequency items was likely because many of the target words were consistent with the core sense. These results indicate that for L2 learners, polysemous senses are more acceptable when they express a combination of higher intra-word frequency and semantic similarity to the core sense.

In the preceding three chapters, the TPM results have been analysed for the effect of three factors: semantic similarity, intra-word frequency, and L2 proficiency as

measured by vocabulary size. The next chapter will address the possible effect of the learners' L1 on the results of the TPM. The participants in the study reported here were all Arabic L1 speakers. It is worthwhile to consider how their responses might have been influenced by the way the senses of the polysemous English target words were expressed in an equivalent way with polysemous Arabic words.

Chapter 10: The effect of L1 influence

In this thesis, the responses to the Test of Polysemous Meanings were all provided by Arabic-speaking learners of English. Since these participants all shared the same L1, there is a question whether Arabic-semantics were an influence on their responses. Specifically, if two L2 polysemous senses translate into the same L1 word form, these L2 senses may be easier to learn than if they translated into two different L1 word forms. In order to investigate this possibility the responses to the TPM provided in Chapter 7 were reanalysed for the effect of L1 influence. The polysemous senses in the acceptable items on the TPM were translated into Arabic and then grouped according to whether they translated the same as or different from the L1 equivalent word form. In the analysis, the effect of L1 influence is considered in relation to the effects of intra-word frequency and semantic similarity.

10.1 Introduction

In the preceding chapters, the investigation of L2 knowledge of polysemous senses was approached by analysing the results of acceptability judgements for the influence of several competing factors. First, influence for the factor of intra-word meaning frequency was demonstrated by the higher number of acceptability judgements for more frequent senses of polysemous words. Moreover, L2 learners with a greater overall vocabulary size were shown to make acceptability judgements closer to the native speaker baseline. While the effect of a greater vocabulary size complemented the effect of intra-word frequency, the factor of semantic similarity was shown to exert an independent influence on the learners' acceptability judgements. The learners judged distractor items as acceptable more often if the sense of the polysemous word was semantically related to the core sense, in contrast to semantically unrelated distractor items. Moreover, when the senses of acceptable items were controlled for intra-word frequency, the L2 learners found polysemous senses acceptable more often when they expressed greater semantic similarity to the core sense of the word.

A final factor needs to be considered in the investigation of L2 knowledge of polysemous senses. The responses to the Test of Polysemous Meanings (TPM) were

gathered from L2 learners who spoke Arabic as their first language (L1). We should consider the possibility that the learners judged L2 senses as acceptable more often if the senses were equivalently polysemous in their L1. For example, in English *chicken* refers to both the ANIMAL and its MEAT. If a learner's L1 equivalent for *chicken* also referred to both senses, as is the case in Arabic, then a learner of English, who knew the ANIMAL sense, might be influenced by her L1 to transfer the concept of MEAT to the L2 word form *chicken.*

The *chicken* example illustrates a further point about some polysemous words because many other MEATS in English are also referred to by the name for the animal. For this reason, an English speaker might find it easy to remember that the L2 word for *beef* was referred to by the translation equivalent of *cow*. While *beef* in English doesn't follow the polysemous pattern of ANIMAL FOR MEAT, knowing the pattern itself might facilitate learning. There are many polysemous patterns in English which recur across other languages. Srinivasan and Rabagliati (2015) conducted a cross-linguistic study comparing 27 polysemous patterns in English with 14 other languages. A pattern like ANIMAL FOR MEAT was considered generative because it was reasonable to coin a novel expression, such eating *seagull*. Other patterns were common across many languages but not generative, such as MATERIAL FOR OBJECT. In English, a *glass* refers to a drinking vessel, whereas in Spanish, the translation equivalent of *glass* refers to a car's windshield. There were also patterns which were less common across languages, such as in *Milk the cow* where a substance refers to its removal from the source.

When attempting to explain the constraints which allow for these recurrent patterns across languages, Srinivasan and Rabagliati proposed a model of 'conventions-constrained-by-concepts', whereby the senses are learned individually "but the conceptual structure makes some types of relations between senses easier to grasp than others," (Srinivasan & Rabagliati, 2015, p. 125) . This model is consistent with the hypothesis of semantic similarity for L2 learning (cf. Section 8.4.3 on page 199) because both argue that recognizing the pattern or similarity between senses facilitates comprehension. However, the patterns of polysemy presented by

Srinivasan and Rabagliati are much more specific than the general pattern of similarity explored in Chapter 9. While Srinivasan and Rabagliati identify 27 different patterns of polysemy, the analysis in Chapter 9 grouped the polysemous senses as either metaphors or metonyms. For example, the TPM includes the item "The president sat at the *head* of the table," which follows the polysemous pattern of BODY PART FOR OBJECT. However, in Chapter 9, this sense was categorised as a metaphor which exploits the UP-DOWN image schema.

Another point of difference between the polysemous words across the two studies is that the words which Srinivasan and Rabagliati investigated are often not as generally polysemous as the TPM target words. For example the words *window* and *Vietnam* are polysemous according to the patterns FIGURE FOR GROUND and PLACE FOR EVENT. However, these words are not highly polysemous outside of these restricted patterns. In contrast, the polysemous words used on the TPM include a minimum of five distinct senses in the LDOCE. While few of these senses follow a regular pattern, they may be more indicative of polysemy in the language as a whole. That being said, there are three TPM senses which follow the pattern of BODY PART FOR OBJECT: *face* of the watch, *head* of the table, *arm* of the chair. The results of these three items will be addressed specifically in the discussion.

In the example above, *chicken* is ambiguous in both the learner's L2 and her L1. However, much of the research into L1 influence on knowledge of L2 polysemous senses has looked at translation ambiguity, where the word in one language is ambiguous, while in the other language, the senses are represented by separate word forms. This research was reviewed in Section 2.4.3. One finding from translation ambiguity research showed that the mapping of concepts in the L1 continues to influence semantic judgements in the L2. In a semantic acceptability task, Jiang (2002, 2004) found that L2 learners recognised semantic similarity faster between two L2 words which translated into the same L1 word form. This finding was corroborated by Elston-Güttler and Williams (2008). They conducted an anomaly detection task, where bilinguals had to decide whether a target word was used correctly in an L2 sentence. In the critical condition, the L2 target word was a

mistranslation of an L1 polysemous word: the German *Blase* translates as both *blister* and *bubble*, but 'His shoes were uncomfortable due to a bubble,' would be unacceptable in English. The bilinguals made more errors and longer responses for the critical condition. These studies are taken as evidence of how polysemous words in the L1 can influence comprehension in the L2.

Furthermore, there is some evidence that the L2 can also have an effect on the L1. Degani, Prior & Tokowicz (2011) found that when proficient bilinguals were compared to monolinguals, the bilinguals rated two L1 words as more similar if the meanings of the words were expressed as a single polysemous word in their L2. However, it doesn't appear that this result would be reproduced with L2 learners. Frenck-Mestre & Prince (1997) compared proficient bilinguals to less proficient L2 learners in a lexical priming task with homographs as ambiguous L2 words. In their task, a homograph such as *fire* could be primed with the word *hot* for its dominant meaning, *gun* for its subordinate meaning, or *late* for an unrelated word. They found that in proficient bilinguals, homographs were primed with both dominant and subordinate L1 meanings, but with intermediate L2 speakers only the dominant meaning was primed. This indicates that the L2 exerts less influence on the L1 for intermediate L2 learners.

These studies into translation ambiguity provide evidence that concepts are shared across languages and, moreover, the mapping of a concept to a word form in one language can affect the mapping of that concept to the word form in the other language. These studies have primarily involved proficient bilinguals and not learners. In other words, their results are more about the outcome of learning than about the process of learning itself. Only Frenck-Mestre & Prince (1997) included intermediate learners in their research, but even here there was no discussion of how translation ambiguity might inhibit or facilitate language learning.

The effect of translation ambiguity on L2 learning was considered by Kellerman (1986) who investigated the likelihood that an L2 learner would transfer an L1 sense to her L2. The investigation of transfer asks the following question: if an L1 word has

senses A and B, and a learner knows the L2 word for sense A, will she transfer sense B to the same L2 word?

Kellerman proposed that learners will transfer polysemous senses from their L1 to their L2 based on the sense's prototypicality. He defined prototypicality as a combination of the sense's similarity to a core sense and its subjective frequency. The measures of similarity and subjective frequency were elicited from L2 learners using a Likert scale. In my critique of his study (see Section 2.5.5), I commented that while the measure of similarity was relatively straightforward, the measure of frequency was less reliable. As Kellerman himself noted, it was unclear whether the participants were considering the lexical frequency of the sense or the frequency of the sense's real-world object. With the absence of strong intra-word frequency data in the learners' L1, it is difficult to measure this component of Kellerman's model of prototypicality.

The other component of Kellerman's model was similarity to the core sense. Kellerman found that he could measure this component with more confidence. His hypothesis was that a language user would transfer a polysemous sense from her L1 to her L2 based, in part, on the sense's similarity to a core sense. However, the hypothesis of L1 transfer does not consider whether equivalency across languages benefits L2 learning. L1 transfer only considers whether a learner is likely to use an L2 word in the same way as she uses the L1 equivalent in her native language. For example, in a learner's L1 there is a polysemous word with senses A and B. The learner has just learned the L2 word form for sense A. The question of L1 transfer considers whether the learner will use the L2 word form for sense B *without evidence from the L2.*

L1 influence, the phenomenon to be addressed in this chapter, is different from L1 transfer. For the purposes of the investigation, L1 influence will be characterised to include the benefit of evidence from the L2. To illustrate with another example, an L2 learner has just learned the L2 word form for sense A. This L2 word can be used for senses A, B and C. However, in the learner's L1, the equivalent word form can

253

only be used for senses A and B. Sense C is mapped to a different word form in the L1. The question of L1 influence asks whether sense B is easier to learn in the L2 than sense C. An example of L1 influence can be drawn between the English word *head* [BODY] and the equivalent French word *tête* [BODY]. These two words are also equivalent in a secondary sense because both *head* and *tête* have a NAIL sense, as in "Hit the nail on the head." However, the two words are not completely equivalent. While the English word *head* has the sense of PERSON IN CHARGE, the French word *tête* cannot be used in this sense. The question of L1 influence asks whether the French learner of English can learn the NAIL sense more easily than the PERSON IN CHARGE sense. The question is not whether she is more likely to transfer the sense from her L1, but whether she needs less evidence from the L2 to learn the sense. Based on this argument, a hypothesis is proposed for L1 influence on receptive knowledge of polysemous senses. It states that an L2 learner will respond with more 'Hits' to acceptable items on the TPM if the sense of the target word translates into the same L1 word-form as the L1 equivalent. The L1 equivalent is identified as the direct translation of the core sense for the L2 target word.

For Kellerman, similarity to the core and L1 transfer were based on subjective ratings of participants about their own L1 without recourse to the L2. However, the investigation of L1 influence presented here will be based on different data sources from those of Kellerman. In this analysis, the participants' L1 is Arabic and their L2 is English. The core sense of the (English) L2 word is based on the dictionary entry from the Oxford Dictionary of English (ODoE). The L1 equivalent of the core sense is based on the direct translation of the L2 word[34]. The equivalency of the two word forms is based on whether the polysemous senses of the word in the L2 translate into the same L1 word form as the L1 equivalent. Finally, the L2 polysemous senses are controlled by a measure of intra-word meaning frequency. As discussed in Sections 4.3.1 and 4.3.2, this measure was established by the order of the L2 senses for a given

---

34. The details of the method are explained in 10.3 below.

word form as they are listed in the Longman Dictionary of Contemporary English (LDOCE).

The investigation of L1 influence presented in this chapter uses the responses to the TPM from Chapter 7. Evidence of L1 influence is gathered by noting how frequently the learners responded with 'Hits' to acceptable items which translated into the same L1 word form as the core sense. These results will be compared to senses which translated into a different word form from the core sense.

10.2 Research Question

The TPM is composed of three acceptable items for each target word. When the target word for these items is translated into the learners' L1, some senses will translate into the L1 equivalent word, while others will translate into a different L1 word from the L1 equivalent. The example given above was that core sense of *head* [BODY] in English is the translation equivalent of *tête* [BODY] in French. The French equivalent can be used in the sense of head of a NAIL, but not in the sense of PERSON IN CHARGE. Using this example, the hypothesis would predict that a French learner of English would find "He hit the nail on the head," more acceptable than "The president is the head of the company." However, the validity of this hypothesis is based on two conditions. The first is that the learner already knows the core sense of the L2 word and the second is that both extended senses are balanced for intra-word frequency. The second condition is because it has already been established that higher frequency senses are more likely to be judged as acceptable. In order to address this potential confound, senses will be matched for frequency.

To investigate the hypothesis for L1 influence on receptive knowledge of polysemous senses, the TPM items are separated into two groups: those items for which the sense of the target word translates into the same L1 word-form as the L1 equivalent and those items for which the sense translates into a different L1 word-form from the L1 equivalent. For the sake of clarity, the two groups will be referred to as 'the same as the L1 equivalent' and 'different from the L1 equivalent'. The

255

following research question is proposed to investigate the hypothesis for L1 influence on receptive knowledge of polysemous senses:

RQ: *Is there a greater proportion of 'Hits' for responses to items which are 'the same as the L1 equivalent' in comparison to responses to items which are 'different from the L1 equivalent'?*

In this analysis the L1 equivalent is defined as the translation of the core sense, and a learner's responses are only recorded if he or she has judged the core sense as acceptable.

10.3 Translation and categorisation of the acceptable items on the TPM

A native Arabic speaker was recruited to provide translations for the senses of the 20 target words for the TPM presented in Chapters 6 and 7. The translator was a fluent bilingual in Arabic and English. She had completed her Bachelor's degree in an Arabic medium and her post-graduate degrees in an English medium. Her PhD was in the field of applied linguistics and she had been living in the UK for the last 10 years. The translator provided translations to the senses of the 20 target words used on the TPM. The senses were those listed in the LDOCE and all the senses were translated from English to Arabic. The translator used *Al-Mawrid: A Modern English-Arabic Dictionary* (Ba'albaki, 2002) as a reference.

The first step in organising the data for analysis was to identify a translation equivalent for each target word. In Chapter 9, the core sense of a polysemous word was identified as the sense which provided the most coherence to the category of the polysemous word. The centrality of the core sense made it the best candidate for the L1 equivalent. As mentioned above, the items which share the same translation as the L1 equivalent can be compared to those which translate differently. In order to balance the items for intra-word frequency, items will only be compared to those of the same intra-word frequency category. While the measurement of intra-word frequency is relative to other senses of the same target word, there should be an attempt to balance the senses between target words as well. In Chapter 7, the measurement of intra-word frequency, taken from the LDOCE, was shown to be a

significant factor in the participants' acceptability judgements of the TPM items. Thus, while an actual corpus count would be a more satisfactory measure of sense frequency, controlling the items by intra-word frequency should help to isolate the participants' responses to the effect of L1 influence.

Table 10.1 presents the translations for the acceptable items on the Test of Polysemous Meanings. For three of the target words (*board*, *class* and *form)*, the TPM senses didn't include the core sense, and thus there was no evidence that the learners knew this sense or not. For this reason, the responses to the items for these three words were not included in the analysis. Each sense for the remaining 17 target words was translated into Arabic, the learners' L1. These translations were then categorised according to translation equivalency.

Table 10.1
Translations of the senses used in the TPM

Target word	Intra-word frequency	Sense	L1 translation	Translation equivalency
air	High	GAS	هواء	L1 equivalent
	Middle	PLANES	طيران/ طائرة	Different
	Low	BEHAVIOUR	تظاهر	Different
arm	High	BODY	ذراع	L1 equivalent
	Middle	WEAPONS	سلاح	Different
	Low	FURNITURE	يد الكرسي	Different
body	High	PEOPLE/ANIMALS	جسم	L1 equivalent
	Middle	GROUP	جماعة	Different
	Low	BODY OF SOMETHING	كم	Different
branch	High	OF A TREE	فرع	L1 equivalent
	Middle	OF GOVERNMENT	قسم	Different
	Low	OF A RIVER/ROAD	فرع	Same
case	High	SITUATION	وضع	L1 equivalent
	Middle	LAW/CRIME	قضية	Different
	Low	DISEASE	حالة	Different
character	High	ALL SOMEBODY'S QUALITIES	طبع	L1 equivalent
	Middle	MORAL STRENGTH	خُلق	Different
	Low	LETTER/SIGN	احرف/ اشكال	Different

course	High	EDUCATION	مقرر تعليمي	Different
	Middle	DIRECTION	اتجاه	L1 equivalent
	Low	MEDICAL TREATMENT	مجموعة جرعات	Different
cover	High	PROTECTION	غطاء	L1 equivalent
	Middle	WEATHER	غطاء	Same
	Low	BED	غطاء	Same
cut	High	REDUCTION	تخفيض	Different
	Middle	HOLE/MARK	قطع	L1 equivalent
	Low	SHARE OF SOMETHING	استقطاع	Different
face	High	FRONT OF YOUR HEAD	وجه	L1 equivalent
	Middle	PERSON	وجه	Same
	Low	CLOCK	وجه	Same
hand	High	PART OF BODY	يد	L1 equivalent
	Middle	HELP	مساعدة	Different
	Low	WORKER	عامل	Different
head	High	TOP OF BODY	رأس	L1 equivalent
	Middle	CALM/SENSIBLE	عقل	Different
	Low	FRONT/LEADING POSITION	رأس	Same
heart	High	BODY ORGAN	القلب	L1 equivalent
	Middle	EMOTIONS/LOVE	القلب	Same
	Low	THE CENTRE OF AN AREA	قلب	Same
line	High	ON PAPER/ON THE GROUND	خط	L1 equivalent
	Middle	WAY OF DOING SOMETHING	أسلوب	Different
	Low	FAMILY	اسرة	Different
order	High	FOR A PURPOSE	لكى	Different
	Middle	ARRANGEMENT	بالترتيب	L1 equivalent
	Low	SOCIAL/ECONOMIC SITUATION	نظام	Different
point	High	IDEA	نقطة	L1 equivalent
	Middle	PLACE	نقطة	Same
	Low	SHARP END	طرف/حرف	Different
position	High	SITUATION	موقف/ مركز	Different
	Middle	PLACE WHERE SOMETHING IS	مكان	L1 equivalent
	Low	JOB	مركز	Different

In Table 10.2 below, the total count of the translation categories is presented according to intra-word frequency and translation equivalency.

Table 10.2
Count of the TPM senses when organised by intra-word frequency and translation. The shaded area includes the results compared in the analysis.

Intra-word frequency	L1 equivalent	Same translation as L1 equivalent	Different translation from L1 equivalent	Total
High-frequency senses	13	0	4	17
Middle-frequency senses	4	4	9	17
Low-frequency senses	0	5	12	17

In total, 13 of the high-frequency senses and four of the middle-frequency senses were consistent with the core sense and were identified as the L1 equivalent. These senses were not used in the analysis because the L1 equivalent was considered the source of L1 influence and the analysis investigated its effect. Once these items were removed, none of the high-frequency senses shared the same translation as the L1 equivalent. For this reason, no comparison could be made with the results to the high-frequency items. Among the middle-frequency senses, four shared the same translation as the L1 equivalent and nine translated differently. The results for these items were selected for comparison. Finally, among the low-frequency senses, five shared the same translation as the L1 equivalent and 12 translated differently. These senses were also selected for comparison.

10.4 Analysis of the TPM responses for L1 influence

As discussed in Section 7.2.1, the responses to the TPM were gathered from 67 Arabic learners of English. The following four steps were taken in order to organise the responses and calculate the mean number of 'Hits' for comparison:

1. If the participant did not judge the core sense of a target word as acceptable, then his or her responses to the items for this target word were removed. This was

done because the hypothesis of L1 influence was based on knowledge of the core sense.

2. The responses were classified as 'Hits' and 'Misses' according to signal detection theory, as presented in Section 5.3. The number of 'Hits' were organised by two independent variables, translation equivalency (same as the L1 equivalent or different from the L1 equivalent) and level of intra-word frequency (middle-frequency senses or low-frequency senses).

3. The proportion of 'Hits' to total number of responses was adjusted downwards to compensate for creative guessing. As discussed in Section 7.2.4, the downwards adjustment was calculated using the number of 'False Alarms' the participant made in response to illogical distractors.

4. In some cases, the mean number of 'Hits', when adjusted downwards, was zero or a negative number. In such cases the participant's responses were removed from the comparison. The rationale for this action was discussed in Section 7.3.3. For the comparison at the middle frequency level, the results of eleven participants were removed, and for the comparison at the low frequency level, the results of 14 participants were removed.

Figure 10.1 presents the mean proportion of 'Hits', adjusted downwards, when organised by translation equivalency and intra-word frequency level.
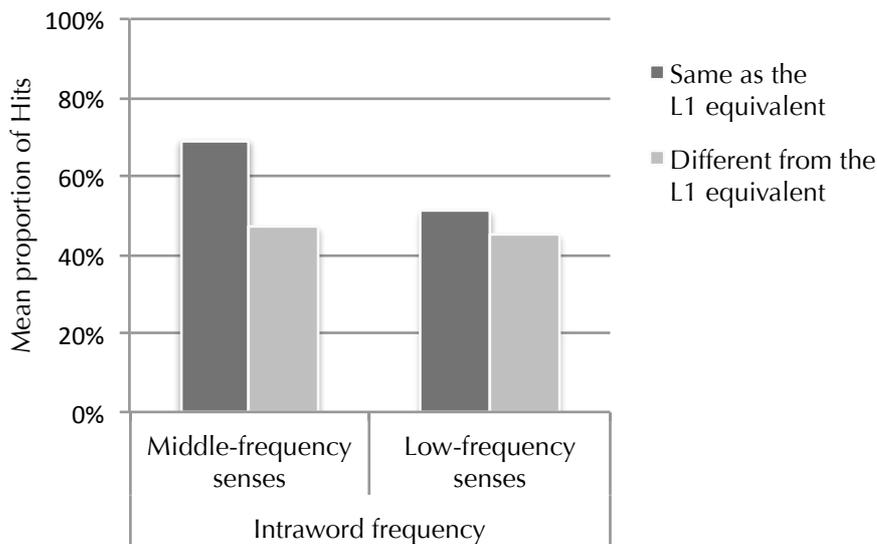
Figure 10.1
Mean proportion of 'Hits' to the TPM items when organised by translation equivalency and intra-word frequency. The results were adjusted downwards and presented as proportions.

In order to address the research question it was necessary to compare the mean number of 'Hits' between items translated as the same as the L1 equivalent and items translated differently. Table 10.3 presents the results to a Shapiro-Wilk test of normality to determine whether parametric or non-parametric tests should be used.

Table 10.3
Shapiro-Wilk Test of Normality for the results of the TPM when the senses are organised by their relationship to the L1 equivalent.

Grouping	Translation	Statistic	Sig.
Middle-frequency senses (*df* = 56)	Same as the L1 equivalent	0.86	0.001
	Different from the L1 equivalent	0.98	0.541
Low-frequency senses (*df* = 53)	Same as the L1 equivalent	0.93	0.048
	Different from the L1 equivalent	0.97	0.006

For the middle-frequency level, the distribution of means for senses which shared translation with the L1 equivalent was significantly different from the null hypothesis of normality at $p < .05$, and the results were significantly skewed, *skew* = -0.58, $z = -1.81$, $p < .05$. Non-parametric tests were used for comparisons within the results for the middle-frequency senses. For the low-frequency level, both groups

261

were significantly different from the null hypothesis of normality. Non-parametric tests were also used for comparisons within the low-frequency senses.

For the middle-frequency senses, a sign test was used to compare the results of the senses with the same translation as the L1 equivalent (M = .69, Mdn = .70, SD = .31) to those senses with a different translation (M = .48, Mdn = .50, SD = .23). There was a significant difference between the mean proportion of 'Hits' for these senses, $z(56)$ = 3.25, p = .001, r = .43. For the low-frequency senses, a Wilcoxon signed ranks test was also used to compare the results of the senses with the same translation as the L1 equivalent (M = .51, Mdn = .43, SD = .28) to those senses with a different translation (M = .45, Mdn = .41, SD = .24). There was no significant difference between the mean proportion of 'Hits' for these senses, $z(53)$ = 1.107, p = .269.

The analysis supported the hypothesis of L1 influence for the middle-frequency level, but it did not support the hypothesis for the low-frequency level.

10.5 Discussion

The L1 influence hypothesis predicted that the participants would find the TPM items more acceptable if the sense of the target word translated into the same L1 word-form as the L1 equivalent. The results confirmed this prediction for senses at the middle frequency level but not at the low frequency level. One possible explanation for this is that the L2 learners are more likely to have encountered senses of the target words at the middle frequency level, which would have given them more evidence that the equivalent sense could be transferred from their L1.

However, it is important to consider whether the results are due solely to the effect of L1 influence and not to other factors. While the analysis of the results attempted to control for the effect of intra-word frequency, it did not control for semantic similarity. Previously, the participants were shown to respond to metaphor extensions with more 'Hits' than metonym extensions of the core sense (see Section 9.4, on page 236). They were also shown to respond with the fewest 'Hits' to items where the meaning of the target word was inconsistent with the core sense. In the analysis of L1 influence, the groups of items contained different proportions of

metaphors, metonyms and inconsistent meanings. Figure 10.2 shows the proportion of items classified according semantic similarity for the four groups of items used in the comparisons for L1 influence.
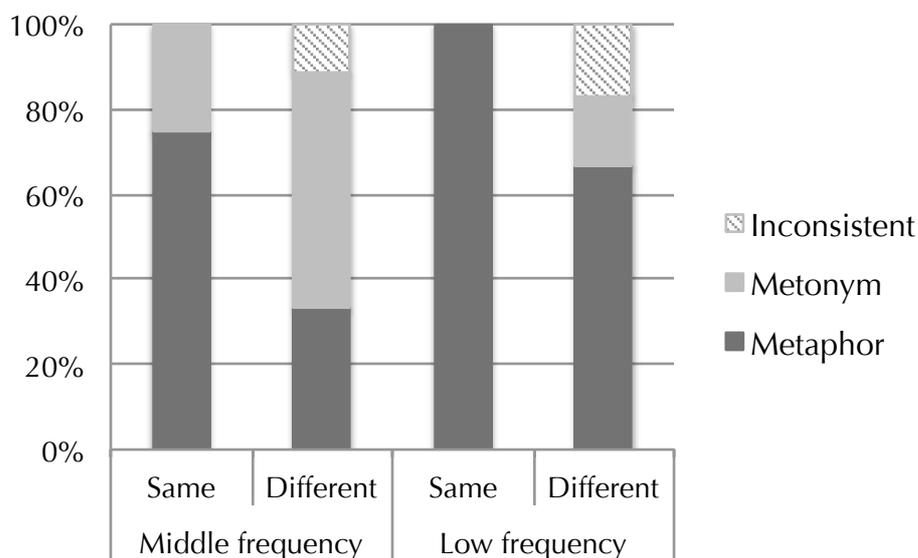
Figure 10.2
The proportion of semantic-similarity categories for the acceptable items when organised by translation (same as the L1 equivalent or different from the L1 equivalent) and intra-word frequency (middle-frequency items and low-frequency items).

Based on semantic similarity, there were more metaphor items among the same translation categories, which might have been a factor in the higher acceptability results for the same translation items. In contrast, items in the different translation category included items with inconsistent meanings. This might have been a reason why the participants found the items in this category less acceptable. In sum, there is reason to believe that the results of the analysis were not entirely due to the factor of L1 influence but also due to the factor of semantic similarity. It would be worthwhile to investigate the effect of L1 influence when the items are controlled for semantic similarity.

Table 10.4 shows the number of TPM items when organised by the three factors of intra-word frequency, translation equivalency and semantic similarity. (The items were originally categorised for semantic similarity in Table 9.2 on page 224).

Table 10.4
The number of TPM items when grouped by intra-word frequency, translation equivalency and semantic similarity. The shaded area indicates the items used in the subsequent comparison.

Intra-word Frequency	Compared to L1 Equivalent	Metaphor	Metonym	Inconsistent
Middle	Same	3	1	0
	Different	3	5	1
Low	Same	5	0	0
	Different	8	2	2
Middle & Low	Same	8	1	0
	Different	11	7	3

A comparison of the metaphor items which translate the same as the L1 equivalent to the items which translate differently should give an indication of the strength of L1 influence when controlled for semantic similarity. It would not be feasible to compare items in the metonym category because there are no low-frequency items which express both a metonym extension and the same translation as the L1 equivalent. Even for the metaphor category, the number of items by frequency condition is quite small, and it would be better to group the middle and low-frequency items together. This would leave eight items which translate the same as the L1 equivalent and 11 items which translate differently. Such a grouping would include enough items to be representative of the metaphor category.

As in the previous analysis, responses to the TPM were first presented in Chapter 7. The following steps were taken to select the responses for the comparison of 'Hits' for items expressing a metaphor extension between those which translated the same as the L1 equivalent and those which translated differently.

1. Responses to the middle and low-frequency items were selected if the participant responded that the core sense of a target word was acceptable. Again, this was done because the hypothesis of L1 influence was based on knowledge of the core sense.

2. The items expressing a metaphor extension of the core sense were then selected from the set of responses.

3. These responses were then organised by translation equivalency: the same as the L1 equivalent and different from the L1 equivalent.

4. As in the previous analysis, the mean number of 'Hits' for these groups of items was adjusted downwards. The downwards adjustment was based on the number of 'False Alarms' the participant made to the illogical distractors.

5. Participants were excluded from the analysis if their adjusted mean number of 'Hits' was zero or a negative number. Of the original 67 participants, twelve participants were removed.

A Shapiro-Wilk test of normality showed that the distributions of the results were not significantly different from the normal distribution: $W(55) = .98$, $p = .35$ for the metaphor items which shared the same translation as the L1 equivalent, and $W(55) = .97$, $p = .11$ for the metaphor items which translated differently from the L1 equivalent. A paired t-test was used to compare the results of the senses with the same translation ($M = .58$, $Mdn = .56$, $SD = .22$) to those senses with a different translation ($M = .55$, $Mdn = 56$, $SD = .24$). There was no significant difference between the mean number of 'Hits' for these senses, $t(54) = 0.835$, $p = .27$.

The results of the comparison indicate that for the items expressing a metaphor extension from the core sense, the effect of L1 influence was not strong enough to be measured with the current design. There may be a more measurable effect with a larger number of items. The result might also be different with items expressing a metonym extension or an inconsistent meaning from the core sense.

The results of the analysis for L1 influence may seem counterintuitive. One might expect a sense to be more acceptable to an L2 learner if there was an equivalent mapping for that sense in her L1. Indeed, such an expectation was predicted based on the research from translation ambiguity studies (Degani et al., 2011; Elston-Güttler & Friederici, 2005; Elston-Güttler & Williams, 2008; Jiang, 2002; Jiang, 2004).

The sum of this research leant support to the bi-directional influence of L1 or L2 ambiguity on the lexical activation of concepts in the other language. However, as previously pointed out, these studies all dealt with proficient bilinguals, whereas the learners taking the TPM were of lower L2 proficiency. It is likely that the results reported in these translation ambiguity studies are the implicit effect of shared concepts across the bilinguals' L1 and L2.

In contrast to the proficient bilinguals, the lower proficiency L2 learners are not expected to know many of the figurative senses for the target words on the TPM. As a consequence, they are more likely to engage in an explicit strategy of translation in order to interpret the meaning of an unfamiliar sense. However, if the learners engaged in an explicit strategy of translation, then why did the results not support a stronger benefit for shared translation with the L1 equivalent? There are two possible explanations, each with support from previous studies. The research of Frenck-Mestre and Prince (1997) offers evidence that the learners may have incorrectly translated the sense, while the research of Kellerman (1986) and MacArthur and Littlemore (2008) offers evidence for unacceptable judgements despite correct translation.

As discussed previously, Frenck-Mestre and Prince compared proficient bilinguals to L2 learners. They found that an ambiguous L2 word would prime the dominant sense in the L1 for both groups; however, it would only prime a subordinate L1 sense for the bilinguals. The implication of this study is that when translating the target word on the TPM, the L2 learners may have only translated the core sense of the word and not the extended sense. This point can be illustrated with the following item from the TPM:

(1)   I have a house in the *heart* of the city.

The sense of *heart* in (1) has equivalency in Arabic. However, the implication of Frenck-Mestre and Prince's research is that the learner may have mistranslated *heart* to the dominant BODY sense in Arabic because the subordinate CENTRE OF AN AREA sense was not activated, despite its equivalency in Arabic.

Kellerman's research offers an alternative explanation for the results. Kellerman presented his L2 learners with different senses of a polysemous word in the learners' L1. He asked which senses were likely to translate into the L2 using the equivalent L2 word form. Unknown to the learners was that all the senses could be directly translated into the L2. Kellerman found that L2 learners were less likely to transfer polysemous senses from their L1 to their L2 if the senses lacked subjective frequency and were unprototypical in relation to the core sense.

In the present study, the L2 learners showed a similar reluctance to judge infrequent, unprototypical senses as acceptable. This reluctance was despite the fact that the equivalent word in their L1 could be used in this sense. The implication of Kellerman's research is that when an L2 learner judges a sense to be unacceptable, she has not necessarily misinterpreted the meaning of the sense. To return to the *heart* example at (1), an Arabic learner of English who rejects this sense hasn't necessarily misunderstood the meaning. She may have interpreted the meaning correctly based on its equivalency in her L1. However, despite understanding the meaning, she is just as likely to reject the sense because she hasn't encountered it before (low frequency) and it seems unusual (low prototypicality).

This finding was partially supported by MacArthur and Littlemore. They found that L2 English learners were unable to predict how nouns, such as *snake*, might be used as verbs, despite a similar figurative use with the equivalent word in their L1. However, the same L2 learners were successful at inferring such figurative uses when they read them in sentence length contexts. What isn't clear from MacArthur and Littlemore's qualitative study is whether the learners' successful inferencing was due to L1 influence or to the semantic similarity to the core sense. It should be pointed out that the task in MacArthur and Littlemore's study was notably different from the task on the TPM. In MacArthur and Littlemore's study, the learners expected the use of the word to be acceptable, while the learners who took the TPM, expected the some of the target words to be used unacceptably. This difference in

tasks between the two studies might easily have affected the inference decisions made by the L2 learners.

The research from Frenck-Mestre & Prince and from Kellerman offer alternate explanations for the weak influence of the L1 reported in the results of the TPM. The explanations are not mutually exclusive and could be cumulative. At times a learner may incorrectly translate the extended sense of the target word, and yet, when she does correctly translate the sense, she may reject it for being unprototypical and infrequent. It is important to note that results from Frenck-Mestre & Prince and from Kellerman proceed from very different research designs in comparison to the TPM. First, Frenck-Mestre and Prince's research involved a priming methodology which is used to tap into implicit cognitive processes involved with lexical retrieval. In contrast, the task on the TPM engages with more explicit cognitive processes involved with inferencing. Second, Kellerman's research on L1 transfer did not employ evidence from L2 language. In contrast, the L2 learners who took the TPM had to interpret the meaning of sentences in their L2. For these reasons the strength of the support from these studies is qualified when using them to explain the results of the TPM.

A final consideration of L1 influence pertains to whether shared patterns of polysemy across languages facilitate L2 learning or not. These shared patterns of polysemy were introduced in this chapter (Section 10.1, page 250) in reference to Srinivasan and Rabagliati cross-linguistic study (Srinivasan & Rabagliati, 2015). While the TPM was not designed for this type of analysis, among the TPM acceptable items, three low-frequency senses follow the pattern of BODY PART FOR OBJECT: *face* of the watch, *head* of the table, *arm* of the chair. All three translated into the same word as the L1 equivalent and the participants responded to these results more 'Hits' than average (*n* = 67, *mean no. of 'Hits'* = 38): *head* of the table was above average at 43 'Hits'; *face* of the watch was above average 48 with 'Hits'; and the *arm* of the chair was about average at 36 'Hits'. While it is difficult to say whether there is any benefit to shared patterns of polysemy across languages based only on these examples, there has been some research on L1 acquisition of patterns of polysemy.

268

Evidence of cognitive predispositions to certain polysemous patterns in L1 development would offer some reason to further investigate the effect on L2 learning.

Srinivasan and his colleagues conducted research into the L1 acquisition of three polysemous patterns in English. The first pattern investigated was CONTAINER FOR REPRESENTATIONAL CONTENTS as in "Carry a *book*," [OBJECT] compared to "Read a *book*" [CONTENT] (Srinivasan & Snedeker, 2011). In their experimental task, the researches used novel words for polysemous concepts (*'blicket'* for *book*). Initially, the concept is described as a physical object (a physically long book). When a story emphasized the abstract properties (a short book in content), children will reject a statement about the physical qualities ("Ernie read the long book"), if it contradicts these abstract qualities. The researchers argue that this may be the result of foundational, generative properties of the lexicon or conceptual system. The pattern was common across the 14 languages investigated, which indicates that results would be common across most other languages.

Three other patterns were researched for L1 learning, also using similar experimental task with novel words . In one study, the pattern SUBSTANCE FOR PLACING SUBSTANCE AT GOAL ("*Seed* the garden") was compared to the pattern SUBSTANCE FOR TAKING SUBSTANCE FROM SOURCE ("*Seed* the watermelon") (Srinivasan & Barner, 2013). The results provided evidence that children have a goal bias and that the SUBSTANCE FROM SOURCE pattern was counter-intuitive. Again how common the patterns are across languages support the results of the children's preferences: the SUBSTANCE AT GOAL pattern was common across all but two of the 14 languages investigated, while the cross-linguistic SUBSTANCE FROM SOURCE pattern was the least common across languages with six languages lacking evidence of the pattern. A final study provided evidence that children spontaneously generalize instrument-activity flexibility to new words *(Srinivasan, Al-Mughairy, Foushee, & Barner, 2017)*. When investigating the pattern INSTRUMENT FOR ACTIVITY, the researchers presented the children with a novel word for an instrument as a noun (*"gork"* for *shovel*), and elicited the children to produce the a verb form of the

word ("*gorking*"). Cross-linguistically, evidence for this pattern was found across all languages researched except Hungarian. In sum, evidence from L1 learners suggests that if a pattern is common across languages then it might also have generative properties for children. There is then a possibility that these common patterns may also be easy for L2 speakers to learn.

Overall, the factor of L1 influence has proven to exert a much weaker effect on the results of the TPM than the factors of semantic similarity and intra-word frequency. All of these factors will be reviewed in the general discussion that follows. The discussion will also review the limitations of the research design associated with the TPM, as well as the implications that this research has for teaching L2 polysemous words.

Chapter 11: General Discussion

11.1 Introduction

The studies presented in chapters 4-10 have focussed on the design and implementation of an instrument called the Test of Polysemous Meanings (TPM), which was developed to measure the knowledge of English polysemous words for intermediate L2 learners. The main thrust of the analysis was to see if linguistic descriptions of polysemy are able explain why L2 learners know or don't know certain senses. The power of this explanation was placed in comparison to the effects of intra-word frequency of polysemous senses, the language learner's L2 vocabulary size, and the influence of the learner's L1. Because of the incremental nature of the studies in this thesis, I have included interim discussions of the ways in which study findings and specific research literature have related to these aims. Specifically, I discussed the effect of the following factors on the acceptability judgements of L2 learners for uses of polysemous senses: semantic similarity between polysemous senses (sections 7.5, 8.4.3 and 9.5), intra-word frequency of polysemous senses (sections 7.5 and 8.4.2), and L1 influence (section 10.5). In this final discussion chapter I will draw together the themes and issues arising from these into three strands: factors affecting semantic similarity, possible uses of translation from the learners' L1 to their L2, and the implications of the findings for the construct of depth of L2 vocabulary knowledge.

One reason why it is important to establish the contribution of cognitive linguistic descriptions to the development of polysemous word knowledge, is that there have been several attempts to design L2 teaching methods informed by these descriptions. The results of those teaching methods have been inconsistent, with some researchers finding their method to be effective (Csábi, 2004; Verspoor & Lowie, 2003), while others finding more limited effectiveness (Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007). The inconsistent results call into question the validity of the common underlying theory: is the description of polysemy described by cognitive linguistics able to explain how L2 learners develop their knowledge of polysemous senses? I argued that it would be useful to see if the same theory could

271

explain L2 learners' knowledge of polysemous senses outside of specific teaching methods (see Section 3.9 on page # and Section 4.1 on pages 107-108). Such an explanation would support the underlying theory. A reason could then be sought for why the different teaching methods produced inconsistent results.

11.2 Factors affecting semantic similarity

The different senses of polysemous words can refer to very different real world objects and events. Despite these differences, the senses are said to form a coherent semantic category because of the abstract similarities between them. The results of the TPM have shown that L2 learners found senses more acceptable if they expressed semantic similarity to a core sense than if the meaning was inconsistent with the core sense. However, the effect of intra-word frequency makes it difficult to evaluate whether a sense was judged more acceptable because of its semantic similarity or because the participant had encountered this sense more frequently. The responses to the logical distractors can help to evaluate the strength of semantic similarity. The distractor items were unacceptable to native speakers and so the L2 learners could not have previously encountered the target word used in the sense of the logical distractor.

One possible explanation for the strength of semantic similarity is the concreteness of the polysemous sense or the concreteness of the context. An earlier analysis of concreteness did not provide significant results (see note 18 on page #). However, the concreteness ratings used in that analysis were associated with the word form and not with individual senses. A more qualitative analysis of the logical distractors shows that concreteness plays a strong role in whether the learners found the semantic similarity acceptable or not. Table 11.1 lists the four logical distractors with greatest proportion of 'False Alarms' compared to the five logical distractors with and least proportion of 'False Alarms'. The selection of these nine distractor items from the total of 20 was made to highlight any salient differences.

Table 11.1
A list of the logical distractors with most and least proportion of 'False Alarms'.

Acceptability to L2 learners	Item	Proportion of "False Alarms"
Most acceptable	The horse was black, with white marks on its *arms*.	67.16%
	The centre part of an apple is called its *heart*.	58.21%
	Our staff combine efficient service with a personal *hand*.	55.22%
	The house was old and its *cover* needed to be repaired.	53.73%
Least acceptable	I met her by *case* in Oxford Street.	23.88%
	My suitcase was so full I didn't have *air* for anything else	22.39%
	The phone *class* to telephone Europe is 2 Riyals per minute.	19.40%
	I thought she was upset because she had a sad *head*.	17.91%
	The cook put a *point* of salt into the soup.	17.91%

It seems apparent that the L2 learners found the logical distractors most acceptable when both the core sense and the context were highly concrete. For example, in the first distractor, the target word is *arm* and the context is the body of a horse; both referents are highly concrete. This is in contrast to those senses they found the least acceptable. For the target words *case, class* and *point* the core sense was more abstract. While the core sense of *head* is concrete, it was construed in the more abstract context of emotions. The core sense of *air* is not abstract, but as a gas it is less concrete than solid objects and thus less imageable. In the distractors judged least acceptable, there is less semantic consistency than among the most acceptable distractors. While it appears that concreteness makes a logical distractor more acceptable, it is not entirely clear what makes one less acceptable.

The effect of concreteness in the logical distractors aligns well with the prediction promoted by de Groot's Distributed Feature Model (1992). This model proposed that translation in the bilingual lexicon is mediated by meaning activation. The meaning of a word in either the L1 or the L2 is mapped to a set of semantic features in conceptual memory. Concrete meanings are likely to share more features across languages than abstract meanings. The model was developed to explain the effect of concreteness in translation latency, whereby concrete words were translated faster and more correctly than abstract words. The argument is that concrete words refer to objects that are shared across languages. In contrast, abstract words refer to concepts

273

which are more culturally and linguistically determined and are less likely to share parallel features.

The learners found the logical distractors more acceptable if they referred to concrete concepts through both the target words and the sentence context. According to the Distributed Feature Model, these concrete concepts are expressed by more semantic features in the learners' conceptual memory. The greater number of activated features makes the meaning of the sentence more imageable which facilitates an acceptable construal.

In contrast to the logical distractors, concreteness does not explain the responses to the acceptable uses which express semantic similarity. Consider the following two acceptable items which express semantic similarity to the core sense,

(1)   He's been in good *form* all this season. (Judged "Correct" by 53 of 67 participants)

(2)   Cloud *cover* in the morning should clear later. (Judged "Correct" by 27 of 67 participants)

The participants found (1) to be twice as acceptable as (2), and yet (2) is by far the more concrete sentence. It is likely that prior experience with the acceptable uses complicates the results, making it difficult to measure the effect of concreteness[35]. Furthermore, while concreteness might make the similarity to the core sense more apparent to the learner, it does not necessarily follow that the learner will find the use more acceptable. A learner might be able to understand the meaning of the sense because it is concrete and imageable, but reject it as unacceptable for other reasons. For example, a learner might construe "cloud cover" correctly but reject the use of *cover* as unacceptable because there is no physical contact between the clouds and

---

35. As reported in footnote 17 in on page #, an analysis using concreteness ratings from native speaker normative data did not find any statistically significant relationship in the TPM responses for the effect of concreteness.

the land below, or because clouds are not solid objects but ephemeral bodies in the sky.

The difficulty of determining why some senses are more acceptable than others is apparent in the unexpected findings for the metonym and metaphor senses. The senses expressing a metonym extension were proposed to express greater semantic similarity to the core sense than the metaphor extensions (see Section 9.4). This proposition of the greater similarity was due to the shared domain between metonyms and the core sense, as well as to the support of the results from Klepousniotou (2002) and Klepousniotou and Baum (2007). Against expectation, the L2 participants judged the metaphor senses as more acceptable than the metonyms.

One reason for the unexpected result is that the learners may have already learned the sense of the word through frequent access. Such frequent access could mean that the extended sense could have separate representation from the core sense. This proposition is supported by Murphy's discussion of the factors affecting shared vs. separate representation (Murphy, 2007). The first proposition is that the semantic content of a word is represented in some type of semantic space. If two uses of a word are similar to each other, then their representation in this semantic space will be adjacent or overlapping. Murphy argues that there are a couple of reasons why two uses of the same word could be represented separately in the semantic space. First, if the two uses are very distinct from each other then it is likely that there will be separate representation. Second, if two uses of a word are accessed frequently enough they may be separately represented despite their similarities. With frequent access, the speaker would have more opportunity to identify the details between the two uses and, in effect, emphasizing what is different between two similar uses.

The influence of frequent access could help explain the results to the following two items. These were the most acceptable metaphor sense and the most acceptable metonym sense to the participants:

(3)   Metaphor: English *classes* start at 5:15.
             (Judged "Correct" by 61 of 67 participants.)

275

(4)　　　　　　Metonym: Can you give me a *hand* to lift this?
　　　　　　　　(Judged "Correct" by 58 of 67 participants.)

While neither of these two senses is the core sense, it is likely that participants were familiar with both. The participants were all students and so the TEACHING sense of *class* was probably the most frequent in their experience. Likewise, the participants were also probably familiar with *hand* in the sense of HELP, since it is the second most frequent sense according to the Longman Dictionary of Contemporary English (LDOCE). According to Murphy's proposal, this familiarity would benefit from separate representation, thus countering the benefit of semantic similarity.

However, frequent access doesn't explain why the metaphor senses should have been judged as more acceptable than the metonym senses when the senses were controlled at the middle frequency level. One possibility is that the L2 participants recognised the semantic connection to the core sense in both the metonyms and metaphors, but rejected the metonym use as unacceptable. For example, in the following two items, the target word *body* was used as both a metonym and a metaphor.

(5)　　　　　　Metonym: The student *body* numbers 5000.
　　　　　　　　(Judged "Correct" by 13 of 67 participants.)

(6)　　　　　　Metaphor: There is now a large *body* of knowledge about childhood.
　　　　　　　　(Judged "Correct" by 24 of 67 participants.)

The metaphor sense was judged nearly twice as acceptable as the metonym sense. What is notable about these examples is that the metonym sense was the middle-frequency item and the metaphor sense was the low-frequency item. The reason that the metonym was judged as less acceptable cannot be attributed to corpus frequency. One explanation for its low acceptability is that the participants may have had a stronger set of restrictions against extending the use of the word within the same domain as the core sense. In contrast, when the word is construed metaphorically in a different domain from the core sense, the participants did not face the same restrictions and felt more free to interpret the meaning of the word.

276

As discussed in Section 9.3, Murphy (1997) found that native speakers judged figurative extensions of new language as more acceptable when the metonym was used conventionally rather than unconventionally. That finding may be relevant to explain the lower acceptability of the metonym senses as in the example in sentence (5). One can only speculate what those restrictions are, but it may be that the participants felt it was unacceptable for the same word to refer two distinct senses that could exist along side each other. For example, how can *body* refer to an individual and a group at the same time? Insight into these restrictions seems beyond the scope of the evidence gathered using the TPM. Nevertheless, the possibility that L2 learners find certain semantic extensions unacceptable opens up potential for further research into how L2 conceptual knowledge develops.

In sum, the learners' acceptability judgements broadly correspond to the model of polysemous senses as a radial category extending from a core sense to more peripheral senses. What isn't well understood are the factors which make one extended sense more similar to the core sense than another. Concreteness may play a role, but it is not the defining factor. Furthermore, based on the results, it appears that some senses categorised as metonyms are less acceptable than those categorised as metaphors. The distinction between metonyms and metaphors was based on whether the sense used the same domain as the core sense (metonym) or whether the sense was construed against a different domain from the core sense (metaphor). It may be more promising to compare the acceptability of conventional metonyms (*chicken* [ANIMAL] vs. *chicken* [MEAT]) to more idiosyncratic metonyms (*hand* [BODY] vs *hand* [PERSON])[36], or to compare metaphors with concrete domains to metaphors with more abstract domains. Future research in this area might profitably investigate this possibility.

---

36. The framework of conventional metaphors and metonyms was developed by Apresjan (1974).

11.3 Possible use of translation

As discussed in Section 10.5, the findings from the study on L1 influence indicated that the Arabic learners did not greatly benefit from L2 senses which shared translation with the L1 equivalent. This finding was somewhat at odds with the research on translation ambiguity, which provides evidence that a sense which shared the L2 word form with the translation equivalent would be more acceptable to bilinguals (Degani et al., 2011; Jiang, 2002, 2004; Paribakht, 2005). To address this difference between the findings and the literature, it is useful to consider two questions. First, should we expect L2 learners to use translation when interpreting the meaning of a familiar word in an unfamiliar sense, and second what would be the specific purpose of translation for the learners?

Questions concerning the interaction between L1 and L2 lexicons have often been investigated with reference to the Revised Hierarchical Model (RHM) of bilingual language processing (Kroll & Stewart, 1994). The RHM was designed as a model of L2 production, but it has also been used as a model of L2 comprehension[37]. The RHM predicts that at lower proficiencies, L2 access to conceptual memory will be mediated by the L1, but as learners become more proficient, they will tend to access the conceptual memory directly. As discussed in Section 11.2 (page 272), reduced translation latency due to the effect of concreteness supported the claim that conceptual memory can be directly accessed in the L2 (de Groot, 1992). While it might be argued that this claim is not applicable to the lower proficiency learners who took the TPM, it is important to consider two points. First of all, unlike much of the research supporting the RHM, the task on the TPM did not require the learners to make a translation. Translation might have been engaged through automatic processes or as an explicit task completion strategy, but whether translation was engaged or not, might have varied from learner to learner. Secondly, there is reason to believe that the learners were sufficiently proficient to directly access conceptual

---

37. For an extended discussion of the application of the Revised Hierarchical Model, see Brysbaert & Duyck, 2010 and Kroll, Van Hell, Tokowicz, & Green, 2010.

memory for the target words. The target word forms used in the TPM were highly frequent in the language, as were the other words in the sentence context of the items (see Chapter 4). While the learners were not expected to be familiar with all the senses of the target words, they were expected to be familiar with the word forms and at least one associated sense. For this reason, it seems likely that conceptual memory associated with the familiar word form would be accessed directly. The problematic issue is that if the polysemous sense was unfamiliar, then the stored conceptual memory of the word may not have been rich enough for the L2 participant to easily construe the meaning of the polysemous word within the context of the sentence.

Construal of an unfamiliar L2 polysemous sense is a situation that is not easily explained by the RHM. The RHM offers two possible routes for the semantic interpretation of a word form, either through the meaning of an L1 translation equivalent or through the direct retrieval of stored conceptual knowledge. However, neither route offers a satisfactory explanation for an unfamiliar polysemous sense. Since the polysemous sense is unfamiliar, the dominant meaning of the translation equivalent will not offer a satisfactory resolution to the ambiguity. In such a case, the learner might attempt to resolve the ambiguity by interpreting the sentence using a subordinate meaning of the L1 equivalent.

An example is useful to illustrate how the translation route might apply to an unfamiliar L2 polysemous senses. I will use French instead of Arabic, because the language is more widely understood by English speakers. In French, the word *couverture* can be used in the following ways,

(7)   "la couverture de son livre," (the cover of his book)

(8)   "la couverture de sa maison," (the roof of his house)

In these examples, the French shows ambiguity (*couverture)* while the English doesn't (*cover* vs *roof)*. We'll consider what problems this poses to an English speaker learning French.

If the learner attempts to use translation, then she will translate *couverture* into the equivalent L1 word form, which is likely *cover* in this case. This approach will satisfactorily map form to meaning in (7), the BOOK sense, but not in (8), the ROOF sense. It seems unlikely that the learner would explicitly run through the list of other senses of *cover* in English, 'music cover', 'insurance cover', 'bed covers', and attempt to apply those meanings to the sentence. As discussed in Chapter 10, Frenck-Mestre and Prince found that L2 semantic primes did not implicitly activate subordinate meanings of homophones with L2 learners (1997).

The other route proposed by the RHM is for the learner to directly retrieve conceptual knowledge from memory. This route fails at first because the learner's knowledge of COVERS does not apply to HOUSES. The implication is that the learner needs to be sufficiently flexible in her conceptual knowledge to extend the range of features she has mapped to *couverture* to allow her not only to construe the meaning correctly on this occasion, but to recall the same extended range of features the next time she attempts to interpret the meaning of *couverture*.

The findings from the responses to the TPM generally support direct retrieval from conceptual memory for senses that are different from the dominant meaning of the L1 equivalent. In the analysis of L1 influence in Chapter 10, I looked at whether the learners found English senses more acceptable if they translated into the same word as the Arabic equivalent of the core sense. When the senses were controlled for the metaphor extension, the Arabic learners of English did not benefit from shared translations between the L1 equivalent and the L2 sense. In other words, the learners found the different translations (like *couverture* [ROOF]) to be as acceptable as shared translations (like *couverture* [BOOK]). The results indicated that semantic similarity, not L1 influence, was the stronger factor influencing whether the learners found the polysemous sense acceptable or not.

As discussed in Section 10.5, this finding is supported by Kellerman's research (1986). According to Kellerman, the learner is likely to transfer the most prototypical sense of the L1 word to translate the ambiguous L2 word. Since Kellerman is looking

at transfer, his work is more applicable to productive use, i.e. how would the learner use her L1 to produce an L2 utterance. The findings from the TPM are more applicable to receptive understanding, i.e. how would the learner use her L1 to interpret an L2 utterance. The evidence from the TPM is that the prototypical sense (the core sense) is used as the L1 equivalent in L2 receptive tasks, in a similar way as Kellerman found it was used in L2 production.

However, evidence from other research in favour of L1 influence mitigates the strong argument of the single translation approach. In studies of translation ambiguity (Degani et al., 2011; Jiang, 2002, 2004), the L1 lexicon was shown to have a continued influence on the L2 lexicon. Given this evidence from translation ambiguity research in favour of L1 influence, why didn't the Arabic learners find the senses more acceptable if they shared translation with the L1 equivalent?

The translation ambiguity studies are different from the TPM in two important ways. First of all, these studies used a priming methodology that taps into how meaning is stored in the lexicon. Second, the subjects in the translation ambiguity research were either bilingual or advanced L2 speakers. What this research shows is that for advanced L2 speakers, activation of a word form in one language primes the equivalent word form in the other language. These speakers were expected to understand all the L2 language in the experiments, and they weren't expected to engage in explicit translation tasks for comprehension. In contrast, the Arabic learners who took the TPM were at an intermediate level of L2 proficiency. The TPM task measured how the language was processed rather than how the language was stored. This is a key difference between the TPM and many of the psycholinguistic studies. It relates to Sandra and Rice's distinction between how senses are structured and how they are processed (see Section 2.4 on page 34). It was expected that the learners might engage in explicit translation while taking the TPM to interpret some of the L2 polysemous senses. It seems likely that the translation ambiguity research is reporting on the interrelationship of the L1 and L2 lexicons, while the TPM

research provides insight on the interpretation process that intermediate learners make when they encounter unfamiliar polysemous senses.

It must be said that there were slightly more 'Hits' for shared translation senses than for senses which translated differently from the L1 equivalent (see Section 10.4 on pages 259-262). With a larger sample of items, this difference might have proved significant, indicating a small benefit for the shared translation condition. It could be that in some cases, two separate entries in the dictionary (the LDOCE) are actually stored as a single sense in the learners' L1 equivalent, while in other cases the senses, despite sharing the same L1 word form, are still stored separately. The variation between shared and separate storage in senses in the L1 could complicate the results differently from word to word and also from speaker to speaker.

11.4 Implications for depth of knowledge

The results of the TPM are also relevant to the distinction between vocabulary size (the number of words a speaker knows) and depth of vocabulary knowledge (how well the speaker knows the individual words). In a review of how the two constructs have been investigated, Schmitt (2014) notes that there have been conflicting reports about the usefulness of the distinction. For example, Vermeer (2001) found that a measurement of depth correlated highly with measurements of vocabulary size, indicating a lack of conceptual distinction between the two constructs. In contrast, Qian (1999, 2002) and Akbarian (2010) found that a measurement of vocabulary depth could explain differences in reading scores beyond measurements of vocabulary size. As Schmitt points out, the reason measurements of size and depth may converge or diverge depends largely on how the measurement of depth is conceptualised. The issue is important because an L2 learner could conceivably develop a large vocabulary through deliberate study of word forms and their meanings without being able to integrate that knowledge into communicative skills of comprehension or production. The question to address here is whether the Test of Polysemous Meanings (TPM), as a measurement of L2 polysemous knowledge, is

able to provide insight into L2 knowledge of vocabulary beyond measurements of vocabulary size.

Depth of vocabulary knowledge is commonly conceptualised by separating it into smaller elements. In Section 2.1.1, knowledge of polysemous words was identified in Nation's framework as part of the meaning category, under the aspect of concepts and referents. There are differing views about how distinct this knowledge is from the construct of vocabulary size in general. Schmitt notes that,

> [l]earning multiple meanings for the same word form [...] may not be much different than learning different words each with their own meaning. If this is true, it might be better to view knowledge of multiple meaning senses as a size measure instead of a depth measure. (Schmitt, 2014, p. 944)

In contrast, Crossley, Salsbury and McNamara (2010) contend that polysemous sense relationships correlate well with the richness of word knowledge, and that investigation of these relationships could add to insight to the development of lexical networks. Furthermore, a study by Malt and Sloman (2003) offers an example of how measures of vocabulary size are limited in relation to knowledge of conceptual development. They asked L2 learners to name a variety of common objects (containers and dishes) and found strong divergences from the naming practices of native speakers. Importantly, non-native speakers could only approach more native-like naming practices by using a fewer number of words. This indicates that development of conceptual knowledge in an L2 is not fully captured by measurements of vocabulary size.

In the study presented in Chapter 8, the L2 participants who took the TPM also took the Vocabulary Levels Test (VLT) (Nation, 1990; Schmitt et al., 2001). The VLT is a measurement of vocabulary size which associates knowledge of a word form with a single dominant meaning. The results from the test were not used in a correlational analysis because the VLT was not designed as an interval measure. Instead, the results were used to separate the participants into two groups, a group who scored over the 2000-word threshold and a group who scored under that threshold. Despite

the lack of a correlation analysis, there was a strong correspondence between the results of the VLT and those of the TPM when they were analysed for intra-word frequency. The group who scored above the 2000-word threshold on the VLT also indicated knowledge of more polysemous senses on the TPM. These results support the position that a measurement of polysemous meaning senses is akin to a measurement of size. However, there are three reasons why the TPM can claim to measure a conceptually different construct from a test of vocabulary size like VLT. These reasons are outlined below.

First of all, the results from the TPM provide evidence that semantic similarity is as strong a factor in learning a polysemous sense as intra-word frequency. When the results to items expressing a metaphor extension were compared, there was no significant difference between the high-frequency senses and the middle-frequency senses. This distinguishes the TPM from a test of vocabulary size. Measures of size take into account the frequency of the word form in a corpus, but beyond this they do not differentiate between different qualities of form-meaning links. In contrast, the data from the TPM was able to be analysed for the semantic similarity between senses. Such an analysis provided insight into the effect of these meaning relationships on the acceptability judgements of L2 speakers.

Second, while knowledge of polysemous senses was shown to develop according to the hypothesis of intra-word frequency (see Section 5.1), the comparison between the proficiency groups indicated that this development was not entirely uniform. This can be seen in the divergent results of the low and high proficiency groups in Figure 11.1 below. In keeping with the hypothesis of intra-word frequency, there was a significant difference between the acceptability judgements for the middle and low-frequency senses for the high-proficiency group; however, there was no significant difference the acceptability judgements between these intra-word frequency categories for the low-proficiency group. These results were discussed in Section 8.4.2. This finding is something of an anomaly because one might expect that the low-proficiency group would know proportionally fewer senses across each of the

three categories. Why did knowledge of the middle-frequency senses diverge so strongly between the high and low-proficiency groups?
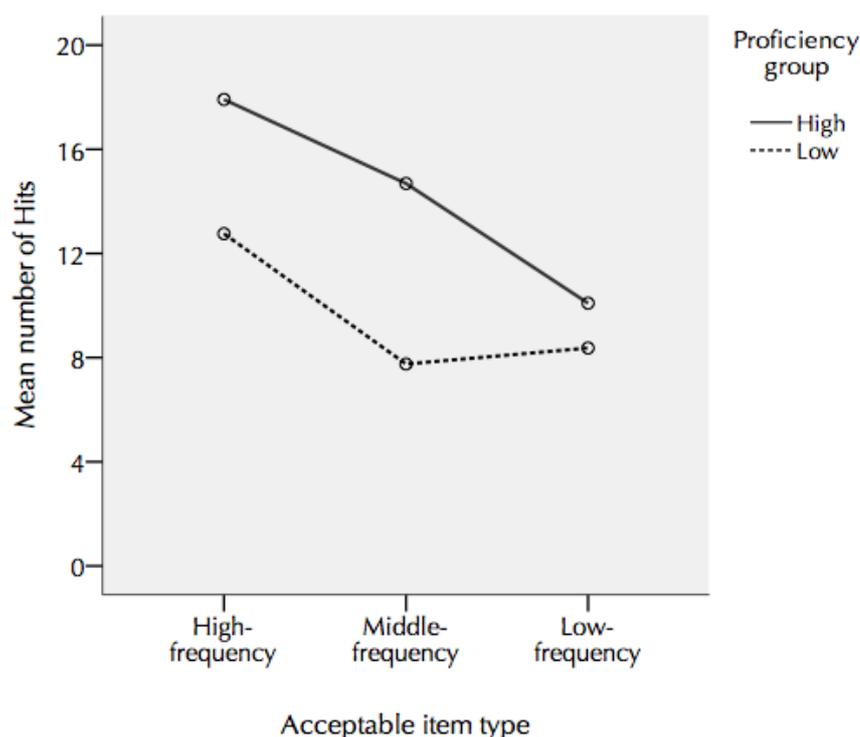
Figure 11.1
Mean number of 'Hits' to the acceptable items by proficiency group.

There are three possible explanations for this result. One possibility is that a learner might have to encounter a sense a certain number of times before she finds its use acceptable. By this argument, the learners in the low-proficiency group had only encountered the high-frequency senses frequently enough to overcome a learning threshold. In contrast, their responses to the middle and low-frequency senses were based largely on guesswork and inference. Consequently, their results to the middle and low-frequency sense were lower and comparable.

Another possibility is that there are individual differences that distinguish the two groups. L2 learners who develop larger vocabularies might be better attuned to how words are used in different contexts. This facility might enable them to learn

285

different senses of the same word form. In contrast, a third explanation is that some learners might have difficulty in seeing how a single L2 word form can have different meanings beyond the associated translation equivalent to their L1. This difficulty might limit their ability to learn a new sense despite its frequency of occurrence. A test of vocabulary size would not be able to reveal this finding because in these tests an individual word form is associated with a single meaning. As a result, tests of vocabulary size are insensitive either to learners who recognise subtle differences in meaning, or to learners who are limited by their L1 equivalent to only a single meaning.

The final reason why the TPM is distinct from tests of vocabulary size is because the results to the test can indicate whether the L2 learners have knowledge of unacceptable uses of the polysemous words. The TPM contains two types of distractor items which native speakers judged to be unacceptable. When the L2 learners' responses to these items are compared with the VLT scores, there appears to be a relationship between the learners' vocabulary size and their understanding of the restrictions on the use of polysemous words. The high-proficiency group responded with more 'Correct Rejections' to both the logical and illogical distractors. The relationship between the factors of vocabulary size and restrictions on use might be explained by the number of connections between words in the learners' lexicon. As vocabulary size increases, there are more connections between words that co-occur in the same domain (Meara & Wolter, 2004). A word might be said to be used unusually when it occurs among other words outside of its normal domain of connections. This may be why a learner in the high-proficiency group, with a larger vocabulary size, found the distractor items more unacceptable.

Knowledge of the semantic restrictions on a word's use is potentially interesting for the investigation of productive skills. This point can be explained by comparing an acceptable judgement to an unacceptable judgement. When a learner responds that a word is used acceptably, the interpretation is that she was able to make a meaningful construal of the word in agreement with her knowledge of how the word is used. However, there are two possible interpretations to explain why a learner responds

286

that a word is used unacceptably. Her response could indicate that she hasn't understood how the word is used, or it could also indicate that she wouldn't use the word like this, even if she could make sense of it. The latter explanation contains the more interesting implications for production.

It is generally accepted that productive use of vocabulary follows receptive understanding. From teaching experience, what a student can write lags behind what she can read. However, the research is inconsistent on what proportion of an L2 learner's receptive vocabulary is also productive (Schmitt, 2014). The use of the logical distractors on a test of polysemous word knowledge has the potential to offer insight into this question. To say, "I wouldn't use the word in that way," is an expression of productive knowledge. It may be that a learner with a stronger notion of when a word is used unacceptably is more likely to be able to use that word productively. However, to reliably investigate that possibility, the design of the TPM would need to be altered to allow for test takers to indicate a measure of confidence for their judgements. An L2 learner who correctly and confidently rejects a logical distractor would indicate that she has understood the restrictions of the word's use. The possibility is that this measure of 'Correct Rejections' may correlate with a measure of productive vocabulary. Whether a measure of confidence is a viable option is discussed at more length below, in Section 11.5.1.

It is reasonable to expect that other tests of vocabulary size will correlate quite strongly with measures of multiple meanings such as the TPM. However, as has been discussed here, the TPM was able to reveal certain qualities of L2 vocabulary knowledge not available through the measurement of vocabulary size. The quality of the meaning links to the polysemous word form was shown to affect whether the L2 learner found the polysemous meaning acceptable or not. This relationship is particular to polysemous words and distinguishes them from words associated with only a single meaning. Also, the responses to the TPM indicated that there may be individual differences between learners; some may be more attuned to recognising multiple meanings while others might find it difficult to break from the meaning associated with the L1 equivalent. Finally, there is the possibility that correct

rejections of distractor items may be useful to the investigation of productive knowledge of vocabulary. In each of these cases, the possible insight provided by the TPM needs to be supported by another separate measure, such as a test of vocabulary size or a test of productive vocabulary. A correlation analysis of the results of two such tests has the potential to be fruitful for further research into the development of L2 conceptual knowledge.

11.5 Data collection methods

The Test of Polysemous Meanings (TPM) was designed to elicit acceptability judgements from intermediate L2 learners of English on acceptable items categorised by frequency and distractor items categorised as either logical or illogical. The design of the instrument was presented in Chapter 4. Overall, the TPM was an effective method for investigating the knowledge of polysemous senses for learners at this level of proficiency. However, in the light of the findings using the TPM in the six studies presented in this thesis, it is appropriate to reflect on specific aspects of the test design, and to consider whether the test might be adapted to improve its fitness for purpose in future studies of this kind. In the following sections six aspects of test design are addressed.

11.5.1 Was the 'Correct/Incorrect' judgement sensitive enough?

The main task of the participants was to judge whether the target word was used acceptably in a given sentence by deciding if the use of the word was 'Correct' or 'Incorrect'. The benefit of this method was that participants were able to make their judgements quickly, allowing for more items to be tested without fatigue.

One limitation of the acceptability task was that it didn't allow for degrees of acceptability. For example, it might be predicted that the core sense would be the most acceptable use of the target word, that the inconsistent meanings would be considered the least acceptable, and that the metaphor extensions would be somewhere in between those extremes. Such a prediction is in keeping with studies of prototypicality effects (Rosch, 1978).

One alternative would have been to ask the participants to sort the five TPM items sharing the same target word from most acceptable to least acceptable. Such a sorting task would have been similar to the one conducted by Sandra and Rice (1995). The results would certainly have been richer than those offered by the TPM. For example, there was some question whether the participants responded to the low-frequency senses by guessing as they did to the logical distractors. This question could be tested empirically with the sorting task. If the learners did guess at the meaning of the low-frequency senses, then they would have been sorted similarly to the logical distractors. One drawback to the sorting task is that it would have taken more time for a participant to respond to a single test item, leading to fewer target words being tested over a given amount of time.

To offset the restrictions of the 'Correct/Incorrect' judgement, the TPM included a 'Don't Know' option if the participants felt they couldn't make a decision. This option was used less frequently than expected and its use varied greatly between individuals. It is not exactly clear why the participants didn't choose to use this option more often.

One alternative to the 'Don't Know' option would have been to allow the participants to rate their confidence of judgement on a Likert scale. This possibility was raised in the discussion on depth of vocabulary knowledge in the previous section (11.3.4). Unlike the 'Don't Know' option, the Likert scale could be a required aspect of the task. As with the sorting task, a Likert scale would allow the researcher to test predictions about the acceptability of different categories of senses. However, such a task might be unreliable if the participants misinterpreted the scale. For example, would the participants understand that strong confidence of acceptability should be rated the same as strong confidence of unacceptability? In other words, there is a risk that a measure of confidence could be misinterpreted as a measure of acceptability.

Overall, the acceptability judgements provided good insight into what senses the learners could construe meaningfully. However, the task provided less insight into

the construal process itself. One option for gaining insight into the participants' decision making processes would be to conduct a retrospective interview with the participants. Schmitt (1999) used this method to investigate what senses of a word learners knew and didn't know. One of the benefits of such an interview would be to highlight how the participants interpreted unfamiliar senses and what unexpected interpretations the participants made. The interviewer could ask such questions as, "Were you familiar with this use of the target word?", "How did you interpret the meaning of the target word in this sentence?", or "Why did you think the word was used incorrectly?"

 It is not entirely clear how effective the interview would be if conducted in the learners' L2. The construal processes may be too subtle for explicit and accurate elicitation in their L2, and the processes may even be too implicit for the participants to elaborate in their L1[38].

11.5.2 Could the range of target words be expanded?

The target words were selected because they represented a wide variety of polysemous senses. As presented in Section 4.3.1, there was a minimum of five distinct senses per target word excluding idiomatic uses. This was a necessary condition, but the target words were also selected according to two other parameters which could be altered depending on the objectives of the test. First, in order to ensure that the participants would know at least one sense of the target words, the words were chosen from the first 1000 most frequent words in English according to the Brown Corpus. Second, all the words selected were nouns.

First of all, with the design of the TPM it was important that the participants know at least one sense of the target words so that they would have some knowledge upon

---

38. An attempt was made to record L2 learners performing a concurrent think aloud while completing the TPM. However, the data did not prove to be rich enough to differentiate the decisions of one learner from another. This might have been due to the subtle nature of the processes involved or because the participants did not have the necessary proficiency in their L2 to satisfactorily express themselves to the level required by the think aloud protocol.

which to construe unfamiliar senses. However, the participants should not be expected to know all of the senses, or else there would be no way to differentiate one participant from another or one category of polysemous sense from another category. If participants with a larger vocabulary size were tested, then the test should be adapted to include words that occur less frequently in the language.

Secondly, other parts of speech could be selected. Restricting the selection of target words to nouns made the analysis of the results more straightforward. However, there has been substantial linguistic research into both verbs and prepositions as polysemous words. It is by no means a foregone conclusion that polysemous verbs and prepositions will be learned in a comparable way to nouns. Verbs are more strongly bound with their grammatical patterns and their concepts cannot be so easily isolated (Wray, 2015, p. 15). Also, prepositions rely far more on their context to elaborate their meaning than nouns do. While the design of the TPM has proven successful with nouns, it remains a question whether it would also work as well with other parts of speech.

11.5.3 Could the test items present intra-word frequency better?

The acceptable items themselves were created by selecting example sentences from the Longman Dictionary of Contemporary English (LDOCE) (2010). The LDOCE was chosen because its example sentences were written for intermediate learners of English and the senses of the words were listed according to intra-word frequency. The choice of senses represented the most frequent sense in the language, followed by a middle-frequency sense and a low-frequency sense. This process created items which were more appropriate to the level of the participants than the items in Verspoor and Lowie's instrument (2003), which had been used unsuccessfully in a replication study.

One limitation of the design for the acceptable items was the variation in intra-word frequency between target words. The senses of two different target words might have the same frequency of occurrence, but for one target word that sense could be classified as a high-frequency sense while the other could be classified as the middle-

frequency sense. As discussed in Chapter 9, the ideal solution would be to use interval data of actual frequency counts by sense and conduct a correlational analysis. The difficulty with this option is that, to my knowledge, there isn't a large scale corpus in the public domain that offers this data. Part of the problem, as shown by the variation between dictionaries, is that there isn't a consensus around what distinguishes one sense from another. The method employed by research in psycholinguistics is to use norms of sense-dominance in place of intra-word frequency counts[39]. However, an underlying assumption of using this type of normative data is that the participants know all of the senses. This is not the case with L2 learners who may know only a few senses of the word. For the time being, intra-word frequency may need to remain a limitation until more complete corpus data becomes available.

Hanks (2013) has proposed a lexicographical method of identifying polysemous senses that differs from the conceptual approach used in Chapter 9. This means that his method lacks equivalency with the studies which attempted to teach L2 polysemous words using a method taken from cognitive linguistics (Csábi, 2004; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). However, Hanks' method may be more amenable to the challenges posed by conducting intra-word frequency counts of different meanings in a large corpus. As discussed in Chapter 1, Hanks identifies distinct senses as different *norms* of use. A *norm* refers to how a specific word is used in a frequently recurring lexical pattern. The method can be briefly illustrated using the three acceptable items of the target word *heart* on the TPM:

(9)   Regular exercise is good for the *heart*.

(10)   Edith loved her boy with all her *heart* and soul.

(11)   We've got to take *heart* from the fact that we played well.

---

39. The most relevant of these norming studies is probably Durkin & Manning, 1989.

The word form *heart* occurred 177 times in the Brown corpus. It was preceded by the definite article 61 times, as in (9), and by a possessive pronoun 54 times, as in (10). However the word was never preceded by a verb, as in (11). According to Hanks' framework, the phrase "take heart" would not be considered a *norm* but an *exploitation* because it departs from the established norms of how *heart* is used. Furthermore, Hanks' approach also allows for the identification of metonyms and metaphors through the alteration of lexical sets. Sentence (9) includes the lexical unit "regular exercise". This unit might be part of a larger lexical set of [[PHYSICAL ACTIVITY]], which could also include 'jogging', 'dancing', etc. A metaphor could be achieved by replacing [[PHYSICAL ACTIVITY]] with [[SOCIALLY REWARDING ACTIVITY]], i.e., "Volunteer work is good for the heart."[40]

Hanks lays out a step by step procedure for identifying *norms* (Hanks, 2013, p. 119). Briefly, the first step involves identifying statistically significant collocates of the target word. A collocate is statistically significant if the target word co-occurs with another word more often than the frequency of the two words would have them occurring by chance (Church & Hanks, 1990). Considerations are taken for part of speech, the order of the words, and the number of words between the two words. Next, these collocates are sorted by clause roles and grouped for meaning discrimination to form lexical sets. Finally, the number of occurrences for the lexical sets are counted and exploitations of the norms are noted.

The benefit of Hanks' approach is that counting the different meaning senses appears to be more achievable than through the approach used in the field of cognitive linguistics. According to Hanks a Pattern Dictionary of the Present-Day Language will present the intra-word frequency in a large corpus as in the following example for the verb *spoil* (Hanks, 2013, p. 427):

1. Someone or something spoils an event that should be enjoyable: 60%
2. Something spoils a view or other pleasant location: 18%

---

40. However, for such a metaphor to be established as a *norm*, the lexical pattern would need to be regularly repeated in the language as attested by corpus analysis.

3. People spoil a child (i.e., overindulge him or her): 11%
4. Food spoils (= goes bad): 3%
5. Someone is spoiled for choice (idiom): 3%
6. A voter spoils a ballot paper in an election: 1%

Such an approach would be immensely valuable to a more accurate assessment of the intra-word frequency of polysemous senses.

11.5.4 Could the test items represent semantic categories in a more balanced way?

A challenge arose when the results were analysed for semantic similarity. In order to conduct the analysis, the items were reclassified according to four categories based around semantic similarity to a core sense. There was a greater number of senses representing the metaphor extension and very few items representing the inconsistent meaning category. Table 11.2 shows the number of items in each semantic category when organised by intra-word frequency.

Table 11.2
Number of occurrences of the acceptable items for each semantic category when categorised by intra-word frequency.

Intra-word frequency	Consistent	Metonym	Metaphor	Inconsistent
High frequency	13	0	7	0
Middle frequency	4	8	6	2
Low frequency	0	2	14	4

In order to compare the interaction of semantic similarity and intra-word frequency, it would be more desirable to have a balanced number of items represented across the intersection of the categories of semantic similarity and intra-word frequency, as indicated in Table 11.3.

Table 11.3
A balanced number of occurrences of the acceptable items for each semantic category when categorised by intra-word frequency.

Intra-word frequency	Consistent	Metonym	Metaphor	Inconsistent
High frequency	*n* items	*n* items	*n* items	*n* items
Middle frequency	*n* items	*n* items	*n* items	*n* items
Low frequency	*n* items	*n* items	*n* items	*n* items

The challenge for a balanced design is that meanings which are inconsistent with the core sense do not generally occur frequently in the language, whereas senses which are consistent with the core sense do not generally occur infrequently. It may be impossible to truly balance the items, but a researcher should be aware of this limitation at the outset and adapt accordingly.

11.5.5 Are there other ways of designing the logical distractors?

The rationale behind the design of the logical distractors was inspired by the 'false friend' errors made by L2 learners. As explained in Section 4.3.3, the meanings of the target words were looked up in bilingual dictionaries of languages which bore historical similarity to English. The meaning of the logical distractors was taken from meanings of equivalent words in these languages. Those meanings were not shared by the English target words.

The design of the logical distractors limits the possibility of replicating the study with speakers of languages related to English. If a learner was the speaker of a language that was used to define the logical distractor, then the learner might find the distractor acceptable because of transfer from her L1. It would be difficult to differentiate between L1 influence and the influence of semantic similarity.

Are there other ways of designing the logical distractors? One alternative would be to use the method employed by Rodd et. al. (2012). They transferred a single semantic feature of the target word's core sense to a novel, unacceptable meaning. For example, *ant* expresses the feature of SMALL. The researchers used this feature to create the novel sense of 'an *ant* is a small listening device'. It may be possible to use this method to create logical distractors. However, in this method, the meaning was established by a paragraph length description. It is unclear whether the meaning could easily be established in the sentence length context that the TPM items used.

Another method would be to exploit the meanings of near synonyms to the target word. While two synonyms can be used interchangeably in many contexts, there are cases where only one is acceptable. This situation could be exploited for the creation of a logical distractor. For example, the core senses of the words *air* and *atmosphere*

share many similarities. However, a*tmosphere* has the sense of "the pervading tone or mood of a place, situation, or creative work" (Stevenson, 2010). Because *air* cannot be used in this way, the following distractor could be created, "Their conversation passed in a friendly *air*". This distractor could be considered logical because it is employing the metaphor PHYSICAL ENVIRONMENT IS EMOTIONAL ENVIRONMENT.

One challenge with the synonym approach would be to find consensus among native speakers for the unacceptability of the distractors. There appears to be a tendency to look for meaning in any utterance (Wilson & Sperber, 2005). Using the synonym approach, it might be more challenging to create logical distractors that native speakers consistently find unacceptable.

11.5.6 Could the illogical distractors be improved?

The illogical distractors expressed no meaningful correspondence with any sense of the target word and they were judged as unacceptable by native speakers. The number of times they were judged as acceptable acted as a useful measure of the uncontrolled factors affecting the participants' judgements on the TPM. However, there was a fair amount of variation in the number of 'False Alarms' for these items (see Section 7.3.3). While it is not exactly clear why certain items were judged more acceptable than others, it would be useful to pilot a larger number of these items with the L2 learners and select those which were consistently identified as unacceptable. This extra step would ensure that there was less variation in how the participants responded to the illogical distractors.

11.6 How effective was the TPM as a measurement of learning?

The task of the TPM required the participants to make semantic judgements on the acceptable use of polysemous words. It is a question whether the results of those judgements can accurately describe the general learning progression of L2 learners. One way to address this concern is to see if the results satisfy predictions made by models of L2 vocabulary learning. Jiang's (2000) psycholinguistic model (see Section 2.2 on page 15) predicts that learners will use the lemma of the L1 translation equivalent for semantic knowledge about the word. (For the purpose of analysis, the

L2 core sense was considered to overlap with the L1 translation equivalent.) The results attest to this prediction because the TPM senses were judged more acceptable as they bore more semantic similarity to the core sense.

The model also predicts that new semantic content would be developed when the word was used in contexts where the translation equivalent would not be be used. This is because such contexts would be outside the normal application of the L1 lemma. The results for the logical distractors supported this prediction. The logical distractors were used in contexts which were unacceptable to both the L2 target word and the L1 translation equivalent. The participants judged the logical distractors as more acceptable than the illogical distractors, indicating that they interpreted these senses meaningfully. The assumption is that with repeated encounters to the new senses, the learners would develop new semantic content and establish the logical distractor as a learned sense.

One might question whether there is evidence of new semantic content, with reference Wolter's (2009) meaning-last model of vocabulary acquisition. Wolter argued that unfamiliar word forms would be learned as part of a larger lexical chunk before being isolated and interpreted for specific semantic content. This model could apply to unfamiliar senses of known polysemous words; however, the logical distractors, being novel uses of the word, would not express familiar collocations. It is unlikely that the participants had encountered before how the target words were used in the logical distractors, thus they couldn't be expected to be familiar with the collocations presented in the item.

Despite the evidence which supports learning, there are certain limitations to the experimental design which mitigates the strength of the conclusions. One issue is that the cross-sectional approach to the analysis assumes that all participants will develop their knowledge in similar ways. However, it could be that only some of the participants developed new semantic content for the polysemous words, while others are limited by the L1 lemma. In the analysis on the effect of vocabulary size, the low-proficiency group judged acceptability of the middle-frequency senses, low-

frequency senses and logical distractors about equally. This could indicate that these participants had difficulty making meaningful interpretations because they were limited by an L1 translation. Furthermore, the TPM provides no indication of the rate of learning. Some learners may take more time and practice than others to develop new semantic content.

A second limitation is that the design doesn't provide evidence for the appropriateness of the learners' interpretations. While the illogical distractors were used to make a quantitative adjustment to account for creative guessing, a qualitative measure of the learners' interpretations would be beneficial. For example, a retrospective interview would show how each participant interpreted the senses, whether the participants were consistent in their interpretations, and how appropriate their interpretations were.

A final limitation to the design is that judgement of 'Correct' and 'Incorrect' might not be nuanced enough to identify which items best facilitate learning. It could be that only very acceptable judgements led to learning, when a clear construal of the new sense is made. For example, the analysis might prove quite different with a judgement that offers four options: very acceptable, somewhat acceptable, somewhat unacceptable, very unacceptable. With such a design, the logical distractors might be judged as somewhat acceptable while the illogical distractors judged as very unacceptable. With such a result, one might be less inclined to attribute new learning based on the results to the logical distractors and more inclined to attribute greater unacceptability to the illogical distractors.

In sum, the TPM does provide evidence that learners develop new semantic content for unfamiliar senses that bear similarity to known semantic understanding. However, the evidence would be stronger with retrospective interviews and more nuanced judgement of acceptability.

11.7 Implications for research into teaching polysemous senses

While the TPM was designed to measure L2 learners' knowledge of polysemous words, the findings may have some relevance to second language teaching as well.

The first point of relevance is that in most cases the core sense is likely to be better known than other senses, and that the L2 learners are likely to find senses more acceptable if they share greater semantic similarity with the core sense. In those cases where a non-core sense is more frequent than the core sense, there is still a benefit to teaching the core sense because it provides coherence with other senses of the polysemous which have the potential to aid L2 learners in develop their knowledge of other senses. In Section 2.6, I reviewed several studies which investigated methods for teaching L2 polysemous words (Csábi, 2004, 1986; Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007; Verspoor & Lowie, 2003). They all looked at the benefit that knowledge of the core sense has for learning new senses. Furthermore, the centrality of the core sense to learner knowledge of polysemous words, further advocates the value to place the core sense first in L2 learners' dictionaries.

The second point of relevance from the findings is that the choice of target word and target sense will have an effect on the outcomes of any teaching intervention. Even when the polysemous senses were controlled for frequency and semantic similarity, there was still a good deal of variation between how acceptable the learners found one sense compared to another. One reason for this variation is the role that concreteness may play in the acceptability judgements, both the concreteness of the sense and the context in which the sense is used. Concreteness is one example of a factor affecting why a learner would find a semantic extension acceptable or not. It would be important to consider issues of acceptability when designing a teaching intervention, since the success of the teaching method may depend in part on the choice of the target word and senses.

The final point of relevance considers the degree of cognitive effort the TPM required of the participants. The task on the TPM asked the learners to make straightforward decisions about acceptability. The possible lack of cognitive nuance of the TPM might have had an effect on the results. With more cognitive effort, the participants might have recognised more semantic similarity and found the logical distractors more acceptable. If this was the case then it might be appropriate to ask

299

how easy or difficult it was for a learner to recognise the semantic similarity between an extended sense and a core sense. The learners who are more able to make this connection might be more successful at learning new senses.

Several studies, reviewed in Section 2.6, investigated whether teaching tasks using image schemas were more effective than translation tasks. The success of the image schema task appears to be related to the cognitive effort it requires of the learner. In Csábi's study (2004), the learners made subtle semantic distinctions to complete a gap-fill task. In other studies (Khodadady & Khaghaninizhad, 2012; Morimoto & Loewen, 2007), the learners read a description of how an image schema could define the meaning of the words. While some extensions may be readily available to the learners, depending for example on factors like concreteness, other semantic extensions may require more reflection and cognitive effort to be meaningful.

Finally, little is known about how individual differences between learners affect successful inferencing of unfamiliar polysemous senses. It might be the case that some learners exploit explicit strategies to interpret the meaning of an unfamiliar sense. If so, then these strategies could be formalised into specific teaching methods. These strategies may be more successful than those derived from the practice used in cognitive linguistics to identify different polysemous senses.

Chapter 12: Conclusion

At the start of this thesis, I set out to investigate whether the findings of Verspoor and Lowie (2003) could be confirmed through an approximate replication. In their study, Verspoor and Lowie investigated an L2 learning strategy for unfamiliar senses of polysemous words. The findings of the replication did not support those of the original, and I attributed the reason for this difference to the lower proficiency level of the participants who took part in the replication. As a result, I developed a new instrument, more appropriate to the proficiency of the participants, to investigate the knowledge L2 learners have of polysemous senses and the factors affecting their knowledge. As a way of conclusion, I will present some main findings and some implications for further research.

An analysis of the findings for L1 influence showed that the participants did not find senses significantly more acceptable if an equivalent word in their L1 shared similar form to meaning mappings. This indicated that L2 learners do not generally map each meaning of a word in their L2 to a specific translation in their L1. If translation is used, it is more likely to an L1 equivalent word whose meaning can be generalised across different contexts and uses. In contrast, vocabulary size was found to be a stronger indicator of whether the learners found a polysemous word acceptable or not according to native speaker norms. Specifically, as the learners knew more words in their L2, not only were they shown to know more polysemous senses, but they were better able to identify unacceptable uses of a word.

Turning from the differences between the learners, there were also lexical factors affecting why different senses would be judged more or less acceptable. Initially, the intra-word meaning frequency of a polysemous sense was shown to be a good predictor of whether L2 learners judged a word to be acceptable or not. This finding was complicated by the relationship between intra-word frequency and semantic similarity. The senses expressing high intra-word meaning frequency were often consistent with the core sense and senses expressing low intra-word meaning frequency were often inconsistent with the core sense. It was argued that the

301

combination of intra-word meaning frequency and semantic similarity was a stronger predictor of acceptability than simply frequency alone.

Some considerations should be taken into account when creating an instrument that develops upon the findings of this study. While the extended senses in this study were categorised as either metonyms or metaphors, it would be useful to consider ways of refining that categorisation. First of all, the metaphor extensions were linked to the core sense by image schemas, but a consideration of how different contexts play a role in the degree of semantic similarity may lead to a better understanding of what makes a figurative sense more or less acceptable. Secondly, it would also be worthwhile to consider whether a metonym extension is conventional or unconventional. For example, a conventional extension from *chicken* [ANIMAL] to *chicken* [MEAT] is likely more acceptable than an unconventional extension such as *body* [PERSON] to *body* [GROUP]. This method would align well with Hanks' (2013) approach of distinguishing polysemous senses based on recurring lexical patterns.

In this thesis I have demonstrated that polysemy can be used not only as a gauge of learner competence, but as a means of revealing information about storage and processing in the learner's developing L2 lexicon. The research method presented in this thesis has potential to be developed and extended in order to further enrich our understanding of the relationship of conceptual knowledge to L2 lexical competence.

References

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

Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/ EAP learners. *System, 38*(3), 391-401. doi:10.1016/j.system.2010.06.013

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark: International Reading Association.

Appel, R. (1998). Linguistic minority children's knowledge of multiple word meanings; Polysemy and the testing of lexical skills. *ITL. Institut voor Togepaste Linguistik*, 119-20), 79-89.

Apresjan, J. D. (1974). Regular polysemy. *Linguistics, 12*(142). doi:10.1515/ ling.1974.12.142.5

Ba'albaki, M. (2002). *al-Mawrid: a modern Arabic-English dictionary*. Beirut: Dar El-Elm Lilmalayin.

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

Beréndi, M., Csábi, S., & Kövecses, Z. (2008). Using conceptual metaphors and metonymies in vocabulary teaching. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 65-100). Berlin: Mouton de Gruyter.

Boers, F. (2000a). Enhancing metaphoric awareness in specialised reading. *English for Specific Purposes, 19*(2), 137-147.

Boers, F. (2000b). Metaphor awareness and vocabulary retention. *Applied Linguistics, 21*(4), 553-571.

Boers, F. (2013). Cognitive linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching, 46*(2), 208-224.

Boers, F., & Demecheleer, M. (1998). A Cognitive Semantic Approach to Teaching Prepositions. *ELT Journal, 52*(3), 197.

Boers, F., Demecheleer, M., & Eyckmans, J. (2004). Etymological elaboration as a strategy for learning idioms. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 53-78). Amsterdam: John Benjamins Publishing.

Boers, F., & Lindstromberg, S. (2007). Cognitive linguistic applications in second or foreign language instruction: Rationale, proposals, and evaluation. In G.

303

Kristiansen (Ed.), *Cognitive linguistics: Current applications and future perspectives*. Berlin: Mouton de Gruyter.

Boers, F., & Lindstromberg, S. (2008). Opening chapter: How cognitive linguistics can foster effective vocabulary teaching. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 1-61). Berlin: Walter de Gruyter.

Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: The case of the Euralex French tests. *Applied Linguistics*, *21*(4), 490.

Bowerman, M., & Choi, S. (2001). Shaping Meanings for Language: Universal and Language Specific in the Acquisition of Spatial Semantic Categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475-511). Cambridge: Cambridge University Press.

British Museum. (n.d.). Retrieved March 10, 2012 from http://www.britishmuseum.org/explore/highlights/highlight_objects/aes/g/glass_bottle_in_the_form_of_a.aspx

Brown, S. W. (2008). Polysemy in the mental lexicon. *Colorado Research in Linguistics*, *21*, 1-12.

Browne, C., Culligan, B., & Phillips, J. (2013). The New General Service List. Retrieved from http://www.newgeneralservicelist.org

Brugman, C., & Lakoff, G. (1988). Cognitive topology and lexical networks. Lexical ambiguity resolution. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology & artificial intelligence* (pp. 477-507). Palo Alto: Morgan Kaufman.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(03), 359-371. doi:10.1017/S1366728909990344

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 1-8.

Cambridge Advanced Learner's Dictionary. (2011). Retrieved 30 September, 2011 from http://dictionary.cambridge.org/dictionary/british/

Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In *Semantics: Theory and application* (pp. 181-206). Washington: Georgetown University Press.

CEFR. (n.d.). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Retrieved February, 2016 from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Church, K. W., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Comput. Linguist.*, *16*(1), 22-29.

Clausner, T. C., & Croft, W. (1999). Domains and image schemas. *Cognitive Linguistics*, *10*, 1-32.

Cobb, T. (n.d.). Web Vocabprofile: an adaptation of Heatley, Nation & Coxhead's (2002) Range. Retrieved 30 September, 2011 from http://www.lextutor.ca/

Coleman, L., & Kay, P. (1981). Prototype semantics: The English word lie. *Language*, *57*(1), 26-44.

Condon, N. (2008). How cognitive linguistic motivations influence the learning of phrasal verbs. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 133-158). Berlin: Mouton De Gruyter.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, *34*(2), 213-238. doi:10.2307/3587951

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684.

Croft, W. (1998). Linguistic evidence and mental representations. *Cognitive Linguistics*, *9*, 151-174. doi:10.1515/cogl.1998.9.2.151

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Cronbach, L. J. (1943). Measuring knowledge of precise word meaning. *Journal of Educational Research*, *36*, 528-534.

Crossley, S. A., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in english second language speakers. *Language Learning*, *60*(3), 573-605.

Csábi, S. (2004). A cognitive linguistic view of polysemy in english and its implications for teaching. In M. Achard & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 233-256). Berlin: De Gruyter Mouton.

de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001-1018. doi:doi:10.1037/0278-7393.18.5.1001

Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, *23*(1), 18-28. doi:10.1080/20445911.2011.445986

Degani, T., & Tokowicz, N. (2010a). Ambiguous words are harder to learn. *Bilingualism-Language and Cognition*, *13*(3), 299-314. doi:10.1017/S1366728909990411

Degani, T., & Tokowicz, N. (2010b). Semantic ambiguity within and across languages: An integrative review. *Quarterly Journal of Experimental Psychology*, *63*(7), 1266-1303. doi:10.1080/17470210903377372

Dronjic, V., & Helms-Park, R. (2014). Fixed-choice word-association tasks as second-language lexical tests: What native-speaker performance reveals about their potential weaknesses. *Applied Psycholinguistics*, *35*(1), 193-221. doi:10.1017/S014271641200029X

Durkin, K., & Manning, J. (1989). Polysemy and the subjective lexicon: Semantic relatedness and the salience of intraword senses. *Journal of Psycholinguistic Research*, *18*, 577-612.

Ellis, N. C. (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Elston-Güttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, *52*(2), 256-283. doi:10.1016/j.jml.2004.11.002

Elston-Güttler, K. E., & Williams, J. N. (2008). First language polysemy affects second language meaning interpretation: evidence for activation of first language concepts during second language reading. *Second Language Research Second Language Research*, *24*(2), 167-187. doi:10.1177/0267658307086300

eNotes Study smarter. (2012). Retrieved February 25, 2012 from http://www.enotes.com/scissors-reference/scissors

Fillmore, C. J. (1975). *An alternative to checklist theories of meaning*. Proceedings from the Annual Meeting of the Berkeley Linguistics Society, Berkley.

Fillmore, C. J. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Linguistic structures processing* (pp. 55-88). Amsterdam: North Holland.

Fillmore, C. J. (1982). Frame semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111-137). Seoul: Hanshin Publishing Co.

Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language*, *67*(4), 407-425. doi:10.1016/j.jml.2012.07.010

Francis, W. N., & Kucera, H. (1979). Brown corpus manual. *Brown University*, *15*.

Frenck-Mestre, C., & Prince, P. (1997). Second Language Autonomy. *Journal of Memory and Language*, *37*(4), 481-501. doi:10.1006/jmla.1997.2526

Gao, L., & Meng, G. (2010). A study on the effect of metaphor awareness raising on Chinese EFL learners' vocabulary acquisition and retention. *Canadian Social Science*, *6*(2), 110-124.

Gathercole, V., & Moawad, R. (2010). Semantic interaction in early and late bilinguals: All words are not created equally. *Bilingualism: Language and Cognition*, *13*(04), 385-408.

Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, *4*(3), 223-272.

Geeraerts, D. (2007). Where does Prototypicality Come From? In V. Evans, B. Bergen, & J. Zinken (Eds.), *The cognitive linguistics reader* (pp. 169-186). London: Equinox Publishing Ltd.

Goddard, C. (2000). Polysemy: A problem of definition. In Y. Ravin & C. Leacock (Eds.), *Polysemy: Theoretical and computational approaches* (pp. 129-151). Oxford: Oxford University Press.

Greidanus, T., & Nienhuis, L. (2001). Testing the Quality of Word Knowledge in a Second Language by Means of Word Associations: Types of Distractors and Types of Associations. *The Modern Language Journal*, *85*(4), 567-577. doi:10.1111/0026-7902.00126

Gutiérrez Pérez, R. (2008). A cross-cultural analysis of heart metaphors. *Revista Alicantina de Estudios Ingleses*, *21*, 25-56. doi:10.14198/raei.2008.21.03

Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Cambridge, MA: MIT Press.

Heatley, A., Nation, P., & Coxhead, A. (1994). Range. Retrieved 18 September, 2016 from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Huang, L. (2004). *An exploratory study on the assessment of semantic knowledge of English polysemy*. Proceedings from International Conference on English Instruction and Assessment.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47-77. doi:10.1093/applin/21.1.47

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*(04), 617-637. doi:10.1017/S0272263102004047

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*(3), 416-432.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, reason and imagination*. Chicago: University of Chicago Press.

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

Kellerman, E. (1986). An eye for an eye: Crosslinguistic constraints on the development of the L2 lexicon. In E. Kellerman & M. S. Smith (Eds.), *Crosslinguistic influence in second language acquisition* (pp. 35-48). Oxford: Pergamon Press.

Khodadady, E., & Khaghaninizhad, M. S. (2012). Acquisition of french polysemous vocabularies: Schema-based instruction versus translation-based instruction. *Porta Linguarum*, *17*, 29-46.

Klein, D., & Murphy, G. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*(2), 259-282. doi:10.1006/jmla.2001.2779

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*, 205-223.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1-24. doi:10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology-Learning Memory and Cognition*, *34*(6), 1534-1543. doi:10.1037/a0013012

Kövecses, Z., & Szabó, P. (1996). Idioms: A view from cognitive linguistics. *Applied Linguistics*, *17*(3), 326-355.

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*(2), 149-174. doi:10.1006/jmla.1994.1008

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373-381. doi:10.1017/S136672891000009X

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (1). Stanford: Stanford University Press.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Béjoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: MacMillan.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20-34). Cambridge: Cambridge University Press.

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

Lindstromberg, S. (1996). Prepositions: Meaning and method. *ELT Journal*, *50*(3), 225-236.

Longman Corpus Network. (n.d.). Retrieved February, 2017 from http://www.pearsonlongman.com/Dictionaries/corpus/index.html

Longman Dictionary of Contemporary English. (2010). Harlow: Pearson Longman.

MacArthur, F., & Littlemore, J. (2008). A discovery approach to figurative language learning with the use of corpora. In F. Boers & S. Lindstromberg (Eds.), *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology* (pp. 159-188). Berlin: Mouton de Gruyter.

MacLennan, C. H. G. (1994). Metaphors and prototypes in the learning teaching of grammar and vocabulary. *International Review of Applied Linguistics*, *32*(2), 97-110.

Malt, B. C., & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, *6*(1), 47-67. doi:10.1017/S1366728903001020

Mandler, J. M. (2005). How to build a baby III: Image schemas and the transition to verbal thought. In B. Hampe (Ed.), *Image schemas in cognitive linguistics* (pp. 137-163). Berlin: Mouton de Gruyter.

Meara, P. (1992). EFL Vocabulary Tests. Retrieved September 30, 2011 from http://www.lognostics.co.uk/vlibrary/meara1992z.pdf

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142-154. doi:10.1177/026553228700400202

Meara, P., & Wolter, B. (2004). V_Links: Beyond vocabulary depth. In K. Haastrup & B. Henriksen (Eds.), *Angles on the English-speaking world* (pp. 85-96). Copenhagen: Museum Tusculanum Press.

Merriam-Webster's Advanced Learner's English Dictionary. (2008). Retrieved March 10, 2012 from http://www.learnersdictionary.com

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. L., & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use* (pp. 57-78).

Morimoto, S., & Loewen, S. (2007). A comparison of the effects of image-schema-based instruction and translation-based instruction on the acquisition of L2 polysemous words. *Language Teaching Research*, *11*(3), 347-372.

Mumsnet Talk. (n.d.). Retrieved March 10, 2012 from http://www.mumsnet.com/

Murphy, G. (1997). Polysemy and the creation novel word meanings. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Conceptual structures and processes in creative thought* (pp. 235-265). Washington DC: American Psychological Association.

Murphy, G. (2007). Parsimony and the psychological representation of polysemous words. In M. Rakova, G. Pethö, & C. Rákosi (Eds.), *The cognitive basis of polysemy: New sources of evidence for theories of word meaning* (pp. 47-70). Frankfurt am Main: Peter Lang.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In I. S. P. Nation & N. Schmitt (Eds.), *Vocabulary Description, Acquisition and Pedagogy* (pp. 6-39). Cambridge: Cambridge University Press.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguist Philos*, *3*(2), 143-184. doi:10.1007/BF00126509

OED Online. (2011). Retrieved 1 October, 2011 from http://www.oed.com

Oxford Advanced Learner's Dictionary. (2011). Retrieved March 10, 2012 from http://oald8.oxfordlearnersdictionaries.com

Oxford Spanish Dictionary. (2011). Retrieved 30 September, 2011 from http://oxfordlanguagedictionaries.com/

Oxford-Duden German Dictionary. (2011). Retrieved 30 September, 2011 from http://oxfordlanguagedictionaries.com/

Oxford-Hachette French Dictionary. (2011). Retrieved 30 September, 2011 from http://oxfordlanguagedictionaries.com/

Oxford-Paravia Italian Dictionary. (2011). Retrieved 30 September, 2011 from http://oxfordlanguagedictionaries.com/

Paribakht, T. S. (2005). The influence of first language lexicalization on second language lexical inferencing: A study of farsi-speaking learners of english as a foreign language. *Language Learning*, *55*(4), 701-748.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review/La revue canadienne des langues vivantes*, *56*(2), 282-308.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language learning*, *52*(3), 513-536.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355-371.

Read, J. (1994, July). *Refining the word associates format as a measure of depth of vocabulary knowledge*. Proceedings from 19th Annual Congress of the Applied Linguistics Association of Australia, Melbourne, Australia.

Read, J. (1995, March). *Validating the word associates format as a measure of depth of vocabulary knowledge*. Proceedings from 17th Language Testing Research Colloquium, Long Beach, CA.

Read, J. (1997). Vocabulary and testing. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 303-320). Cambridge: Cambridge University Press.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, *10*(1), 77-89.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: effects of semantic relatedness. *Memory & Cognition*, *40*(7), 1095-1108. doi:10.3758/s13421-012-0209-1

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245-266. doi:10.1006/jmla.2001.2810

Rosch, E. (1978). Principles of categorization. In B. B. Lloyd & E. Rosch (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, N.J.: L. Erlbaum Associates.

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. Albany: SUNY Press.

Rundblad, G., & Annaz, D. (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547-563. doi:10.1348/026151009X454373

Sandra, D., & Rice, S. (1995). Network analyses of prepositional meaning: Mirroring whose mind—the linguist's or the language user's? [1]. *Cognitive Linguistics*, *6*, 89-130. doi:10.1515/cogl.1995.6.1.89

Schmitt, N. (1994). Vocabulary testing: Questions for test development with six examples of tests of vocabulary size and depth. *Thai TESOL Bulletin*, *6*(2), 9-16.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*(2), 281-317.

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, *16*(2), 189-216. doi:10.1177/026553229901600204

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913-951. doi:10.1111/lang.12077

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55-88.

Shillaw, J. (2009). Putting yes/no tests in context. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners: papers and perspectives in honour of Paul Meara* (pp. 13-24). Bristol: Multilingual Matters.

Skutnabb-Kangas, T. (2000). *Linguistic genocide in education or worldwide diversity and human rights?* London: Routledge.

Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, *159*, 11-24. doi:http://dx.doi.org/10.1016/j.cognition.2016.10.019

Srinivasan, M., & Barner, D. (2013). The Amelia Bedelia effect: World knowledge and the goal bias in language acquisition. *Cognition*, *128*(3), 431-450. doi:http://dx.doi.org/10.1016/j.cognition.2013.05.005

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, *157*, 124-152. doi:http://dx.doi.org/10.1016/j.lingua.2014.12.004

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, *62*(4), 245-272. doi:http://dx.doi.org/10.1016/j.cogpsych.2011.03.002

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*(2), 139-152. doi:10.1080/09571730802389975

Stallman, A. C., Pearson, P. D., Nagy, W. E., Anderson, R. C., & Garcia, G. E. (1995). *Alternative approaches to vocabulary assessment* (Tech Rep. No. 607). Urbana-Champaign, IL: Center for the Study of Reading, University of Illinois.

Stevenson, A. (Ed.). (2010). *Oxford Dictionary of English*. Oxford: Oxford University Press.

Summers, D. (1996). Corpus lexicography - the importance of representativeness in relation to frequency. *Longman Language Review*, *3*, 6-9.

Survey Monkey. (2011). Retrieved September 30, 2011 from http://www.surveymonkey.com/

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*(1), 49-100. doi:10.1207/s15516709cog1201_2

Taylor, J. R. (2012). Polysemy. In *The mental corpus: How language is represented in the mind* (pp. 219-244). New York: Oxford University Press.

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, *4*(3), 273-290.

Tuggy, D. (1999). Linguistic evidence for polysemy in the mind: A response to William Croft and Dominiek Sandra. *Cognitive Linguistics*, *10*(4), 343-368.

Tyler, A., Mueller, C. M., & Ho, V. (2010). Applying cognitive linguistics to instructed L2 learning: The English modals. *AILA Review*, *23*(1), 30-49.

Van Herwegen, J., Dimitriou, D., & Rundblad, G. (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities*, *34*(4), 1300-1311. doi:http://dx.doi.org/10.1016/j.ridd.2013.01.017

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: the same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457-479.

Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics*, *14*(4), 344-363.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217-234.

Verspoor, M., & Lowie, W. (2003). Making Sense of Polysemous Words. *Language Learning*, *53*(3), 547-586.

Wesche, M. B., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, *53*(1), 13-40.

West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.

Wierzbicka, A. (1996). *Semantics: Primes and Universals*. Oxford: Oxford University Press.

Wilson, D., & Sperber, D. (2005). Relevance theory. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607-632). Oxford: Blackwell. doi:10.1111/b.9780631225485.2005.00029.x

Wittgenstein, L. (1980). *Philosophical Investigations*. Oxford: Wiley-Blackwell.

Wolter, B. (2009). Meaning-last vocabulary acquisition and collocational productivity. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical Processing in Second Language Learners: Papers and Perspectives in Honour of Paul Meara* (pp. 128-140). Clevedon: Multilingual Matters.

Wray, A. (2015). Why are we so sure we know what a word is? In J. Taylor (Ed.), *The Oxford handbook of the word* (pp. 725-750). Oxford: Oxford University Press.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language learning and communication*, *3*(2), 215-229.

Yahoo Answers! (n.d.). Retrieved October 02, 2012 from http://answers.yahoo.com/

Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology, 33*(2), 251-256.

Appendices

Appendix 1 Replication Study

1.1 Replication Test S1 - Core sense cue

Vocabulary Study – Test S1   Name:

            Class:

            TOEFL (or IELTS):

Directions:

In this study, you will find pairs of sentences: a and b. The underlined word in these sentences is the same but the meanings are different. For sentence a. you will find the Arabic translation of the word. For sentence b you will guess what the Arabic meaning of the word is. For example:

ex.   a.   She is as light as a feather.   light is خفيف

       b.   That book is very light reading.   light is

In sentence b. light has a different meaning from sentence a.

In sentence b. light means "سهل", so you would write "سهل" in the answer space: "light is _____سهل_____"

Questions:

1   a. Boost me up this tree and I will get you an apple.   يرفع boost is

     b. The landlord will boost the rent.   boost is

2   a. We grappled with him and took the gun away from him.   يتصارع، يتعارك grapple is

     b. He now watches many of his students grapple with the same dilemma.   grapple is

3   a. What's that bulge in your pocket?   بروز، انتفاخ bulge is

     b. A breakway dunk by Raheed Wallace ended a 120 run by the Bullets that gave bulge is
        them their 5-point bulge.

4   a. Using a spoon, skim off any air bubbles from the top of the custard mixture. يقشد، يزيل رغوة skim is

     b. A moment more and the helicopter rises along the slope of a mountain, up and skim is
        over the tree line until we are skimming over peaks of rock that are jagged as
        flints.

5   a. Wind it round the screws until the wire is taut.   مشدود taut is

     b. Eyes blinking, showing no signs of being emotionally taut, her husband looked taut is
        like an ordinary man defending the ordinary lies he had made up to hide an
        ordinary affair.

6   a. Explosions scorch floors and shatter windows.   يحطم، يهشم shatter is

     b. I hope to make you laugh, but I also hope to shatter the ideas you have, he shatter is
        said. You're forced to question your own prejudices.

7 a. Then her mother nudged her and pointed to the seven helicopters flying overhead.

   nudge is (يلكز(بالمرفق

   b. In that race, there were other people who wanted to run, but the party leaders nudged them not to enroll, because ``they didn't have a chance.'' ``It wasn't as formal a process,'' Mr. Byrne said.

   nudge is

8 a. The gardener set fire to the piles of weeds he had raked up.

   rake is يجمع أعشاب ,يدمِّ

   b. Relief foundations raked in $13 million last year.

   rake is

9 a. The principle in clocks is that a number of wheels, locked together by cogs, are forced to turn round.

   cog is ترس مسنن

   b. Clark has been a major cog in the St. Louis attack since being acquired in a July 31 trade with the Baltimore Orioles.

   cog is

10 a. His father originally sent him solid golden nuggets.

   nugget is (كتلة صلبة (من معدن ثمين

   b. The new LSS does that with a choice of V6 engines and with a body, interior and suspension that make the car a true nugget in today's rushing stream of fancy cars.

   nugget is

11 a. My mother hates gutting fish.

   gut is يخرج أحشاء

   b. But while President Clinton fought against the gutting of environmental laws, he offered little of no resistance on civil liberties.

   gut is

12 a. An owl hooted among the pines.

   hoot is (ينعق (البوم

   b. We drove smiling, hooting, fists thrust through the open windows in our decorated cars through the streets.

   hoot is

13 a. Gaudi liked to work in metal and may have actually forged his own sword.   يصنع بطرق معدن forge is

    b. Mr. Murstein said he was surprised that Disney, which has forged a close   forge is
        relationship with the Giuliani administration as it has invested in Times
        Square, had been turned down by the city.

14 a. We pegged a tent to the ground for the kids.   يثبت بوتد . يشد أوتاد peg is

    b. But he added that he was still considering pegging his country's currency to   peg is
        the dollar, a move the I.M.F. strongly opposes.

15 a. She confessed last year that she had smothered the baby because she could   smother is يخنق
        not stand its crying.

    b. Nearly four months after the agreement, Mayor Giuliani smothered a strike by   smother is
        lawyers for the Legal Aid Society.

16 a. Glaciers are usually covered with perennial snow.   perennial is مدار العام

    b. Losing weight--the perennial New Year's resolution may make you more   perennial is
        attractive and less prone to disorders like diabetes and high blood pressure.

17 a. The salmon fight their way back up the river to spawn.   spawn is يضع بيض

    b. The Democratic convention is spawning with secret agents who have to   spawn is
        protect the President.

1.2 Replication Test S3 - Extended sense cue

Directions:

Write the Arabic translation of what the underlined word means in the sentence.

ex.   That book is very light reading.   light is

In the example, light means "سهل", so you would write "سهل" in the answer space: "light is _____سهل_____"

Questions:

1   a.   The tax cut can boost the economy.   boost is يدعم ، يحسن

    b.   The landlord will boost the rent.   boost is

2   a.   But those explanations have not made it any easier for them to grapple   grapple is يواجه
          with the stark reality of losing four young men in such a terrible fire.

    b.   He now watches many of his students grapple with the same dilemma.   grapple is

3   a.   After the war there was a bulge in the birth rate.   bulge is زيادة (مفاجئة)

    b.   A breakway dunk by Raheed Wallace ended a 120 run by the Bullets that   bulge is
          gave them their 5-point bulge.

4   a.   Just skimming the newspaper, I saw a headline about the Pope's visit.   skim is يتصفح

    b.   A moment more and the helicopter rises along the slope of a mountain, up   skim is
          and over the tree line until we are skimming over peaks of rock that are
          jagged as flints.

321

5   a.   The taut and provocative film ``Rough Treatment'' was made in 1988 by   taut is حاد ، مسبب للتوتر
         the Polish director Krzysztof Kieslowski.

    b.   Eyes blinking, showing no signs of being emotionally taut, her husband   taut is
         looked like an ordinary man defending the ordinary lies he had made up to
         hide an ordinary affair.

6   a.   In New York you can hear a concert of ambulances, fire engines and   shatter is يحطم جدار الصمت
         police cars that shatter the air with their noise.

    b.   I hope to make you laugh, but I also hope to shatter the ideas you have, he   shatter is
         said. You're forced to question your own prejudices.

7   a.   The speedometer moved up to nudge sixty.   nudge is يلامس

    b.   In that race, there were other people who wanted to run, but the party   nudge is
         leaders nudged them not to enroll, because ``they didn't have a chance.''
         ``It wasn't as formal a process,'' Mr. Byrne said.

8   a.   We have been raking through all her papers.   rake is يفتش/يبحث بدقة

    b.   Relief foundations raked in $13 million last year.   rake is

9   a.   The Truth Squad is just one cog in the Democratic machine created to   cog is جزء صغير
         mock and contradict the Republicans during their convention.

    b.   Clark has been a major cog in the St. Louis attack since being acquired in a   cog is
         July 31 trade with the Baltimore Orioles.

322

10	a.	They came up with the nugget that he had been involved in dubious business speculations.	nugget is معلومة ثمينة/قيمة
	b.	The new LSS does that with a choice of V6 engines and with a body, interior and suspension that make the car a true nugget in today's rushing stream of fancy cars.	nugget is
11	a.	``Decani is awful,'' Mr. Holbrooke said, standing out- side homes that had been gutted by grenades. ``It looks like western Bosnia when I went there in 1992.''	gut is يفرغ شيء من محتواه
	b.	But while President Clinton fought against the gutting of environmental laws, he offered little of no resistance on civil liberties.	gut is
12	a.	The fans hooted at Mike Richter, who wasn't at his best during the competition.	hoot is يصيح باستهجان
	b.	We drove smiling, hooting, fists thrust through the open windows in our decorated cars through the streets.	hoot is
13	a.	I learnt how to forge someone else's signature.	forge is يزوّر
	b.	Mr. Murstein said he was surprised that Disney, which has forged a close relationship with the Giuliani administration as it has invested in Times Square, had been turned down by the city.	forge is
14	a.	I pegged him as a big spender.	peg is يصف، يسم
	b.	But he added that he was still considering pegging his country's currency to the dollar, a move the I.M.F. strongly opposes.	peg is

323

15 a.   Ms. Thompson smothered the boy with kisses.   يغطي بوابل من (القبلات. إلخ) smother is

    b.   Nearly four months after the agreement, Mayor Giuliani smothered a strike   smother is
          by lawyers for the Legal Aid Society.

16 a.   Daffodils, tulips, and snowdrops are what we call perennials.   معمرة/دائمة الخضرة (نبتة) perennial is

    b.   Losing weight--the perennial New Year's resolution may make you more   perennial is
          attractive and less prone to disorders like diabetes and high blood
          pressure.

17 a.   The sickness will spawn epidemics, and then the epidemics will spread to   spawn is ينتج بكثرة
          the United States. In our own defence, we must do everything in our
          power to help these hurricane victims.

    b.   The Democratic convention is spawning with secret agents who have to   spawn is
          protect the President.

1.3 Worksheet - Core sense condition

1 a. <u>Boost</u> me up this tree and I will get you an apple.

      يرفع is <u>boost</u>

   b. The landlord will <u>boost</u> the rent.

      . يرفع يزيد is <u>boost</u>

2 a. We <u>grappl</u>ed with him and took the gun away from him.

      يتصارع ، يتعارك is <u>grapple</u>

   b. He now watches many of his students <u>grapple</u> with the same dilemma.

      يواجه/يصارع (مشكلة) is <u>grapple</u>

3 a. What's that <u>bulge</u> in your pocket?

      بروز ، انتفاخ is <u>bulge</u>

   b. A breakway dunk by Raheed Wallace ended a 120 run by the Bullets that gave them their 5-point <u>bulge</u>.

      تقدم is <u>bulge</u>

4 a. Using a spoon, <u>skim</u> off any air bubbles from the top of the custard mixture.

      يقشد ، يزيل رغوة is <u>skim</u>

   b. A moment more and the helicopter rises along the slope of a mountain, up and over the tree line until we are <u>skim</u>ming over peaks of rock that are jagged as flints.

      ينزلق فوق is <u>skim</u>

5 a. Wind it round the screws until the wire is <u>taut</u>.

      مشدود is <u>taut</u>

   b. Eyes blinking, showing no signs of being emotionally <u>taut</u>, her husband looked like an ordinary man defending the ordinary lies he had made up to hide an ordinary affair.

      مشدود ، متوتر is <u>taut</u>

6 a. Explosions scorch floors and <u>shatter</u> windows.

      يحطم ، يهشم is <u>shatter</u>

   b. I hope to make you laugh, but I also hope to <u>shatter</u> the ideas you have, he said. You're forced to question your own prejudices.

      يقضي على is <u>shatter</u>

7	a.	Then her mother <u>nudged</u> her and pointed to the seven helicopters flying overhead.	<u>nudge</u> is (يلكز(بالمرفق
	b.	In that race, there were other people who wanted to run, but the party leaders <u>nudged</u> them not to enroll, because ``they didn't have a chance.'' ``It wasn't as formal a process,'' Mr. Byrne said.	<u>nudge</u> is يحث
8	a.	The gardener set fire to the piles of weeds he had <u>raked</u> up.	<u>rake</u> is يدمّ. يجمع أعشاب
	b.	Relief foundations <u>raked</u> in \$13 million last year.	<u>rake</u> is (يجمع (أموال طائلة) . يدر(أرباح طائلة
9	a.	The principle in clocks is that a number of wheels, locked together by <u>cog</u>s, are forced to turn round.	<u>cog</u> is ترس مسنن
	b.	Clark has been a major <u>cog</u> in the St. Louis attack since being acquired in a July 31 trade with the Baltimore Orioles.	<u>cog</u> is عنصر . جزء
10	a.	His father originally sent him solid golden <u>nugget</u>s.	<u>nugget</u> is (كتلة صلبة (من معدن ثمين
	b.	The new LSS does that with a choice of V6 engines and with a body, interior and suspension that make the car a true <u>nugget</u> in today's rushing stream of fancy cars.	<u>nugget</u> is شيء ثمين/قيم
11	a.	My mother hates <u>gut</u>ting fish.	<u>gut</u> is يخرج أحشاء
	b.	But while President Clinton fought against the <u>gut</u>ting of environmental laws, he offered little of no resistance on civil liberties.	<u>gut</u> is (يدمر(الجزء الداخلي من
12	a.	An owl <u>hoot</u>ed among the pines.	<u>hoot</u> is (ينعق (البوم
	b.	We drove smiling, <u>hoot</u>ing, fists thrust through the open windows in our decorated cars through the streets.	<u>hoot</u> is يطلق بوق/زمور السيارة
13	a.	Gaudi liked to work in metal and may have actually <u>forged</u> his own sword.	<u>forge</u> is يصنع بطرق معدن
	b.	Mr. Murstein said he was surprised that Disney, which has <u>forged</u> a close relationship with the Giuliani administration as it has invested in Times Square, had been turned down by the city.	<u>forge</u> is يكون . يشكل

326

14 a.   We pegged a tent to the ground for the kids.   يثبت بوتد . يشد أوتاد peg is

    b.   But he added that he was still considering pegging his country's currency to the dollar, a move the I.M.F. strongly   يثبت (معدل أو سعر) peg is
          opposes.

15 a.   She confessed last year that she had smothered the baby because she could not stand its crying.   يخنق smother is

    b.   Nearly four months after the agreement, Mayor Giuliani smothered a strike by lawyers for the Legal Aid Society.   يخمد (إضراب ، إلخ) . يقمع smother is

16 a.   Glaciers are usually covered with perennial snow.   مدار العام perennial is

    b.   Losing weight--the perennial New Year's resolution may make you more attractive and less prone to disorders like   دائم perennial is
          diabetes and high blood pressure.

17 a.   The salmon fight their way back up the river to spawn.   يضع بيض spawn is

    b.   The Democratic convention is spawning with secret agents who have to protect the President.   يعج ب spawn is

1.4 Worksheet - Extended sense condition

1 a. The tax cut can <u>boost</u> the economy.   يحسن , يدعم , <u>boost</u> is

    b. The landlord will <u>boost</u> the rent.   يرفع يزيد , <u>boost</u> is

2 a. But those explanations have not made it any easier for them to <u>grapple</u> with the stark reality of losing four young men in such a terrible fire.   يواجه <u>grapple</u> is

    b. He now watches many of his students <u>grapple</u> with the same dilemma.   يواجه/يصارع (مشكلة) <u>grapple</u> is

3 a. After the war there was a <u>bulge</u> in the birth rate.   زيادة (مفاجئة) <u>bulge</u> is

    b. A breakway dunk by Raheed Wallace ended a 120 run by the Bullets that gave them their 5-point <u>bulge</u>.   تقدم <u>bulge</u> is

4 a. Just <u>skim</u>ming the newspaper, I saw a headline about the Pope's visit.   يتصفح <u>skim</u> is

    b. A moment more and the helicopter rises along the slope of a mountain, up and over the tree line until we are <u>skim</u>ming over peaks of rock that are jagged as flints.   ينزلق فوق <u>skim</u> is

5 a. The <u>taut</u> and provocative film ``Rough Treatment'' was made in 1988 by the Polish director Krzysztof Kieslowski.   مسبب للتوتر , حاد <u>taut</u> is

    b. Eyes blinking, showing no signs of being emotionally <u>taut</u>, her husband looked like an ordinary man defending the ordinary lies he had made up to hide an ordinary affair.   مشدود , متوتر <u>taut</u> is

6 a. In New York you can hear a concert of ambulances, fire engines and police cars that <u>shatter</u> the air with their noise.   يحطم جدار الصمت <u>shatter</u> is

    b. I hope to make you laugh, but I also hope to <u>shatter</u> the ideas you have, he said. You're forced to question your own prejudices.   يقضي على <u>shatter</u> is

7 a. The speedometer moved up to <u>nudge</u> sixty.   <u>nudge</u> is يلامس

    b. In that race, there were other people who wanted to run, but the party leaders <u>nudge</u>d them not to enroll, because ``they didn't have a chance.'' ``It wasn't as formal a process,'' Mr. Byrne said.   <u>nudge</u> is يحث

8 a. We have been <u>rakin</u>g through all her papers.   <u>rake</u> is يفتش/يبحث بدقة

    b. Relief foundations <u>raked</u> in $13 million last year.   <u>rake</u> is يجمع (أموال طائلة), يدر(أرباح طائلة)

9 a. The Truth Squad is just one <u>cog</u> in the Democratic machine created to mock and contradict the Republicans during their convention.   <u>cog</u> is جزء صغير

    b. Clark has been a major <u>cog</u> in the St. Louis attack since being acquired in a July 31 trade with the Baltimore Orioles.   <u>cog</u> is عنصر ، جزء

10 a. They came up with the <u>nugget</u> that he had been involved in dubious business speculations.   <u>nugget</u> is معلومة ثمينة/قيمة

    b. The new LSS does that with a choice of V6 engines and with a body, interior and suspension that make the car a true <u>nugget</u> in today's rushing stream of fancy cars.   <u>nugget</u> is شيء ثمين/قيم

11 a. But while President Clinton fought against the <u>gut</u>ting of environmental laws, he offered little of no resistance on civil liberties.   <u>gut</u> is يفرغ شيء من محتواه

    b. ``Decani is awful,'' Mr. Holbrooke said, standing out- side homes that had been gutted by grenades. ``It looks like western Bosnia when I went there in 1992.''   <u>gut</u> is يدمر(الجزء الداخلي من)

12 a. The fans <u>hoot</u>ed at Mike Richter, who wasn't at his best during the competition.   <u>hoot</u> is يصيح باستهجان

    b. We drove smiling, <u>hoot</u>ing, fists thrust through the open windows in our decorated cars through the streets.   <u>hoot</u> is يطلق بوق/زمور السيارة

13 a. I learnt how to <u>forge</u> someone else's signature.   <u>forge</u> is يزوّر

    b. Mr. Murstein said he was surprised that Disney, which has <u>forge</u>d a close relationship with the Giuliani   <u>forge</u> is يكون، يشكل
        administration as it has invested in Times Square, had been turned down by the city.

14 a. I <u>pegg</u>ed him as a big spender.   <u>peg</u> is يصف، يسم

    b. But he added that he was still considering <u>pegg</u>ing his country's currency to the dollar, a move the I.M.F. strongly   <u>peg</u> is يثبت (معدل أو سعر)
        opposes.

15 a. Ms. Thompson <u>smother</u>ed the boy with kisses.   <u>smother</u> is يغطي بوابل من (القبلات، إلخ)

    b. Nearly four months after the agreement, Mayor Giuliani <u>smother</u>ed a strike by lawyers for the Legal Aid Society.   <u>smother</u> is يخمد (إضراب، إلخ)، يقمع

16 a. Daffodils, tulips, and snowdrops are what we call <u>perennial</u>s.   <u>perennial</u> is معمرة/دائمة الخضرة (نبتة)

    b. Losing weight--the <u>perennial</u> New Year's resolution may make you more attractive and less prone to disorders like   <u>perennial</u> is دائم
        diabetes and high blood pressure.

17 a. The sickness will <u>spawn</u> epidemics, and then the epidemics will spread to the United States. In our own defence,   <u>spawn</u> is ينتج بكثرة
        we must do everything in our power to help these hurricane victims.

    b. The Democratic convention is <u>spawn</u>ing with secret agents who have to protect the President.   <u>spawn</u> is يعج بـ

330

1.5 Replication Test S2b - Short and long-term post test

Directions:

Write the Arabic translation of what the underlined word means in the sentence.

ex.	That book is very light reading.	light is

In the example, light means "سهل", so you would write "سهل" in the answer space: "light is _____سهل_____"

Questions:

1.   Perhaps advertising might help boost their sales.   boost is

2.   The county is still facing enormous problems as local leaders grapple with huge deficits projected over the   grapple is
      next few years

3.   Washington scored 5 points in a row and, suddenly, it was a 1-point game. Hamilton made two free throws   bulge is
      with 2:16 left for a 7168 bulge, but Femerling scored on a layup cut it to 7170 with 1:59 remaining.

4.   Moments later the airplane skimmed across the landing strip, edging closer and closer to a touchdown- then   skim is
      in a streamer of dust it landed.

5.   Their faces taut and their eyes red, mourners filled the meeting hall at the Springfield Faith Center.   taut is

6.   One hopes that your disturbing news article about poverty will shatter the many illusions and distortions   shatter is
      about the economic boom this country is supposedly enjoying.

7.   The details of the study that nudged the agency to make the change were published in the current issue of   nudge is
      The New England Journal of Medicine.

331

8. But Federal prosecutors here say prison walls have not stopped Mr. Hoover from overseeing an illegal   rake is
   narcotics business that raked in $100 million a year.

9. In the end, that's the cut that really wounds, the terrible knowledge that she was so tiny, so unimportant a   cog is
   cog in the world that she could disappear without comment.

10. Because the agreement gives Tandem a product line that can be sold at lower prices in far higher   nugget is
    volume, ``this deal may end up being the real nugget that turns Tandem back into a competitor in the
    marketplace,'' Mr. Jones said.

11. In 1838, a devastating fire gutted their small shop and soon thereafter David Brown moved west to Illinois,   gut is
    settling on a land grant in his declining years.

12. Readers will take delight in the local descriptions of Beijing, from the ubiquitous street carts selling candied   hoot is
    crab apples to the shining Mercedes-Benzes hooting their way through swarms of Flying Pigeon bicycles.

13. In 1990, Renault and Volvo forged an industrial alliance intended to evolve into a full merger early in 1994.   forge is

14. When a government has declared that it is pegging its currency to another and will defend it by any means   peg is
    necessary but lets it fall anyway, that's devaluation.

15. He's as anxious as you and I to smother things up.   smother is

16. The problems of car parking in eastern midtown is a perennial issue on the agenda of Community Board 6.   perennial is

17. All around them, the area spawns with office workers heading home, parents buying groceries and teenagers   spawn is
    cruising. It is evidence that Downtown Brooklyn is outgrowing itself.

332

2.1 Items used in the pilot study

LINE

High frequency	He raced towards the finishing LINE.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	We were both thinking along the same LINES.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	She comes from a long LINE of actors.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	She ate healthy food because it was good for her LINES.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	He turned on the LINE to listen to music.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

ORDER

High frequency	Samuel trained every day in ORDER to make his presentation better.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	Then they call out our names in ORDER and we answer yes or no.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	The people of South Africa wanted a new social ORDER.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	A new car is in the price ORDER between 30 000 and 150 000 riyals.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	Write your answer down on a piece of ORDER.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

333

BOARD

High frequency	I've put a list of names up on the BOARD.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	The BOARD of Directors met yesterday.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	In the old-age home, she will have to pay for room and BOARD.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	She planted tomatoes in the vegetable BOARD.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	I put the books into my BOARD and went to school.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

COVER

High frequency	There is a plastic COVER over the meal.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	The cloud COVER in the morning should clear later.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	The COVERS had slipped off the bed in the night.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	The house was old and its COVER needed to be repaired.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	The people have a sense of COVER in their neighborhood.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

334

BODY

High frequency	For their BODY size, these birds lay very small eggs.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	The student BODY numbers 5000.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	There is now a large BODY of knowledge about childhood.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	Butter is a fatty BODY that is very good for cooking.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	He looked at the BODY to check the time.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

COURSE

High frequency	Andy is doing a one-year business COURSE.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	The ship was blown off COURSE.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	The doctor directed her to take a new COURSE of medicine.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	The exchange COURSE is 1 US Dollar for 3.63 Qatari Riyals.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	He stepped on the COURSE to check his weight.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

FACE

High frequency	She had a beautiful FACE.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	Gordon is a familiar FACE at the local flower show.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	He liked the FACE of the watch.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	Small children can get into trouble, so you must keep a FACE on them.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	I'll tell you when to get off the FACE.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

HAND

High frequency	She waved her HAND to the crowd.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	Can you give me a HAND to lift this?	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	The farm HANDS wake up at 5:00 in the morning.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	Our staff combine efficient service with a personal HAND.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	The HAND started to grow in the sunshine.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

336

AIR

High frequency	Let's go outside and get some fresh AIR.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	AIR travel was growing rapidly.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	Trudy is always putting on AIRS.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	My suitcase was so full I didn't have AIR for anything else	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	Blackbirds lay their AIR in March.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

ARM

High frequency	Tim's mother put her ARMS around him.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	The growing ARMS trade is a problem for the country.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	Toyota-America is the American marketing ARM of a Japanese company.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	The horse was black, with white marks on its ARMS.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	I arrived here two ARMS ago.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

337

CUT

High frequency	Teachers are expecting further CUTS next year.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	Make a small CUT in the paper.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	She was determined to claim her CUT of the prize money.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	Air bags are a standard CUT in most new cars.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	The morning CUT came streaming in through the windows.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

HEART

High frequency	Regular exercise is good for the HEART.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	Edith loved her boy with all her HEART and soul.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	We've got to take HEART from the fact that we played well.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	The centre part of an apple is called its HEART.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	HEARTS were parked on both sides of the road.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

POSITION

High frequency	Next week we will be in a much better POSITION to talk about it.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	Our hotel was in a central POSITION near St Mark's Square.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	Bill took up his new POSITION as Works Director in October.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	I can't buy a new car because my bank POSITION is too low.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	Turn down the POSITION, it's too loud.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

CHARACTER

High frequency	He has a happy but quiet CHARACTER.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	It takes strength of CHARACTER to admit you are wrong.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	I can't read the CHARACTERS on that sign.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	His red hair and short CHARACTER made him easy to recognize.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	The doctor gave me a CHARACTER for my cold.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

CLASS

High frequency	She is a member of the working CLASS.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	English CLASSES start at 5:15.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	These flowers will give your garden a touch of CLASS.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	The phone CLASS to telephone Europe is 2 Riyals per minute.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	A photo was stuck to the wall with a CLASS.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

BRANCH

High frequency	The top BRANCHES were full of birds.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	All BRANCHes of government are having to reduce spending.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	This BRANCH of the river eventually empties into the Atlantic Ocean.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	Carl began in the fashion BRANCH by running a clothing shop .	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	That car must have cost him a BRANCH.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

340

FORM

High frequency	The bicycle is an environment-friendly FORM of transportation.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	There were six houses arranged in the FORM of a square.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	He's been in good FORM all this season.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	I am on a diet to lose my FORM.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	The cook boiled three FORMS for dinner.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

POINT

High frequency	That's a very interesting POINT.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Middle frequency	No cars are allowed beyond this POINT.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Low frequency	Be careful with that needle - it has a very sharp POINT.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Logical distractor	The cook put a POINT of salt into the soup.	CORRECT	*	INCORRECT	*	DON'T KNOW	*
Illogical distractor	Your hair needs a good POINT.	CORRECT	*	INCORRECT	*	DON'T KNOW	*

341

CASE

High frequency	In this CASE, several solutions could be tried.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Middle frequency	He told the lawyer that he didn't want a court CASE.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Low frequency	Polly carried her CASES upstairs to the bedroom.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Logical distractor	The sales woman put the money into the cash CASE.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Illogical distractor	I showered and put on a clean CASE.	CORRECT	*	INCORRECT	*	DON'T KNOW *

HEAD

High frequency	He turned his HEAD and looked at me.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Middle frequency	I went to sleep early to have a clear HEAD for the exam.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Low frequency	The president sat at the HEAD of the table.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Logical distractor	I thought she was upset because she had a sad HEAD.	CORRECT	*	INCORRECT	*	DON'T KNOW *
Illogical distractor	Come through into the dining HEAD.	CORRECT	*	INCORRECT	*	DON'T KNOW *

3.1 Grouped presentation

In the following sentences your will find one WORD in capital letters.　　ID Number:_____

Do you think the meaning of that WORD is CORRECT or INCORRECT?

You can make a guess. If you don't know, you can choose DON'T KNOW.

Only think about the meaning of the word, not the grammar.

For example:

BREAK

The waves BROKE on the beach.	CORRECT	☑	INCORRECT	☐	DON'T KNOW	☐
He BROKE the juice for lunch.	CORRECT	☐	INCORRECT	☑	DON'T KNOW	☐
He BROKE the window.	CORRECT	☑	INCORRECT	☐	DON'T KNOW	☐

COVER

The house was old and its COVER needed to be repaired.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The cloud COVER in the morning should clear later.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
There is a plastic COVER over the meal.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The COVERS had slipped off the bed in the night.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The people have a sense of COVER in their neighbourhood.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

HEAD

He turned his HEAD and looked at me.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The president sat at the HEAD of the table.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I went to sleep early to have a clear HEAD for the exam.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Come through into the dining HEAD.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I thought she was upset because she had a sad HEAD.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

CLASS

She is a member of the working CLASS.	CORRECT	□	INCORRECT	□	DON'T KNOW □
English CLASSES start at 5:15.	CORRECT	□	INCORRECT	□	DON'T KNOW □
A photo was stuck to the wall with a CLASS.	CORRECT	□	INCORRECT	□	DON'T KNOW □
The phone CLASS to telephone Europe is 2 Riyals per minute.	CORRECT	□	INCORRECT	□	DON'T KNOW □
These flowers will give your garden a touch of CLASS.	CORRECT	□	INCORRECT	□	DON'T KNOW □

CASE

In this CASE, several solutions could be tried.	CORRECT	□	INCORRECT	□	DON'T KNOW □
He told the lawyer that he didn't want a court CASE.	CORRECT	□	INCORRECT	□	DON'T KNOW □
He had a bad CASE of the flu.	CORRECT	□	INCORRECT	□	DON'T KNOW □
I met her by CASE in Oxford Street.	CORRECT	□	INCORRECT	□	DON'T KNOW □
I showered and put on a clean CASE.	CORRECT	□	INCORRECT	□	DON'T KNOW □

FORM

I am on a diet to lose my FORM.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
He's been in good FORM all this season.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The bicycle is an environment-friendly FORM of transportation.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The medicine comes in a liquid FORM.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The cook boiled three FORMS for dinner.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

ORDER

We prefer to travel in first ORDER.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Write your answer down on a piece of ORDER.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Samuel trained every day in ORDER to make his presentation better.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Then they call out our names in ORDER and we answer yes or no.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The people of South Africa wanted a new social ORDER.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

LINE

She comes from a long LINE of actors.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
She ate healthy food because it was good for her LINES.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
We were both thinking along the same LINES.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
He raced towards the finishing LINE.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
He turned on the LINE to listen to music.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

HEART

Edith loved her boy with all her HEART and soul.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Regular exercise is good for the HEART.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I have a house in the HEART of the city.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The centre part of an apple is called its HEART.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
HEARTS were parked on both sides of the road.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

CUT

Air bags are a standard CUT in most new cars.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Make a small CUT in the paper.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
The morning CUT came streaming in through the windows.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
She was determined to claim her CUT of the prize money.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Teachers are expecting further CUTS next year.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐

BODY

Butter is a fatty BODY that is very good for cooking.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
The student BODY numbers 5000.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
He looked at the BODY to check the time.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
For their BODY size, these birds lay very small eggs.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
There is now a large BODY of knowledge about childhood.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐

BOARD

She planted tomatoes in the vegetable BOARD.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The BOARD of Directors met yesterday.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I've put a list of names up on the BOARD.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
In the old-age home, she will have to pay for room and BOARD.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I put the books into my BOARD and went to school.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

POSITION

Bill took up his new POSITION as Works Director in October.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
It was the best hotel in the POSITION.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Turn down the POSITION, it's too loud.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Our hotel was in a central POSITION near St Mark's Square.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Next week we will be in a much better POSITION to talk about it.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

POINT

No cars are allowed beyond this POINT.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Be careful with that needle - it has a very sharp POINT.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
That's a very interesting POINT.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Your hair needs a good POINT.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The cook put a POINT of salt into the soup.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

CHARACTER

It takes strength of CHARACTER to admit you are wrong.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I can't read the CHARACTERS on that sign.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
His red hair and short CHARACTER made him easy to recognize.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The doctor gave me a CHARACTER for my cold.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
It's not in her CHARACTER to be jealous.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

COURSE

He stepped on the COURSE to check his weight.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The doctor directed her to take a new COURSE of medicine.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The plane changed COURSE to avoid the storm.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Andy is doing a one-year business COURSE.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The exchange COURSE is 1 US Dollar for 3.63 Qatari Riyals.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

ARMS

The horse was black, with white marks on its ARMS.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I arrived here two ARMS ago.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Tim's mother put her ARMS around him.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The growing ARMS trade is a problem for the country.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Please don't sit on the ARM of the chair.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

BRANCHES

All BRANCHES of government are having to reduce spending.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
That car must have cost him a BRANCH.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I broke the BRANCH of my glasses!	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The top BRANCHES were full of birds.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
This BRANCH of the river eventually empties into the Atlantic Ocean.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

HAND

The HAND started to grow in the sunshine.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The farm HANDS wake up at 5:00 in the morning.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Can you give me a HAND to lift this?	CORRECT	□	INCORRECT	□	DON'T KNOW	□
She waved her HAND to the crowd.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Our staff combine efficient service with a personal HAND.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

AIR

AIR travel was growing rapidly.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
My suitcase was so full I didn't have AIR for anything else	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Blackbirds lay their AIR in March.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Trudy is always putting on AIRS.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Let's go outside and get some fresh AIR.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

FACE

He liked the FACE of the watch.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Small children can get into trouble, so you must keep a FACE on them.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
She had a beautiful FACE.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Gordon is a familiar FACE at the local flower show.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
You must keep a FACE of $100 in your bank account.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

3.2 Mixed presentation

In the following sentences your will find one WORD in capital letters.   ID Number:_____

Do you think the meaning of that WORD is CORRECT or INCORRECT?

You can make a guess. If you don't know, you can choose DON'T KNOW.

Only think about the meaning of the word, not the grammar.

For example:

BREAK

The waves BROKE on the beach.	CORRECT	☑	INCORRECT	☐	DON'T KNOW ☐
He BROKE the juice for lunch.	CORRECT	☐	INCORRECT	☑	DON'T KNOW ☐
He BROKE the window.	CORRECT	☑	INCORRECT	☐	DON'T KNOW ☐

I put the books into my BOARD and went to school.	CORRECT □	INCORRECT □	DON'T KNOW □

I put the books into my BOARD and went to school. CORRECT □ INCORRECT □ DON'T KNOW □

He had a bad CASE of the flu. CORRECT □ INCORRECT □ DON'T KNOW □

The growing ARMS trade is a problem for the country. CORRECT □ INCORRECT □ DON'T KNOW □

The bicycle is an environment-friendly FORM of transportation. CORRECT □ INCORRECT □ DON'T KNOW □

These flowers will give your garden a touch of CLASS. CORRECT □ INCORRECT □ DON'T KNOW □

It takes strength of CHARACTER to admit you are wrong. CORRECT □ INCORRECT □ DON'T KNOW □

The plane changed COURSE to avoid the storm. CORRECT □ INCORRECT □ DON'T KNOW □

I thought she was upset because she had a sad HEAD. CORRECT □ INCORRECT □ DON'T KNOW □

He raced towards the finishing LINE. CORRECT □ INCORRECT □ DON'T KNOW □

This BRANCH of the river eventually empties into the Atlantic Ocean. CORRECT □ INCORRECT □ DON'T KNOW □

355

Then they call out our names in ORDER and we answer yes or no.	CORRECT □	INCORRECT □	DON'T KNOW □
He looked at the BODY to check the time.	CORRECT □	INCORRECT □	DON'T KNOW □
Be careful with that needle - it has a very sharp POINT.	CORRECT □	INCORRECT □	DON'T KNOW □
Our staff combine efficient service with a personal HAND.	CORRECT □	INCORRECT □	DON'T KNOW □
Let's go outside and get some fresh AIR.	CORRECT □	INCORRECT □	DON'T KNOW □
The centre part of an apple is called its HEART.	CORRECT □	INCORRECT □	DON'T KNOW □
It's not in her CHARACTER to be jealous.	CORRECT □	INCORRECT □	DON'T KNOW □
The BOARD of Directors met yesterday.	CORRECT □	INCORRECT □	DON'T KNOW □
It was the best hotel in the POSITION.	CORRECT □	INCORRECT □	DON'T KNOW □
The horse was black, with white marks on its ARMS.	CORRECT □	INCORRECT □	DON'T KNOW □

Small children can get into trouble, so you must keep a FACE on them.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
He's been in good FORM all this season.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
For their BODY size, these birds lay very small eggs.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
The farm HANDS wake up at 5:00 in the morning.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
The people of South Africa wanted a new social ORDER.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
Regular exercise is good for the HEART.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
His red hair and short CHARACTER made him easy to recognize.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
I broke the BRANCH of my glasses!	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
She was determined to claim her CUT of the prize money.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐
Trudy is always putting on AIRS.	CORRECT ☐	INCORRECT ☐	DON'T KNOW ☐

We were both thinking along the same LINES.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Please don't sit on the ARM of the chair.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The president sat at the HEAD of the table.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The exchange COURSE is 1 US Dollar for 3.63 Qatari Riyals.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The house was old and its COVER needed to be repaired.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The cook boiled three FORMS for dinner.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
Turn down the POSITION, it's too loud.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The HAND started to grow in the sunshine.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
The doctor gave me a CHARACTER for my cold.	CORRECT	□	INCORRECT	□	DON'T KNOW	□
I met her by CASE in Oxford Street.	CORRECT	□	INCORRECT	□	DON'T KNOW	□

You must keep a FACE of $100 in your bank account.	CORRECT □	INCORRECT □	DON'T KNOW □
Edith loved her boy with all her HEART and soul.	CORRECT □	INCORRECT □	DON'T KNOW □
The student BODY numbers 5000.	CORRECT □	INCORRECT □	DON'T KNOW □
Come through into the dining HEAD.	CORRECT □	INCORRECT □	DON'T KNOW □
The cook put a POINT of salt into the soup.	CORRECT □	INCORRECT □	DON'T KNOW □
He stepped on the COURSE to check his weight.	CORRECT □	INCORRECT □	DON'T KNOW □
English CLASSES start at 5:15.	CORRECT □	INCORRECT □	DON'T KNOW □
Tim's mother put her ARMS around him.	CORRECT □	INCORRECT □	DON'T KNOW □
The people have a sense of COVER in their neighborhood.	CORRECT □	INCORRECT □	DON'T KNOW □
Can you give me a HAND to lift this?	CORRECT □	INCORRECT □	DON'T KNOW □

Teachers are expecting further CUTS next year.	CORRECT □	INCORRECT □	DON'T KNOW □
That car must have cost him a BRANCH.	CORRECT □	INCORRECT □	DON'T KNOW □
I have a house in the HEART of the city.	CORRECT □	INCORRECT □	DON'T KNOW □
He told the lawyer that he didn't want a court CASE.	CORRECT □	INCORRECT □	DON'T KNOW □
The medicine comes in a liquid FORM.	CORRECT □	INCORRECT □	DON'T KNOW □
I went to sleep early to have a clear HEAD for the exam.	CORRECT □	INCORRECT □	DON'T KNOW □
My suitcase was so full I didn't have AIR for anything else	CORRECT □	INCORRECT □	DON'T KNOW □
That's a very interesting POINT.	CORRECT □	INCORRECT □	DON'T KNOW □
Next week we will be in a much better POSITION to talk about it.	CORRECT □	INCORRECT □	DON'T KNOW □
She waved her HAND to the crowd.	CORRECT □	INCORRECT □	DON'T KNOW □

She comes from a long LINE of actors.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

She is a member of the working CLASS.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

She had a beautiful FACE.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

In the old-age home, she will have to pay for room and BOARD.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

HEARTS were parked on both sides of the road.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

Write your answer down on a piece of ORDER.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

The morning CUT came streaming in through the windows.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

In this CASE, several solutions could be tried.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

Butter is a fatty BODY that is very good for cooking.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

The COVERS had slipped off the bed in the night.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

The doctor directed her to take a new COURSE of medicine.	CORRECT □	INCORRECT □	DON'T KNOW □
I can't read the CHARACTERS on that sign.	CORRECT □	INCORRECT □	DON'T KNOW □
He turned on the LINE to listen to music.	CORRECT □	INCORRECT □	DON'T KNOW □
He turned his HEAD and looked at me.	CORRECT □	INCORRECT □	DON'T KNOW □
AIR travel was growing rapidly.	CORRECT □	INCORRECT □	DON'T KNOW □
All BRANCHES of government are having to reduce spending.	CORRECT □	INCORRECT □	DON'T KNOW □
I am on a diet to lose my FORM.	CORRECT □	INCORRECT □	DON'T KNOW □
She planted tomatoes in the vegetable BOARD.	CORRECT □	INCORRECT □	DON'T KNOW □
I arrived here two ARMS ago.	CORRECT □	INCORRECT □	DON'T KNOW □
There is now a large BODY of knowledge about childhood.	CORRECT □	INCORRECT □	DON'T KNOW □

He liked the FACE of the watch.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

I showered and put on a clean CASE.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

Bill took up his new POSITION as Works Director in October.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

She ate healthy food because it was good for her LINES.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

The phone CLASS to telephone Europe is 2 Riyals per minute.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

Your hair needs a good POINT.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

The cloud COVER in the morning should clear later.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

Make a small CUT in the paper.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

We prefer to travel in first ORDER.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

The top BRANCHES were full of birds.   CORRECT   □   INCORRECT   □   DON'T KNOW   □

Gordon is a familiar FACE at the local flower show.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Air bags are a standard CUT in most new cars.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Our hotel was in a central POSITION near St Mark's Square.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
A photo was stuck to the wall with a CLASS.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
No cars are allowed beyond this POINT.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
There is a plastic COVER over the meal.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Blackbirds lay their AIR in March.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Andy is doing a one-year business COURSE.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
I've put a list of names up on the BOARD.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐
Samuel trained every day in ORDER to make his presentation better.	CORRECT	☐	INCORRECT	☐	DON'T KNOW ☐

Appendix 4 Vocabulary Levels Test

This is a vocabulary test.   You must choose the right word to go with each meaning.   Write the number of that word next to its meaning.   Here is an example.

l   business
2   clock   _____ part of a house
3   horse   _____ animal with four legs
4   pencil   _____ something used for writing
5   shoe
6   wall

You answer it in the following way.

l   business
2   clock   ___6__ part of a house
3   horse   ___3__ animal with four legs
4   pencil   ___4__ something used for writing
5   shoe
6   wall

Some words are in the test to make it more difficult.   You do not have to find a meaning for these words.   In the example above, these words are business, clock, and shoe.

If you have no idea about the meaning of a word, do not guess. But if you think you might know the meaning, then you should try to find the answer.

Version 1   The 2,000 word level

1 birth
2 dust   _____ game
3 operation   _____ winning
4 row   _____ being born
5 sport
6 victory

1 choice
2 crop   _____ heat
3 flesh   _____ meat
4 salary   _____ money paid regularly
5 secret   for doing a job
6 temperature

1 cap
2 education   _____ teaching and learning
3 journey   _____ numbers to measure
4 parent   with
5 scale   _____ going to a far place
6 trick

1 attack
2 charm   _____ gold and silver
3 lack   _____ pleasing quality
4 pen   _____ not having something
5 shadow
6 treasure

1 cream
2 factory   _____ part of milk
3 nail   _____ a lot of money
4 pupil   _____ person who is studying
5 sacrifice
6 wealth

1 adopt
2 climb   _____ go up
3 examine   _____ look at closely
4 pour   _____ be on every side
5 satisfy
6 surround

1 bake
2 connect   _____ join together
3 inquire   _____ walk without purpose
4 limit   _____ keep within a certain
5 recognize   size
6 wander

1 burst
2 concern   _____ break open
3 deliver   _____ make better
4 fold   _____ take something to
5 improve   someone
6 urge

1 original
2 private   _____ first
3 royal   _____ not public
4 slow   _____ all added together
5 sorry
6 total

1 brave
2 electric   _____ commonly done
3 firm   _____ wanting food
4 hungry   _____ having no fear
5 local
6 usual

366