INCIDENTAL ACQUISITION OF FOREIGN LANGUAGE VOCABULARY THROUGH MULTI-MODAL SITUATIONS

MARIE-JOSEE BISSON, BSc, MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

DECEMBER 2013

Abstract

There are many advantages to learning a foreign language, such as a better understanding of another culture, a better chance of employment, as well as metalinguistic and cognitive benefits. Unfortunately, language learning can be a long and difficult process, partly because of the amount of words necessary to achieve fluency in a language. However, informal exposure to languages can help language learning. The aim of this thesis was to explore the impact of informal exposure on the incidental acquisition of foreign language vocabulary with complete beginners in a foreign language. This was investigated in the context of multi-modal situations including verbal (auditory and written) and pictorial information.

The first study investigated the acquisition of foreign language vocabulary through watching a film with subtitles, as well as the processing of subtitles using eye-tracking. In the second study, a simpler multi-modal situation (one spoken and written foreign language word form along with a line drawing depicting the meaning of the word) was used to investigate incidental vocabulary acquisition. In addition, this study investigated the use of a potentially more sensitive measure of early vocabulary learning based on the savings paradigm. The third study reported in this thesis explored the number of exposures to multi-modal stimuli necessary for incidental learning to occur. The fourth study, explored the type of connections resulting from the multi-modal incidental learning situation in the context of the current models of FL word learning. Finally, in the fifth study, the allocation of attention to the verbal and pictorial elements of the multi-modal situation were investigated using eye-tracking, as well as the impact of attention on vocabulary acquisition.

The results of the studies presented in this thesis showed that incidental acquisition of foreign language vocabulary happens rapidly during a simple multi-modal situation, and that the pictorial information plays an important role. Furthermore, this type of situation was shown to be effective even with complete beginners in a foreign language and therefore could be exploited to promote informal language learning.

Declaration

I declare that this is my own work conducted during my time as PhD student at the University of Nottingham. Although some of the data presented in Chapter 2 of this thesis was collected during my Master's year and presented in my Master's thesis, more data was collected during my PhD and it was significantly re-analyzed for the PhD thesis. More specifically, for the PhD thesis, data was collected for 18 more participants, and the analyses included seven new dependent variables.

Two of the chapters in this thesis have been published in peer-reviewed journals, and one chapter is under review. The data presented in Chapter 2 was published in *Applied Psycholinguistics* (Bisson, van Heuven, Conklin & Tunney, 2012), and the data presented in Chapter 3 was published in *Plos One* (Bisson, van Heuven, Conklin & Tunney, 2013). The data presented in Chapter 4 is currently under review in *Language Learning* (Bisson, van Heuven, Conklin & Tunney, under review).

Published and submitted manuscripts

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2012). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 1–20.

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLoS ONE, 8*(4), e60912.

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (under review). The role of frequency of exposure in incidental multi-modal acquisition of foreign language vocabulary.

Acknowledgement

I need to thank so many people for their help, advice and moral support during the whole process of completing my PhD. Firstly, thank you to my supervisors, Walter, Kathy and Richard for believing in me and giving me the opportunity to do a PhD with them. An especially big thank you goes to Walter for all his in depth help and advice, and for helping me become a better researcher and a better writer.

I also thank my officemates Cat, Paula and Andrey for listening when I needed to talk, and for offering their advice on so many little things. Thank you also to all my postgrad friends for moral support and many fun climbing nights! I am also really grateful to Cat, Mia and Lauren for letting me stay over in Beeston so many times!

Thank you to Hilde, Frea, Charlotte and Berelim for their help with data collection, and to the members of the BRG for their advice and helpful comments on posters and presentations.

I am also grateful to my family for their support, each in their own way, and finally, the biggest thanks of all goes to Rob for his support, absolutely every step of the way.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Chapter 6

Chapter 1: Introduction

1.1. General introduction

There are many benefits to learning other languages such as an awareness and appreciation of other cultures, as well as metalinguistic and cognitive benefits (e.g. Bialystok, 2008; Sanz, 2013; Yelland, Pollard, & Mercuri, 1993). In addition, in a global economy, being able to speak another language increases one's chances of finding employment (see Graddol, 2006). In view of these many benefits, an increasing number of people speak more than one language to some degree of fluency. Pure monolinguals are therefore becoming difficult to find, and many believe that monolingualism is no longer the rule, but is rather the exception (see e.g., Bhatia & Ritchie, 2007). Many people learn English as a second language, however, other languages such as Mandarin and Spanish, have gained in popularity in recent years (Graddol, 2006).

Although language learning is important and desirable, many find it a difficult and frustrating experience. This became even more apparent to me during my years as a secondary school teacher of French as a foreign language (FL). Furthermore, teaching policies and practices are changing all the time, and teachers must quickly adapt their activities and teaching methods to reflect this (e.g. Lightbown & Spada, 1999). Time in the language learning classroom is normally divided between the four skills of listening, speaking, reading and writing, with pupils engaged mostly in formal learning activities requiring effort, motivation and concentration. In contrast, informal exposure to languages in the language classroom is limited, and activities such as games and FL films are often relegated to extra-curricular activities. However, informal exposure to a FL can benefit FL learners. For example, such exposure in childhood has been shown to help FL learners acquire a more native-like accent later on as adults (Au, Knightly, Jun, & Oh, 2002). Informal exposure to a FL film with FL subtitles has helped advanced learners improve their FL speech perception (Mitterer & McQueen, 2009). Even complete beginners can benefit from informal exposure as shown in a study where exposure to a short FL weather report resulted in an increased sensitivity to the words heard in the weather report compared to other FL words (Veroude, Norris, Shumskaya, Gullberg, & Indefrey, 2010). Thus, informal exposure to spoken FL can give rise to speech perception and production benefits. Furthermore, anecdotal evidence from people exposed to FLs informally on a regular basis, for example through films and TV programs also suggests that it is beneficial (see e.g., Koolstra, Peeters & Spinhof, 2002).

Overall, informal exposure to FLs leads to speech perception and production benefits, requires little effort, and is not restricted to the classroom. Informal exposure can therefore potentially play an important role in language learning, and as such, the overall aim of this thesis was to investigate the impact of informal exposure on FL learning. Although learning a language involves the acquisition of vocabulary, syntax and grammar, because of its crucial role in meaningful communication, this thesis focused on vocabulary acquisition. As Wilkens (1972) said: "Without grammar, very little can be conveyed, without vocabulary nothing can be conveyed" (in Schmitt, 2010, p. 5). Furthermore, researchers agree on the importance of acquiring a large vocabulary for fluent communication (e.g., de Groot & van Hell, 2005; Nation, 2001; Schmitt 2008, 2010). In fact, it is estimated that in order to have enough knowledge to understand authentic written texts, learners need to know approximately 8000 word families (Nation, 2001; Schmitt 2008, 2010). As each word family includes word forms that are related semantically (e.g., teach, teacher, teachable, etc.) this represents approximately 35 000 words (Schmitt, 2008; see Schmitt 2010 for a discussion on word families). Whilst this number is estimated to be lower for the understanding of spoken discourse, it is still more words than can be realistically taught in an instructional setting (de Groot & van Hell, 2005; Horst, 2005), hence the importance of informal exposure for vocabulary learning. Much research has been conducted on reading as a potential source of vocabulary acquisition (see Table 1.2). This is often referred to as learning from context, as the meaning of new words must be derived using the context of the sentence (Nation 2001). However, this situation is only suitable for advanced language learners, as beginners do not have enough vocabulary knowledge to be able to derive the meaning of new words from the context of a sentence. It is estimated that learners must know at least 95 % of the words in a text in order to do this (see Nation, 2001; Schmitt, 2008, 2010 for a discussion on lexical coverage). Therefore, most research on word learning through reading has been conducted with advanced FL learners, whereas very little is known about the potential benefits of informal exposure for beginner learners. To fill this gap in literature, this present thesis focused on the early stages of vocabulary learning with complete novices in a FL.

Before moving onto theoretical issues, it is necessary to clarify the distinction between FL and second language learning. FL learning normally refers to the learning of a language that is not widely spoken in one's country. In contrast, second language learning is often used to describe the learning of a language other than the native language (NL) in multilingual countries. For example, in a bilingual country like Canada, where French and English are both widely spoken, native speakers of French learning English are often referred to as second language learners (see e.g., de Groot and Van Hell, 2005 for an explanation of the distinction). In this thesis, studies conducted in both FL and second language settings will be discussed. However, as the distinction between the two is not crucial for the aims of thesis, and for the sake of clarity, the learning of a language other than the NL will be referred to as FL learning throughout the thesis.

1.2. Theoretical Issues

1.2.1. Levels of knowledge

When children acquire words in their NL, they are also learning about what these words represent. Thus, they are developing both their linguistic and conceptual competences (see Bialystok, 2001). NL learning continues throughout adulthood however, for example when learning technical terms or coming across novel words in a text. Learning words in a FL is similar to learning NL words as adults, since once meaning representations are established, word learning involves acquiring new word forms, and linking these to existing meaning representations.[1] Nation (2001) described the different types of knowledge involved in knowing a word, as well as the distinction between receptive and productive knowledge (see Table 1.1). For example, knowledge of the spoken and written form of a word involves knowing what the word sounds like and looks like at the receptive level, whilst it involves knowing how to pronounce and spell a word at the productive level. Understanding the meaning attached to a specific word form, and being able to select the appropriate word form to express a specific meaning represents the knowledge of form-meaning links at the receptive and productive level respectively (Nation, 2001). One last level of knowledge described by Nation (2001), is the knowledge of word usage, which includes the grammatical functions of a word, knowing which words it co-occurs with

[1] Meaning representations across languages do not always overlap completely, (see de Groot and van Hell, 2005) thus there may still be some meaning acquisition required when learning FL words.

(collocations) and knowing where, when and how frequently to expect to come across and use a word. So far, all these levels of knowledge are relevant to both NL and FL learning, as in both cases one must learn new word forms, and link these to known meanings. However, one of the differences between vocabulary learning in the NL and the FL is the amount of effort required, or "the learning burden" (Nation, 2001, p. 23-24). The effort required to learn FL words varies according to the familiarity of the learner with the phonology and orthography of the language, the similarity with their NL, and the regularity of the sound and spelling patterns of the words (Nation, 2001). For advanced FL learners, the learning burden is lessened, because they are already familiar with the phonology and orthography of the language. However, beginner FL learners are faced with a more difficult task as they lack experience with the FL. Even though the learning burden is not exactly the same, in view of the similarities between learning words in the NL (at least later on in life) and FL word learning, both NL and FL learning studies will be discussed throughout this thesis.

Table 1.1: What is involved in knowing a word

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be use to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency…)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note. R = receptive knowledge; P = productive knowledge. Adapted from "*Learning Vocabulary in Another Language",* by I. S. P. Nation, 2001, p. 27. Copyright 2001 by Cambridge University Press.

1.2.2. Types of learning

In the previous Section, I discussed the different levels of knowledge involved in 'knowing a word', as well as the similarities between NL and FL learning. However, how is word knowledge acquired? Firstly, an important distinction must be made between the terms 'acquisition' and 'learning' as put forward by Krashen (1982). In his view, 'acquiring' a FL is seen as a process similar to acquiring a NL, i.e., through mere exposure to the language in an informal manner. In contrast, 'learning' a FL normally refers to the explicit teaching and learning of rules or lists of words. In Krashen's view, in order to increase the likelihood of becoming a fluent speaker of a language, the emphasis should be put on acquisition rather than learning. Krashen's ideas were very influential in the field of language learning, although many do not adhere to his terminology (see e.g. Gass, 1999; Hulstijn, 2001; Williams, 2005). In this thesis, I use the terms 'acquisition' and 'learning' interchangeably, and I will refer to the specific type of learning involved.

Researchers have long established a distinction between implicit and explicit learning. Implicit learning can be described as learning without intention, and without a clear awareness of what has been acquired (Perruchet & Pacton, 2006; Reber, 1989; Rebuschat & Williams, 2011; Williams, 2005, 2009). In contrast, in explicit learning paradigms, learners are willingly trying to commit some information to memory and therefore the learning is intentional and the learners are aware. In explicit language learning instruction, for example, learners are often provided with a FL word along with its NL translation equivalent and they have to learn to associate them (de Groot & van Hell, 2005). There is some debate as to what can be learnt implicitly, and studies of implicit language learning have so far shown implicit learning of syntax in natural language (e.g. Rebuschat & Williams, 2011) and artificial grammar learning (e.g. Reber 1967). In terms of vocabulary learning, it is thought that the learning of word forms can happen implicitly, for example through reading, but that the learning of form-meaning links happens explicitly (e.g. Ellis, 1994). However, Leung and Williams (2011) showed that the learning of association between grammatical forms and meaning can occur implicitly. Researchers also use the term 'incidental learning' to describe cases where the learning was not required, i.e. participants were not explicitly asked to learn (e.g. Pellicer-Sánchez & Schmitt, 2010; Williams, 2010). Rather, learning occurred through mere exposure to FL material during a variety of tasks, following which, a surprise word recognition or recall test is completed (Horst, 2005; Hulstijn, 2001; Laufer, 2001). Furthermore, in contrast to research on implicit learning, researchers using incidental learning paradigms are more interested in the amount of learning that occurred, rather than the levels of awareness of the participants. Although incidental learning is sometimes described as learning without intention (e.g., Schmitt, 2010; Williams, 2009), researchers have often used this type of paradigm in language classroom settings where presumably the overarching goal is to learn the language and therefore, in those cases, it is unclear whether the learning was intentional or not (see Gass, 1999 for a discussion). However, the focus of the learners is often elsewhere, for example on comprehension, and therefore learning occurs as a by-product of another task (Hulstijn, 2001; Schmitt, 2010). Overall, even though there is a lot of variability in the use of the term, in the studies I will review in the next section, 'incidental learning' is used to describe learning that was not explicitly required and that happened through exposure to the language. That is all that can be assumed, as the intentions and awareness of the learners are not reported.

Despite the difficulty in defining incidental learning, its importance is well established since the high number of words necessary to achieve fluency in a language (see Section 1.1) would be difficult to learn solely explicitly through, for example, classroom instruction (Horst, 2005). Furthermore, incidental learning provides a learning context more similar to what we experience as children learning a NL, where knowledge about vocabulary is just picked up through repeated exposure. In addition, as the requirement for incidental FL learning is mainly FL input, this type of learning can happen anywhere. In other words, it is not necessarily linked to a formal learning situation. Furthermore, incidental learning does not require a conscious effort to learn and therefore might be more attractive to people for whom the prospect of explicitly learning lists of words is unappealing. Many studies have investigated the incidental acquisition of vocabulary, and a selection of these will be reviewed in more detail in the next section. I will start by reviewing incidental learning studies from exposure to either written or spoken input. Following this, studies investigating the impact of bimodal written and spoken input will be reviewed, before discussing studies using exposure to multiple inputs including both verbal (spoken and /or written) and pictorial information.

1.3. Incidental vocabulary learning studies

1.3.1. Unimodal input: written

In the field of FL vocabulary acquisition, many studies have investigated the incidental acquisition of vocabulary through reading (e.g., Brown, Waring, & Donkaewbua, 2008; Horst, 2005; Hulstijn, Hollander, & Greidanus, 1996; Kweon & Kim, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Rott, 1999; Vidal, 2011; Waring, & Takaki, 2003). This is often referred to as learning from context, as the meaning of the new words must be derived from the meaning of adjacent words and from the context of the sentence. Table 1.2 presents an overview of incidental learning studies from reading. Overall, the amount of exposures to the new words in these studies varied from one exposure (Hulstijn et al., 1996; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Vidal, 2011) to over twenty exposures (Brown et al., 2008; Kweon & Kim, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006), with each study using a variety of frequencies of exposures and number of target words, as well as different measures of word learning. For example, Pellicer-Sánchez and Schmitt (2010) asked participants to read a 150-page novel containing 34 unknown target words split into 5 frequency bands, with the exposure to the words in the story varying from 1 to 28 occurrences. The vocabulary gains were measured following the reading of the novel using a spelling recognition, word class knowledge, meaning recognition and meaning recall tests. Across all tests of word knowledge and frequency of occurrences, the authors concluded that knowledge gains were found on about 9 of the new words. Furthermore, an overall frequency of exposure effect was found, such that learners had significantly more knowledge across all the tests on the words that occurred 10 or more times in the novel compared to words that occurred 8 times or less. Unfortunately, only 5 words occurred 10 times or more, whilst 29 words occurred 8 times or less, because the study involved authentic text and therefore the number of words in each frequency band could not be controlled. In addition, a control group who only completed the vocabulary tests would have been useful to confirm that the vocabulary gains were due to the reading of the novel.

Rather than using a novel, Vidal (2011) asked native Spanish speakers to read a series of university lectures in English. Subsequently a vocabulary knowledge scale (including form recognition, meaning recall, NL translation and using the word in a sentence) was used to assess vocabulary gains on 36 target words (another group of participants listened to the lectures rather than read them; the results of this group will be discussed in Section 1.3.2). The authors compared the scores on their vocabulary knowledge scale prior to and following the reading of the lectures. Overall, knowledge gains were found on approximately 19 words,[2] and the frequency of exposure (from 1 to 6 repetitions) was found to be a strong predictor of vocabulary gains with 47% of the variance explained by this predictor. Similarly to the Pellicer-Sánchez and Schmitt study, a limitation of this study is that some frequency bands included more words than other (between 3 and 10 words per frequency band). However, because both these studies used authentic material as FL input it was not possible to control for this aspect of their material.

In order to have more control over the FL input, some reading studies wrote material specifically to investigate the incidental acquisition of vocabulary through reading (e.g. Rott, 1999), or altered existing texts slightly (e.g. Hulstijn et al., 1996). For example, Rott (1999) wrote a set of short paragraphs each including novel words that participants read once per week. Using a between-subject design, participants were exposed to 6 target words either 2, 4 or 6 times before completing a productive vocabulary task ("supply a definition") and a receptive vocabulary task ("select a definition") to assess their acquisition of word knowledge. Results showed that scores on items from 2 exposures were significantly higher than scores on a control set of items that participants had not been exposed to. Furthermore, authors found a frequency of exposure effect overall, with participants exposed to novel words 6 times consistently outperforming participants who had been exposed to the words 2 or 4 times. Interestingly, in contrast to the participants in Pellicer-Sánchez and Schmitt (2010) and Vidal (2011), who were advanced FL learners,

[2] Number estimated using Figure 2 of Vidal (2011).

participants in the Rott study were intermediate FL learners. In addition, the Rott study found incidental learning following only a few exposures to novel words in a meaningful context. This is much faster vocabulary acquisition from reading than the 8-10 exposures required, as estimated by Schmitt (2011). However, it is worth noting that Rott wrote the material specifically for the study and therefore each encounter with the target words would presumably have been in a very clear context, and this may have facilitated learning.

Table 1.2: Unimodal input: reading studies

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
Brown et al., 2008	Pseudowords in FL (intermediate FL learners)	**Reading**, listening and reading-while-listening (within subject design) to 3 stories	Meaning reco. MC tests with 5 options and FL to L1 translation imm. after each story, 1 week later, and 3 months later.	2-3, 7-9, 10-13, 15-20, with seven words in each freq. band for each story.	84 (28 per story so results presented as average out of 28)	Imm. post-test about 13 words recognized, and 4 translated correctly; sig. differences between each freq. bands, but not Bonferonni corrected, unclear if sig. higher than chance; no control group or control items.
Horst, 2005	FL words (elementary to high intermediate learners)	Extensive reading of between 3-33 graded readers over 6 weeks period *M* = 10.5 books per participant	Self-report of word meaning knowledge (YES, NO, Not sure) pre-post treatment on a selection of words from the books; (different words pre-post tests)	unknown	Tested on 100 words (50 'infrequent' words, and 50 'more frequent' words)	Participants reported knowing 17 more words on average post-treatment
Hulstijn et al., 1996	FL words (advanced FL learners)	Reading a slightly adapted short story (1306 words long)	Form reco., meaning recall in L1 or FL; meaning recall with context cues	8 words appeared once, 8 appeared 3 times	16 target and 16 control words	Sig. effect of freq. on reco., very low scores on recall measures, with no freq. effect: 1.4/16 points for recall, and 4.2/16 points recall with context
Kweon & Kim, 2008	FL words (intermediate learners)	Reading of 3 authentic novels (100 000 words) over 5 weeks	Self-report of word meaning knowledge (YES, NO, Not sure) pre and post treatment	20 or more occurrences (22 words), 7-19 (79 words) or 1-6 (266 words)	367 words (same pre and post test)	Participants reported a 21.4 % increase in word knowledge; freq. effect not tested with stats.
Pellicer-Sánchez & Schmitt, 2010	African words amongst FL (English) text	Reading a novel (67 000 words) over a one month period	Spelling and meaning reco. using MC tests with 5 options; meaning recall; word class knowledge test	28 or more (3 words), 10-17 (2 words), 5-8 (9words), 2-4 (10 words) 1 (10 words)	34 target words	Learning on 9 words (28%) across all tests. Meaning reco. was higher with 43% (see main text for freq. effect results and limitations)

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
Pigada & Schmitt, 2006	FL words with 1 learner only	Reading 4 simplified books (30 000 words total) over one month	Grammatical function (gender for nouns and prepositions for verbs), spelling, and meaning during interview pre-post treatment	1 (17 words), 2-3 (17 words), 4-5 (15 words), 6-10 (13 words), 10+ (5 words), 20+ (3 words)	70 nouns and 63 verbs (133 words total)	Sig. improvements on all tests for nouns: +19%, +13% and +30% on spelling, meaning and gender. For verbs, +27%, +18%, +17% on spelling, meaning and prepositions.
Rott, 1999	FL words with intermediate FL learners	Reading of short stories (4-6 sentences) including one target word each.	Pre-test (FL to NL translation) on 12 target words plus 38 distracters, post-test meaning reco. MC tests with 5 options and meaning recall (supply a definition)	2, 4 or 6 exposures between-subject design (one per treatment day)	6 target and 6 control words per participants (2 lists of words used overall)	Sig. learning for all groups on both tests tests, at each test time. 6-exposure group > 2 and 4-exposure groups on both tests at each test time. No difference between 2 and 4 exposure groups.
Vidal, 2011	FL words with learners ranging from score 328 to 662 in the PBT TOEFL test	**Reading** or listening to 3 University lectures	Vocabulary knowledge scale (form reco., meaning recall, NL translation and using the word in a sentence) pre-post treatment.	1 (6 words), 2 (6 words), 3 (8 words), 4 (3 words), 5 (10 words), 6 (3 words).	36 target words	Gains in knowledge overall on about 19 words on the imm. post-test; sig. learning effects on both imm. and delayed post tests.
Waring & Takaki, 2003	FL pseudowords with intermediate learners	Reading a 5872 word graded reader	Word-form reco., MC meaning reco. tests with 5 options and translation tests imm. after reading; one week and three months later	5 words per frequency bands: 1 exposure, 4-5, 8-10, 13-14, 15-20.	25 target words	Form reco., meaning reco. and recall on 15, 11, and 5 words at imm. test, decreased over time. Limitation: no stats provided, no control group or items.

Note. imm. = immediately; reco. = recognition; freq = frequency; sig. = significant; MC = multiple-choice.

In this section I have discussed so far studies of incidental vocabulary learning with FL learners. However, research on the acquisition of vocabulary through reading has also been conducted with native speakers acquiring novel pseudowords (see Table 1.3). These studies, investigating early learning effects, used electroencephalography (EEG) to record event-related potential (ERPs) components. The main ERP component that these studies focused on was the N400, which is a negative fluctuation in the electrical brain activity happening around 400 milliseconds post stimulus presentation which is thought to index semantic processing (Kutas & Federmeier, 2000). A modulation in the N400 component has been used to demonstrate meaning integration in adult word learning studies, indicating a learning effect (e.g., Batterink & Neville, 2011; Borovsky, Kutas, & Elman, 2010; Dobel et al., 2009; McLaughlin, Osterhout, & Kim, 2004). For example, Batterink and Neville (2011) used the N400 to explore meaning integration across 10 exposures to words embedded in short stories. Overall, they found a bigger reduction in the N400 for novel words with a consistent meaning compared to novel words for which no consistent meaning could be derived (hence controlling for an effect due to the repetition of the word form only). Interestingly, the difference between the two types of novel words emerged as early as the second encounter with the words. In another study, participants read triplets of sentences where the meaning of the new words could be derived from the sentence context (Mestres-Missé, Rodriguez-Fornells, & Münte, 2007). Remarkably, the results showed that following a third exposure, the N400 component for the new words was undistinguishable from that of known words. Further evidence of early learning effects has been found following a single encounter with a new word (Borovsky et al., 2010). This study showed that the N400 component was reduced following a plausible usage of the novel word, compared to an implausible usage, in a test sentence presented immediately after the exposure to the novel word in a meaningful context. Overall, early learning effects were found either during the exposure to the new words, or immediately after the presentation of each word, and therefore it is unclear whether these effects are long lasting. It is important to note that in Borovsky et al. (2010) participants responded to a test sentence after each context sentence. Therefore, they may have become aware of the purpose of the study. Similarly, Batterink and Neville (2011) tested participants on their vocabulary acquisition after each story, and therefore the learning was not entirely incidental. In addition, in Mestres-Missé et al. (2007), participants were specifically told to try and derive the meaning of the words. Nevertheless, what emerges from these ERP studies is that form-meaning links can occur from just a few exposures to a new orthographic word form in a meaningful context.

Table 1.3: Unimodal input: EEG studies

Study	**Material**	**Incidental learning task**	**Measure**	**Freq. of exposure**	**Number of words**	**Results**
Borovsky et al., 2010	Pseudowords NL	Reading sentences	EEG and plausibility ratings of sentences imm. after each initial sentence	1	40 targets 40 control words	After exposure, N400s significantly smaller if novel word as an object was plausible
Batterink & Neville, 2011	Pseudowords NL	Reading four short stories (4000 to 5000 words each)	EEG during reading; lexical decision task, reco. task (semantic relatedness) recall of meaning	10 exposures	26-27 target and and control items	N400 reduction from 2nd exposure, reco. accuracy = 72.4 % recall = 63.8% Did not measure reco. and recall for control words.
Mestre-Misse & Rodriguez-Fornell 2007 EXP 1	Pseudowords NL	Reading sentences, but explicit learning	ERP to novel words across 3 sentences	3 (one per sentence)	65 target 65 control	By the 3rd sentence, the N400 was undistinguishable from known words
EXP 2	Pseudowords NL	Reading sentences, but explicit learning	EEG and accuracy judgment of relatedness to real word	3	65	ERP amplitudes to novel word-real word pair similar to real word pairs

Note. imm. = immediately; reco. = recognition

1.3.2. Unimodal input: spoken

The incidental acquisition of vocabulary has also been investigated in the context of exposure to spoken input (see Table 1.4 for an overview). A landmark study by Saffran, Newport, Aslin, Tunick, and Barrueco (1997), which has become known as the Saffran task, exposed participants to a continuous recording containing 6 pseudowords made from syllables from an artificial language. In this study, participants completed an unrelated computer task (creating computer illustrations), while listening to the unsegmented speech stream containing 300 repetitions of the pseudowords presented without word boundaries. Following the exposure phase, participants had to complete a surprise 2AFC word form recognition test using the pseudowords from the recording, as well as new pseudowords made up of the same syllables as foils. The results showed that both adults and children participants correctly chose words from the tape with 59% accuracy. A further group of participants repeated the task twice on two consecutive days, and achieved 73% and 68% accuracy for adults and children respectively. This study illustrated that both adults and children are able to use statistical information to extract words from a continuous speech stream, and that this process can happen completely incidentally while attention is focused elsewhere. In this study, however, participants could only extract word form information as there was no meaning attached to these words. Further work on the incidental acquisition of phonological word forms was conducted by Davis, Di Betta, MacDonald and Gaskell, (2008), Dumay and Gaskell (2007), and Gaskell and Dumay (2003) and these studies will be discussed in more detail in Section 1.4.1.

In contrast to the above studies that focused on the incidental learning of phonological word (pseudoword) forms, Vidal (2011) studied the incidental acquisition of vocabulary while students listened to university lectures (this study was also discussed in the section on incidental learning from reading (Section 1.3.1), as there were two groups of participants involved: one group who read the lectures, and one group who listened to them). In this study, as in the reading studies reviewed previously (see Section 1.3.1), it was possible to derive the meaning of novel words from the context of the sentence. Using the same vocabulary knowledge scale and the pre-test / post-test design as with the group of students who read the lectures, it was found that knowledge was gained on about 12 words [2]. The frequency of exposure to the words was not as strong a predictor of word knowledge for the group of students who listened to the lectures, as this predictor explained only 24% of the variance. It is worth noting however, that all the vocabulary tests in this study where conducted in the *written* modality, whilst participants in the listening condition had only been exposed to the words through the *listening* modality. This could have had an impact on their word form recognition scores especially, and therefore the results might have underestimated the amount of knowledge actually gained from listening. In contrast, Brown et al. (2008) assessed vocabulary gains through a listening test following an exposure phase where groups of participants listened to a FL story including 28 pseudo FL target words. In this study participants were also exposed to FL material through reading and reading-while-listening in a within-subject design, the results of which will be discussed in the section on reading-while-listening (Section 1.3.3). The frequency of occurrence of the target words in the story varied between 2 to 20 exposures, split into four frequency bands. Although it was found overall that participants recognized on average 8 new words, they found that frequency of exposure did not have an effect on recognition or recall test scores. Furthermore, this study did not include a control group or a set of control items, and therefore it is difficult to conclude that the results of the 4AFC (plus the option of responding I don't know) meaning recognition test were due to a learning effect, or that they were higher than chance. Finally, the design of this study was such that participants would have become aware of the purpose of the study following the first story and test session. Nevertheless, the results of both these studies suggest that incidental learning of form-meaning links does occur from listening input, however, it seems to be more difficult than with written input, and may require more exposures.

Table 1.4: Unimodal input from listening

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
Brown et al., 2008	Pseudowords in FL	Reading, **listening** and reading-while-listening (within subject design) to 3 stories	Meaning reco. MC tests with 5 options and FL to L1 translation imm. after each story, 1 week later, and 3 months later	2-3, 7-9, 10-13, 15-20, with seven words in each freq. band for each story	84 (28 per story, results presented as average out of 28)	Imm. post-test about 8 words recognized and 1 translated correctly. Limitation: no control group or items
Saffran et al., 1997 EXP.1	Non-words	Listening to 21 min. tape of continuous speech stream while performing unrelated task	2 AFC recognition test (which word have you heard before) with new nonsense words as foils made from same syllables	300 times each word	6 nonsense words	Participants correctly chose words from tape about 59% time (for both adults and children)
Saffran et al., 1997 EXP.2	Non-words	Idem, but heard tape twice over two days	idem	idem	idem	73 and 68% accuracy for adults and children
Vidal, 2011	FL words with learners ranging from score 328 to 662 in the PBT TOEFL test	**Listening** or reading to 3 University lectures	Vocabulary knowledge scale (form reco., meaning recall, NL translation and using the word in a sentence) pre-post treatment	1 (6 words), 2 (6 words), 3 (8 words), 4 (3 words), 5 (10 words), 6 (3 words)	36 target words	Gains in knowledge overall on about 12 words on the imm. post-test; sig. learning effects on both imm. and delayed post tests

Note. imm. = immediate; reco. = recognition; sig. = significant; AFC = alternative forced choice; MC = multiple choice

1.3.3. Bimodal input: auditory and written information

One way to improve phonological word acquisition is to provide an orthographic word form during encoding, as suggested by studies using explicit learning tasks (see Hu, 2008; Ricketts, Bishop, & Nation, 2009; Rosenthal & Ehri, 2008). Furthermore, in an implicit learning paradigm, Bird and Williams (2002) found that the bimodal presentation of novel words during a learning phase resulted in increased recognition memory for the auditory word forms during a test phase. In addition, the same study found that with single word presentation, having access to written word forms facilitated the processing of auditory word forms (see Table 1.5 for more details). Bimodal input could also potentially facilitate the segmentation of a continuous speech stream into more manageable chunks. Reading researchers have been using bimodal input to investigate the incidental acquisition of vocabulary, a situation they call 'reading-while-listening' (see Table 1.5 for an overview of these studies). Although some used this situation to ensure that participants in a study read the story in its entirety (e.g. Horst, Cobb, & Meara, 1998; Webb, Newton, & Chang, 2013), others believed that this situation could lead to more incidental acquisition of vocabulary than reading or listening on their own (e.g. Brown et al., 2008). Brown et al. (2008) found no significant differences in recognition or translation scores of participants in a reading-while-listening condition compared to participants in a reading only condition. However, both conditions involving written input outperformed the listening only condition (discussed in Section 1.3.2).

Nevertheless, participants in the reading-while-listening condition correctly recognized 13 words (out of 28) and correctly translated 4 words into NL. Furthermore, Brown et al. (2008) found overall that words with a higher frequency of exposures in the stories were more likely to be remembered in both the reading and reading-while-listening condition. There were significant differences in the number of words correctly recalled and recognized between 2-3 exposures, 7-9 exposures and 10-13 exposures, but encountering the words 15-20 times did not improve the scores further in the reading-while-listening condition. In contrast, a significant increase was found in the reading condition for each frequency band. In addition to the limitations of this study already discussed in Section 1.3.2, it is important to mention that although this study used multiple comparisons (*t*-tests) in the frequency effect analyses, it is not reported whether their alpha values were corrected accordingly. In spite of that, the results of the Brown et al. study suggest that both reading and reading-while-listening lead to vocabulary gains and frequency of exposure effects. Interestingly, participants reported having a preference for the reading-while-listening condition. Further research on reading-while-listening was conducted by Webb et al. (2013), who investigated the incidental acquisition of collocations rather than single word units. Webb et al., (2013) defined collocations as multiple word units that co-occur more often than chance (for example *short distance*, *play it safe*, and *shoot the breeze*). The frequency of exposure to the collocations included in the study varied in four frequency bands (1, 5, 10 and 15 exposures) and the text of the stories was altered such that each participant was exposed to the 18 collocations according to one of the frequency bands (between-subject design). Compared to a control group that was not exposed to the stories, the results showed incidental learning of the form of the collocations from 5 exposures at the receptive level and from 10 exposures at the productive level, whilst incidental learning of the meaning of the collocations occurred from 10 exposures at the productive level and from 15 exposures at the receptive level. Across all tests of collocation knowledge, significantly higher knowledge was consistently found in the group exposed to each collocation 15 times compared to all other groups. Another study used a pre-test/post-test design to investigate the incidental acquisition of vocabulary, with participants reading a story while listening to the teacher reading aloud to them (Horst et al., 1998). Following this, participants completed a 4AFC meaning recognition test and a word association test, each on 45 words from the story. In the word association test, participants had to select which amongst three words from the story did not belong with the others (an odd one out task). Results showed that on average participants selected the correct meaning of 4.6 more words following the reading of the story, and that their scores increased by 1.3 words on the word association measure. Furthermore, a correlation of .49 was reported between the frequency of exposure to the words in the story (from 2 to 17 occurrences) and the learning gains, although it is not reported whether this correlation was significant. Nevertheless, the results of this study showed modest, but significant gains in vocabulary knowledge. However, it is possible that the pre-test increased participants' attention to some of the words, and as with other incidental learning studies, there were no control group or control items. Overall, the bimodal input studies discussed in this section showed vocabulary gains similar to the reading studies. Importantly, the additional spoken input does not seem to be detrimental compared to written input only, and as the explicit learning studies mentioned earlier suggest, the written input seems to facilitate the processing and encoding of the spoken input.

The studies involving reading, listening, and reading-while-listening in a FL reviewed so far in this section highlight the potential of these types of incidental learning situations for promoting incidental FL vocabulary acquisition. Considering only the studies carried out in the context of FL learning, it is puzzling that so much more research has been conducted on the acquisition of FL vocabulary through a reading situation (9 studies) that might favor the learning of orthographic word form and orthographic form-meaning links. In contrast, research on listening (2 studies), which might have had a positive impact on the incidental learning of FL phonological word forms and form-meaning links, or reading-while-listening, which promotes both the learning of phonological and orthographic word forms and form-meaning links (3 studies) is lagging behind. Incidental learning from spoken input appears to be more difficult for learners, however, as explained in Section 1.3.3, providing the orthographic forms of the words seems to facilitate learning.

In the studies of reading, listening, and reading-while-listening reviewed so far, learners could only derive the meaning of the new words from the context of the sentence. Another way of facilitating meaning acquisition for learners is to provide some additional input in their NL. Lambert, Boehler and Sidoti (1981) and Lambert and Holobow (1984) used combinations of languages while students read and/or listened to FL radio programs. Some students had access to both spoken and written input in the FL (which is similar to the reading-while-listening studies reviewed earlier) whilst other students received some of the input in their NL either through the soundtrack or in writing. Although these studies focused on listening comprehension, the first study revealed that participants were better at understanding the meaning of words in context when they had received the FL in writing (Lambert et al., 1981), whilst the second study indicated that having the NL in the spoken input and FL in writing led to better contextual meaning comprehension (Lambert & Holobow, 1984). Taking the results of both studies together, Lambert and Holobow, (1984) concluded that the best input for learners was therefore having the NL spoken dialogue combined with written FL script. Unfortunately, this work with dual language input in the context of radio programs was discontinued. However, as we will see in the next section, with the advent of multimedia, another potential source of incidental learning started to attract researchers' interest.

Table 1.5: Bimodal input from reading and listening

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
Bird & Williams, 2002 EXP 2	Pseudowords in FL with advanced FL learners	Rhyme monitoring task (sound only, text only or sound + text)	Rhyme monitoring task using auditory word forms, and old vs. new auditory word form reco. (d')	3 occurrences	14 targets per word conditions and 14 new word not presented before	RT faster for words originally presented with text only or sound+text, d' scores higher for sound+text, then sound only, then text only
Brown et al., 2008	Pseudowords in FL	Reading, listening and **reading-while-listening** (within subject design) to 3 stories	Meaning reco. MC tests with 5 options and FL to L1 translation imm. after each story, 1 week later, and 3 months later	2-3, 7-9, 10-13, 15-20, with seven words in each freq. band for each story	84 (28 per story, results presented as average out of 28)	Imm. post-test about 13 words recognized, and 4 translated correctly, limitation: no control group or items
Horst et al., 1998	FL words	Reading-while-listening to teacher read an entire book over six 1 hour sessions	Pre-post tests of 4 AFC meaning reco. and word-association (delete the word that doesn't fit out of three)	Between 2 and 17 occurrences	45 words (on average 22 were already known to participants from the pre-test)	Average gain of 4.6 words on the reco. and 1.28 on the association test, .49 correlation between freq. and learning gain. Limitations: see main text
Lambert et al., 1981	FL words (with French immersion students)	Reading and/or listening to radio programs in FL and/or NL (9 conditions)	4 AFC contextual meaning reco., 4 AFC spelling (find the word with a mistake)	unspecified	unspecified but probably same as below.	Best contextual meaning scores for groups with written FL input (NL audio and FL script; FL audio and FL script; and no audio with FL script) with scores ranging from 6 to 8 correct answers, (calculated from their Figure 1). For spelling, FL audio with no script performed worst. Limitation: very different group sizes, different levels of spelling and vocab knowledge prior to experiment

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
Lambert & Holobow, 1984	FL words (French second language learners)	Reading and/or listening to radio programs with FL audio and NL script; FL audio and FL script; FL audio no script; NL audio and FL script; no audio with NL script (5 conditions)	4 AFC contextual meaning reco., 4 AFC spelling (find the word with a mistake)	unspecified	22 questions on contextual meaning reco., 18 questions on spelling.	No differences between groups for spelling; NL audio and FL script performed best for contextual meaning reco. with scores ranging from 4.77 for FL audio and FL script to 7.35 for NL audio and FL script, (i.e. only 2.5 more correct answers)
Webb et al., 2013	FL words with advanced learners	Reading-while-listening to a slightly adapted version of a graded reader	MC pre-post tests of written form reco. with 4 options, and posttests of written form recall (all using initial word as cue), meaning reco. and recall by translation both L2-L1, and L1-L2	0, 1, 5, 10 or 15 exposures (between subject design)	18 target collocations	Sig. learning from 5 exposures for form reco. and from 10 exposures for form recall, and from 10 exposures meaning recall and from 15 exposures for meaning reco. Across all tests sig. higher knowledge in the group exposed to each collocation 15 times compared to all other groups

Note. imm. = immediate; reco. = recognition; sig. = significant; AFC = alternative forced choice; MC = multiple choice; RT = response time

1.3.4. Multi-modal situations (with pictorial information)

The studies reviewed so far all included verbal information that was presented through written and/or auditory modality, and sometimes included a combination of languages in order for learners to extract meaning more easily. Another way of facilitating meaning acquisition of new FL words is to expose learners to FL input in combination with pictorial information. One advantage of using pictorial information is that even complete beginners can access word meaning, as they can derive the meaning of FL words from the pictorial information. Furthermore, in the field of memory research, many studies have demonstrated a *picture superiority effect* whereby information learnt through pictures results in better recall or recognition than learning without pictures (e.g., Paivio & Csapo, 1973; Nelson, Reed & Walling, 1976). Paivio and Csapo (1973) believed that the picture superiority effect occurred because pictures can be encoded both as verbal and non-verbal information (dual-coding theory) and as such either type of information can be used at retrieval. In contrast, Nelson, Reed and McEvoy (1977) assumed that pictures are better remembered because they are more distinctive than words and benefit from a more direct connection to semantic representations (sensory-semantic model). According to either explanation, using pictures during an incidental learning situation may facilitate FL word learning. This section reviews studies of incidental acquisition of FL through exposure to both verbal (written and/or spoken modality), and pictorial information (see Table 1.6 for an overview). These types of situations will be referred to as 'multi-modal' situations throughout the thesis.

Gullberg, Roberts and Dimroth (2012) investigated the incidental acquisition of FL words using a multi-modal situation with FL auditory information and pictorial information. More specifically, they asked participants to watch a short weather report in a FL following which they measured word form recognition and picture to word form matching. The results showed that participants were able to recognize 57.5% of the target words as having occurred in the weather report. In addition, the authors found that participants were more accurate in the picture to word form matching task for the frequently occurring disyllabic words accompanied by a gesture highlighting the meaning of the word (70.5% accuracy). Unfortunately, there were only 8 target words in this task, and as there were 3 factors with 2 levels each (number of syllables: monosyllabic vs. disyllabic; frequency of occurrence: frequent vs. infrequent; gesture highlighting the meaning: yes or no), there was only one item contributing to the mean in each cell thereby reducing the generalizability of the findings. Nevertheless, the result of their word recognition task is interesting as it shows a small but significant early learning effect with complete beginners in FL. Even with the pictorial information, however, it can be quite difficult for learners to extract FL word forms from a continuous speech stream and link them with meaning extracted from the pictorial information. In contrast to the Saffran task where participants were exposed to only 6 target words repeated 300 times (discussed in Section 1.3.2), the frequency of exposure in the Gullberg et al. study ranged from 2 to 16. Therefore, participants in the Gullberg et al. study may have benefitted either from more repetitions of the target words, or from subtitles either in the FL to allow for easier word segmentation, or in the NL to allow for easier mapping of word meaning.

In recent years, FL films with subtitles have become increasingly popular and are easily accessible on the Internet. Furthermore, even when watching a NL film, subtitles in many languages are often available at the click of a button and therefore FL input can easily be added to many NL films. Language researchers quickly became interested in this multi-modal situation as a potential source of incidental vocabulary acquisition, since a combination of FL and NL can be used in conjunction with pictorial information thereby providing an information rich situation. Depending on the type of subtitles, it is possible to have the FL in the soundtrack and NL in the subtitles (standard subtitling), FL in the subtitles and NL in the soundtrack (reversed subtitles) or FL in both (intralingual subtitles). Intralingual subtitling in the NL, sometimes called captioning, is also available during many television programs to make them accessible to the hard of hearing and deaf community.

Early work on the effectiveness of using films with subtitles to promote incidental FL vocabulary acquisition was conducted by d'Ydewalle and Pavakanun (1995). In two experiments, participants (adults in experiment 1 and children in experiment 2) watched a 12-minute cartoon following which they completed a 5 AFC meaning recognition vocabulary test. Both experiments used a between-subject design, and included all possible combinations of FL and/or NL soundtrack and subtitles (9 groups of participants in each study). This study was unique as it included all possible subtitling conditions as well as many possible control conditions (e.g., FL soundtrack with no subtitles, NL soundtrack with no subtitles, no soundtrack with FL subtitles, etc.). The disadvantage of the design of this study was that the number of participants in each condition was small (less than 10 per condition). The authors concluded that the adult data showed evidence of vocabulary acquisition, with the groups of participants with standard and reversed subtitles performing best. However, their statistical analyses did not fully support their conclusion. For example, although they found no interaction between type of soundtrack and type of subtitles in their 3x3 between-subject ANOVA, they still broke down the interaction and performed post-hoc tests on the data. Furthermore, the results of their group with reversed and standard subtitles did not differ significantly from the results of a group who was exposed to intralingual subtitles in the NL, suggesting that the results of the vocabulary tests might not be due to the learning of FL vocabulary. Finally, in their second experiment with children, they found no significant differences between the groups on the vocabulary test scores. In a similar further study, d'Ydewalle and Pavakanun (1997) concluded once more that considerable vocabulary acquisition occurred from watching a short subtitled video, and that reversed subtitles enhanced vocabulary acquisition more than standard subtitles. Unfortunately, no statistical analyses were provided to support their conclusions. In another study, d'Ydewalle and Van de Poel (1999) focused on standard and reversed subtitles only. They used a 10-minute still-motion movie with Danish or French as a FL, following which participants completed a 10-item auditory and a 10-item 3 AFC meaning recognition test. For Danish as FL, a significant increase in written test performance was found in both standard and reversed subtitle groups compared to a control group with NL intralingual subtitles. In the auditory test, only the standard subtitles group performed better than the control group. However, no significant French vocabulary acquisition was found, and authors therefore concluded that incidental vocabulary acquisition might be facilitated when the FL and the NL are similar (participants in this study were native speakers of Dutch, which is more similar to Danish than French). Their results with Danish as a FL however, would have been more convincing had the vocabulary tests included more items. In fact, the significant differences between the groups amount to about one more word correctly recognized. A similar study by Koolstra and Beentjes (1999) involved more items on the vocabulary test (28 items). The results of this study also showed a small increment in vocabulary acquisition with children correctly recognizing the meaning of two more words if they had watched a FL video with NL subtitles (standard subtitles) compared to a control group who watched a different movie, and only one more word when compared to group who watched the FL video with no subtitles. These results need to be interpreted with caution however, as within-subject analyses were conducted despite the design being between-subjects. Taken together, the results of these studies suggest that incidental vocabulary acquisition from watching films with subtitles is possible, however, in view of the limitations of these studies, more research is warranted.

Table 1.6: Multi-modal input (spoken and/or written input with pictorial information)

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
Gullberg et al., 2012 EXP 1	FL soundtrack with complete beginners	Watching a 7 minute weather report (another group watched it twice)	Word form reco. (was this word in the film, YES / NO?)	Either 2 or 8 occurrences in video (so twice this for group 2)	24 target words (and 72 fillers)	Sig. better at responding YES to target words (57.5%) compared to filler items (48.5%). 2-exposure group marginally better than 1 exposure group (60%. vs. 55%). Participants more accurate on frequent words and disyllabic words (bias to reject monosyllabic words). Limitation: unclear if 55% sig. higher than chance.
Gullberg et al., 2012 EXP 2	Idem	Idem	Picture-to-sound matching task (based on weather icons, e.g., sun icon presented with word sun vs. sun icon presented with word rain)	Idem	8 target words (with 3 factors: freq., presented with gesture or not, and number of syllables)	Frequent items accompanied with a gesture led to sig. higher accuracy Limitation: only 8 target words total which means only 1 word per cell in the analysis, and their task is actually based on the icons, not the target words.
Koolstra & Beentjes, 1999	FL soundtrack with NL subs.; FL soundtrack and no subs.; NL soundtrack	Children watched a 15-minute documentary twice	Auditory 4 AFC meaning reco. and word form reco.	Unknown (but at least twice since they viewed video twice)	35 target words (28 included in meaning reco. test, and 13 included in form reco. test)	Meaning of 2 more words correctly recognized in subs. video group compared to control group, and one more word when compared to FL soundtrack group. Limitations: between-subject design but within-subject analysis; unclear if p values Bonferroni corrected
d'Ydewalle & Pavakanun, 1997	FL or NL soundtrack, with or without FL or NL subs.	Watching a 15-minute still motion picture	Form reco., 5AFC meaning reco.	At least 5 occurrences	Unspecified	Authors concluded that considerable language acquisition from watching short subs video, with reversed subs. enhancing vocab. acquisition more than standard subs. Limitation: no statistical analyses reported

Study	Language	Incidental learning task	Measure	Freq. of exposure	Number of words	Results
d'Ydewalle & Pavakanun, 1995 EXP1	FL, NL or no soundtrack, with or without FL or NL subs. (between subject design, 9 conditions)	Watching 12-minute cartoon	5AFC meaning reco.	At least 5 occurrences	47 target words	Authors conclude that there is evidence of FL acquisition with standard and reversed subs. performing best, but their stats. do not completely support their conclusion. Limitation: see main text.
d'Ydewalle & Pavakanun, 1995 EXP2	idem	Grade 9 students watching 12-minute cartoon	5AFC meaning reco.	At least 5 occurrences	47 target words	No significant effect of means of subtitling and /or soundtrack on vocab test results.
d'Ydewalle & Van de Poel, 1999	FL soundtrack and NL subs., NL soundtrack and FL subs., or NL soundtrack and subs.	Children watched a 10 minutes still-motion movie. One group with French one group with Danish as FL	3 AFC meaning reco. half auditory half written	At least 4 occurrences	20 target words (10 auditory form, 10 written form)	For Danish, sig. increase in written test performance for both standard and reversed subs compared to control (about one more word correct). In auditory test, only standard subs group better than control. For French, no sig. differences.

Note. Subs. = subtitles; reco. = recognition; sig. = significant; AFC = alternative forced choice

There is more research showing encouraging learning effects from watching films with subtitles (see Table 1.7 for an overview), although it is either not clear whether participants were aware of an upcoming vocabulary test (Duquette & Painchaud, 1996; Mitterer & McQueen, 2009; Winke, Gass & Sydorenko, 2010), or participants were aware of the purpose of the study from the start of the experiment (Danan, 1992; Boras & Lafayette, 1994; Sydorenko, 2010) and therefore the learning was more intentional than incidental. In other studies, participants would have become aware of its purpose in the course of the study as multiple film excerpts were used with a vocabulary test following each one (Koskinen, Knable, Markham, Jensema, & Kane, 1996; Neuman & Koskinen, 1992; Winke et al., 2010), or participants were alerted to the purpose of the study because of the use of a vocabulary knowledge test at the beginning of the study (Duquette & Painchaud, 1996; Koskinen et al., 1996; Neuman & Koskinen, 1992). Nevertheless, these studies all showed some benefits of watching films with subtitles for FL learners. For instance, one study showed convincingly that it was more helpful for advanced learners to have access to intralingual subtitles compared to standard subtitles in order to retune their understanding of FL accents (Mitterer & McQueen, 2009). In the context of an immigrant population in the USA, intralingual subtitles were found to be beneficial for acquiring science specific vocabulary compared to watching the same science programs with no subtitles (Koskinen et al., 1996; Neuman & Koskinen, 1992). Intralingual subtitles were also found to have a positive impact on oral presentation performances of language students compared to a group of students who did not have access to the subtitles (Boras & Lafayette, 1994). Conversely, in a study comparing the acquisition of FL vocabulary between groups with intralingual subtitles, reversed subtitles or no subtitles, Danan (1992) concluded that reversed subtitles were more beneficial. Overall, the results of these studies favored intralingual subtitling over standard subtitling, intralingual subtitling over FL soundtrack with no subtitles, and reversed subtitling over both intralingual subtitles and no subtitles. Unfortunately, none of the above studies compared all of the subtitling types within the same experiment.

Table 1.7: Other Word Learning Studies using Multi-modal Input with Pictorial Information

Study	Language	Task	Measure	Length of exposure	Number of words	Results
Borras & Lafayette, 1994	FL soundtrack, with or without FL subs.	Watching video and producing an oral presentation describing events from video	Presentations rated on effectiveness (weighted x 3.), accuracy, organization and fluency scored on a 6 point likert scale	Exposure to video twice to start; then students could watch again during oral task preparation	n/a	Subs. group performed sig. better on the oral production task (33.3 points) than no subs. group (27.8 point) (calculated from figures).
Danan, 1992 EXP. 2	FL soundtrack with or without FL subs., or NL soundtrack with FL subs.	Watching 5 minute	Translation (L2-L1) in writing	5 minutes video watched twice, exact frequency unknown	17 target items	Reversed subtitles: 7.91(6.13 prior knowledge); intralingual: 6.33 (5.81 prior knowledge); FL audio only: 3.92 (3.84 prior knowledge). In view of prior knowledge, unclear if they really learnt anything from videos
Danan, 1992 EXP. 3	FL soundtrack with FL subs., or NL soundtrack with FL subs.	idem	idem	idem	idem	Reversed subtitles: 13.85 (9.00 prior knowledge); intralingual: 9.75 (9.85 prior knowledge). Reversed subs. group seems to have learnt about 4 words. But no learning in intralingual group. However, task favoured the reversed group. L2 recall from picture would have been interesting.
Duquette & Painchaud, 1996	FL (French SL learners in Canada)	8 minute video input or listening to soundtrack twice	L2-L1 translation from sentence context (pre-test, imm. post-test and post-test 10 days later	Unspecified	40 items (20 familiar high freq. French words, 20 unfamiliar	No group differences, but both groups showed sig. increase in vocabulary knowledge between pre-post tests (8% increase; about 1.6 words)

Study	Language	Task	Measure	Length of exposure	Number of words	Results
Koskinen et al., 1996	FL with or without FL subs. (science related words in a ESL context, incarcerated population in US)	Exposure to 9 science video programs (3 per unit) over a 9 week period (ranging from 3 to 8.5 minutes)	Pre-test of self-rated knowledge of target words before each unit, post-tests on 4AFC form reco. each week, true/false definition task after each unit, 4AFC meaning reco. at end of study (all tests in FL)	2 viewing of each video, frequency of target words in videos is unknown	10 words per video (90 words total)	No difference between groups on form reco. and sentence task, but subs. group (72.4/90) sig. outperformed no subs group (62.5) on meaning reco. Limitation: Subs. group more proficient, and no control measure
Mitterer & McQueen, 2009	FL soundtrack with NL, FL or no subs. With advanced FL learners	Watching 25 minutes of DVD either in Scottish or Australian accent	Repeat excerpts from material	Unspecified	80 target excerpts (and 80 control excerpts not from DVD)	Participants repeated 5.6 (No subs.), 6.4 (NL subs) and 8.8 (FL subs.) more excerpts accurately than control excerpts
Neuman & Koskinen, 1992	FL (in a ESL context with immigrant and refugee children in the US) with or without FL subs., or R-W-L condition, or reading a textbook.	Watching 9 science video programs (3 per unit) over a 9-week period (ranging from 3 to 8.5 minutes)	Same as Koskinen et al., 1996	Same as Koskinen et al., 1996	Same as Koskinen et al., 1996	Subs. group better than reading group on form reco. by 4 words, and better by about 4 words than no subs group and by 16 words than reading group and textbook group on meaning reco. Limitation: unclear if groups equivalent in proficiency prior to study

Study	Language	Task	Measure	Length of exposure	Number of words	Results
Sydorenko, 2010	FL soundtrack with or without FL subs, and FL subs. only	Watching 3 videos (2-3 minutes long each) twice, tests followed each video	Form reco. (using target words and non-words), L2-L1 translation (half aural, half written), self-rated pre-knowledge of target words test (post study) to exclude known words	Unspecified	28 target words in total, but known words prior to study removed from each participants score	Group with intralingual subs. outperformed group without subs. on translation (35% vs 18% accuracy) Limitations: intralingual group not significantly different from no soundtrack group (24% accuracy). Different numbers of words per participants (authors do not report how many words were excluded based on pre-knowledge)
Winke et al., 2010	FL videos with or without FL subtitles	Spanish 2nd year students watching 3 short videos twice	L2-L1 translation either from written or auditory input (after each video) self-report pre-knowledge test administered post study to eliminate words each participants knew	Unspecified, but at least twice as videos seen twice each	Number of target words unspecified	Intralingual subs. group performed sig. better than no subs group. Limitations: no control group; different numbers of words per participants (authors do not report how many words were excluded based on pre-knowledge)

Note. Subs. = subtitles; reco. = recognition; sig. = significant; AFC = alternative forced choice.

The studies reviewed thus far in this section suggest benefits of watching films with subtitles. However, one concern about films with subtitles is that the reading of the subtitles may detract from watching the film's images (Koolstra et al., 2002), and that watching a film with subtitles might quickly turn into a reading exercise (Jensema, El Sharkawy, Danthuri, Burch, & Hsu, 2000). Using this kind of situation with multiple inputs of information as a potential source of language learning, it becomes difficult to ascertain which of the different inputs people are attending to, and to what extent each type of information is processed. For example, when people are watching a film with subtitles, how much of the subtitles are they actually reading? Furthermore, does the reading of the subtitles leave enough time to process the images in the film? Finally, do people really listen to the FL words in the soundtrack? Multimedia researchers believe that the cognitive load associated with the processing of multiple sources of information may leave little cognitive resources available for learning, especially when some of the information is redundant (Mayer, 2005). Furthermore, it may be that only some of the input is processed, and therefore it is likely that more attentional resources will be allocated to the NL input. The results of the studies on the learning of vocabulary through watching films with subtitles do suggest that the FL soundtrack is processed to some extent if FL vocabulary learning happened in a standard subtitles situation where the only input of FL is through the soundtrack (e.g. Koolstra & Beentjes, 1999; d'Ydewalle & Van de Poel, 1999). It is difficult to quantify the processing of the soundtrack while people watch a film with subtitles. However, using eye-tracking, it is possible to measure how much time is spent processing the subtitles. I will now review studies investigating the processing of subtitles in films using eye-tracking.

Pioneer studies in the 1980's investigated the allocation of attention to the different sources of information in films with standard subtitles. More specifically, they used eye-tracking to measure the amount of time spent looking in the subtitle area as a function of the subtitle's presentation time. In an initial study, d'Ydewalle, Muylle and van Rensbergen (1985) found that participants fixated upon one or two words per subtitle, leading them to conclude that not much reading of the subtitles occurred. However, another study found that participants spent 30% of the subtitle's presentation time looking in the subtitle area (d'Ydewalle, van Rensbergen & Pollet, 1987). Unfortunately, many methodological details are missing from the description of both of these studies (e.g., the number of participants, the language background of the participants and the number of subtitles used in the study). In a study with children from grades 2 to 6, d'Ydewalle and van Rensbergen (1989) found that they too spent time looking in the subtitle area, and interestingly, this varied depending on the type of film: less time was spent in the subtitle area for an action film. In this study, the number of participants is reported (12 in each experiment), however, the other methodological limitations mentioned above are problematic. Despite their limitations, these first few studies on the processing of standard subtitles indicate that some of the presentation time is spent in the subtitle area. This, however, is not surprising as the NL subtitles were necessary to understand the FL film.

The processing of intralingual NL subtitles were investigated while participants watched either a 12-minute English or 20-minute Dutch film (d'Ydewalle, Praet, Verfaillie & van Rensbergen, 1991). In this case, the NL subtitles were not necessary for the comprehension of the film, as a NL soundtrack was available. However, the results showed that participants still read them approximately 20% of the time. The authors argued that participants read the subtitles because it was more efficient than listening to the soundtrack. However, because the same message was available in both soundtrack and subtitles, it is likely that as participants heard the soundtrack, they just followed along with the written text. Although participants did spend some time reading the NL subtitles in a situation where they did not need to, the 20% figure is rather low and it remains unclear how much (or how little) reading of the subtitles would occur if they had been in the FL.

D'Ydewalle and De Bruycker (2007) recently used eye-tracking to investigate the reading of standard and reversed subtitles in children and adults using a 15-minute animation. Eye-movement data was reported while participants read 114 and 138 subtitles for the standard and reversed subtitling conditions respectively. They found overall that reversed subtitles were skipped more often than standard subtitles (21% compared to 4%), were fixated less (0.59 fixation per word compared to 0.91) and that less time was spent in the subtitle area (26% of the subtitle's presentation time compared to 41%). Unfortunately, this study did not include a control group with no subtitles and therefore the above values may reflect some time spent in the subtitle area as a result of other visual aspects of the movie taking place in this region. Nevertheless, what seemed to emerge from this study is that reading behavior does occur in both standard and reversed subtitles and that more time is spent in the subtitle area than previous studies suggested.

To summarize this section, the results of eye-tracking studies investigating the processing of subtitles in films indicate that there seems to be a preference for reading NL subtitles. However, because some reading of FL subtitles also occurred, this bodes well for using this type of situation as a way to promote FL vocabulary acquisition. It is possible that even more reading would occur in an FL intralingual situation as reading along while listening to the soundtrack might help deciphering the FL, however further research is needed in order to ascertain this.

In Section 1.3, studies of incidental FL vocabulary learning through unimodal (reading or listening), bimodal (reading-while-listening) and multi-modal (films with or without subtitles) were reviewed. Overall, what emerges from these studies is that learners of varied levels of proficiencies can learn FL words incidentally. These studies were all interested in *how much* vocabulary can be acquired incidentally, or from *how many* exposures. An important question that was not addressed in these studies however, is *how* are FL words acquired. The next section now reviews theoretical models of word learning.

1.4. Models of word learning

In this section, I review three different models of word learning that have been proposed in the literature starting with the Complementary Learning Systems account of word learning (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010), then the Revised Hierarchical Model (Kroll & Stewart, 1994), and finally the developmental Bilingual Interactive Activation model (Grainger, Midgley & Holcomb, 2010).

1.4.1. Complementary Learning Systems

The Complementary Learning Systems (CLS) account, initially a model of learning and memory, has recently been used to explain word learning (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010). The model includes an initial fast familiarization stage, presumably involving structures in the hippocampus and medial temporal lobes, followed by a slower lexical consolidation process involving the neocortex. In effect, it is thought that new word form representations are initially temporarily stored, and that during a period of sleep, the knowledge is consolidated and integrated with the existing word representation system.

Some of the evidence for this two-stage model of word learning comes from neuroimaging studies, although so far support is only available for the involvement of medial temporal structures during the initial familiarization stage (e.g. Breiteinstein et al., 2005; Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008). In contrast, to my knowledge, there is still no neuroimaging evidence for the changes in cortical involvement associated with a more word-like behavior that is expected following a consolidation period. In a meta-analysis contrasting the levels of activation for brain regions during NL spoken word and NL pseudoword recognition (pseudoword forms that followed the phonotactic rules of the NL), Davis and Gaskell (2009) looked for regions showing an increased activation for pseudowords compared to words, as well as regions showing the opposite pattern, i.e., an increased activation for words compared to pseudowords. For the former, they reported regions in the superior portion of the left temporal lobe, which fell within the dorsal and ventral speech processing pathways, presumably associated with the phonological processing of pseudowords. Together with medial temporal structures, these cortical regions may contribute to the initial familiarization stage. In contrast, they found increased activation for words compared to pseudowords in regions extending both anteriorly and posteriorly from the temporal lobe area, which have been associated with the retrieval of lexical representations. Many neuroimaging studies reported in Davis and Gaskell (2009) revealed decreased activation in brain regions normally associated with the processing of pseudowords following repeated exposures, thereby providing further evidence for the initial familiarization stage (e.g. Breitenstein et al., 2005; Davis, Di Betta, MacDonald & Gaskell, 2008). However, none of the studies reported an increased activation in brain areas normally associated with the processing of words following the repeated exposure to the pseudowords, which would have provided evidence of word-like behavior for the pseudowords. As the authors highlighted however, most studies did not include an offline consolidation period. Furthermore, it is possible that the lack of increase in activation in the brain areas associated with word knowledge following repeated exposure to pseudowords was due to the lack of a semantic component in the tasks. Clearly, more research is needed to provide further neuroimaging evidence for the second stage of the CLS account.

Further support for the CLS account of word learning comes from behavioral studies in which people were repeatedly exposed to pseudoword forms in the NL (Bowers, Davis & Henley, 2005; Davis, et al., 2008; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). In those studies, participants showed good recognition for the novel word forms immediately following initial exposure. However, it was found that pseudoword forms only interfered with the retrieval of known words following a period of sleep, thereby supporting the idea that they had been integrated within the existing word lexicon. In contrast, lexical competition was not found for pseudowords learnt on the same day, providing support for the two-stage account of word learning.

One of the important aspects of the CLS model of word learning is that it can account for the learning of new words in a FL as well as new words in a NL. The studies used to provide evidence for the model have mostly focused on the learning of phonological or orthographic word forms. However, as mentioned in the previous section, there is more to word learning than the initial acquisition of word forms. The two models that I will discuss next were developed specifically to account for FL word learning and include both lexical and semantic components.

1.4.2. Revised Hierarchical Model

One of the most influential models of FL word learning is the Revised Hierarchical Model (RHM) by Kroll and Stewart (1994). This model was developed to explain the links between FL word forms, NL word forms, and meaning representations (see Figure 1.1). Kroll and Stewart used the RHM to make specific predictions regarding the interlanguage lexico-semantic connections of highly fluent late bilinguals. The model predicted that since learners typically explicitly learn to associate new FL word forms with their NL translation equivalents, strong lexical links should exist from FL to NL words. In contrast, weaker lexical links from NL words to FL words were predicted. Furthermore, the model predicted strong conceptual links between

NL words and semantic representation, however, links between FL words and concepts were predicted to be weaker and presumably developed later as fluency in the FL improved. In a crucial experiment, the authors used word translation tasks in both directions (NL-FL and FL-NL) as well as picture naming tasks to test the model's predictions about lexical strength asymmetries. In addition, they included a semantic component to the translation task to investigate the prediction about asymmetries in conceptual links for NL and FL. The results showed a significant interaction between task type (naming vs. translation) and language of the presented word (NL vs. FL), suggesting that participants were faster at translating words from FL to NL compared to translating from NL to FL. Unfortunately, a breakdown of this interaction was not reported in the article. Furthermore, the authors reported that participants were slower at translating words organized in semantic categories compared to words with no semantic relationship (a semantic interference effect) only when participants were translating from the NL to the FL. The authors accounted for this effect by suggesting that translation from the NL to the FL is mediated by semantic representations. However, they reported finding no semantic interference effect when translating from FL to NL, which led them to conclude that this process occurred solely at the lexical level. The results as reported in this article must be interpreted with caution however, as some statistical analyses were not fully reported, some were reported but interpreted mistakenly, and others were reported as significant by item only, and not by participants.

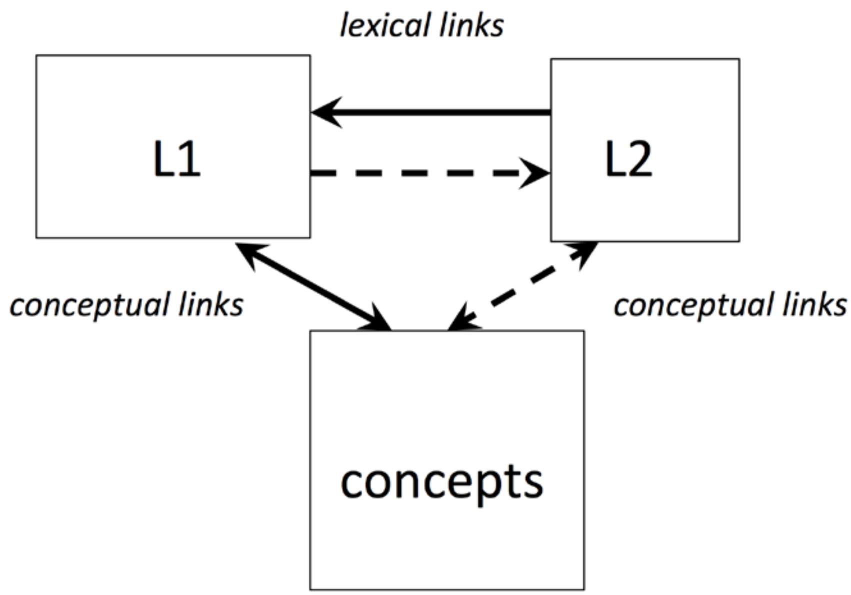

Figure 1.1. The Revised Hierarchical Model of lexical and conceptual representations in bilingual memory. Adapted from "Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations" by J. Kroll and E. Stewart, 1994, *Journal of Memory and Language, 33,* p. 158. Copyright 1994 by Academic Press.

Even though the model was initially presented to explain word production, it was highly influential in inspiring many studies on the bilingual mental lexicon. Many subsequent studies were conducted and provided both evidence in support of and against the model. For example, Brysbaert and Duyck (2010) reviewed studies that provided evidence against the assumption of the model with regards to stronger lexical connections from FL to NL. In addition, there is evidence that even learners with low levels of proficiency can access semantic information during FL to NL translation (see Brysbaert & Duyck, 2010 for a review). In response to this criticism, Kroll, van

Hell, Tokowicz and Green (2010), clarified that the weak link between FL words and semantic representation may be better characterized as asymmetric, depending on the task. They further explained that during FL word comprehension, there is evidence showing that learners with low levels of proficiency can access semantic representations directly, at least for some FL words. In contrast, during FL word production this remains difficult (see Kroll et al., 2010 for a review). Another important point made by Kroll et al., (2010) is that in studies where low proficiency learners are trained intensively on a small number of words, they may in fact start to behave like proficient bilinguals on this small set of words. Notwithstanding the criticisms, against the model, the RHM was very useful in highlighting some of the key issues of the early stages of FL word learning, such as asymmetries in translation directions and changes linked to fluency development.

1.4.3. BIA+d

In the developmental Bilingual Interactive Activation model (BIA-d), Grainger et al. (2010) described in detail the evolution of the connections between languages and semantic representations that is likely to occur as proficiency in a FL develops (see Figure 1.2). This model was designed to account for FL word learning in late FL learners acquiring a FL via classroom instruction, through what the authors call "an initial phase of supervised learning which is progressively replaced by unsupervised learning" (Grainger et al., 2010, p. 277). An important feature of the BIA-d is that both lexicons are integrated, and therefore, as new words are learnt, they are coded according to the language they belong to. This language node allows for inhibitory control during language switching or when operating in single language mode (Grainger et al., 2010). During the supervised phase of language learning, learners are presented with FL word forms and their corresponding NL translations. The authors' assumption is that this leads to connections between FL word forms and NL word forms, and links to the language node associated with the new FL word forms. Much like the RHM, the BIA-d assumes that during this first stage of learning, access to the meaning of the FL word occurs through the NL translation. Importantly, through Hebbian learning processes, connections between the co-activated lexical and semantic representations start to form, such that eventually, the activation of the FL word form co-activates the semantic representation. In other words, a direct route between the FL word and the semantic representation is gradually acquired. Then, during the second phase of language learning (what the authors call the unsupervised stage), inhibitory connections are added to the initial excitatory connections between the translation equivalents. Furthermore, inhibitory connections between the FL node and the NL translation start to develop, as well as inhibitory connections between the FL word forms and similar word forms in the NL.

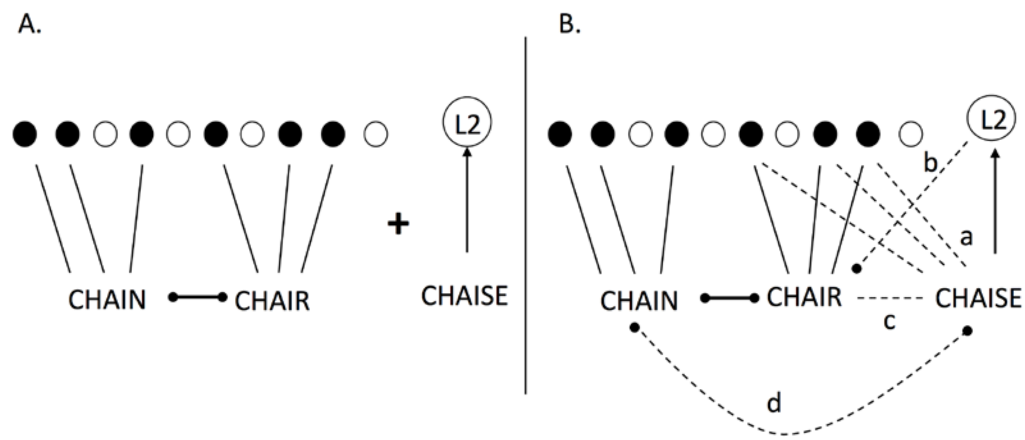

Figure 1.2: The developmental bilingual interactive activation (BIA-d) model of second language vocabulary acquisition. Adapted from "Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d)" by J. Grainger, K. Midgley and P. J. Holcomb, 2010, *Language Acquisition across Linguistic and Cognitive Systems,* p. 277. Copyright 2010 by John Benjamins.

Overall, the BIA-d accommodated the ideas put forward by the RHM in terms of initial reliance on the NL translation and gradual shift toward direct semantic links from FL words (what Grainger et al. call "L2 autonomy"). The BIA-d, however, expanded on these ideas by detailing what happens at the level of individual words. Importantly, Grainger et al. (2010) make it clear that this model would only apply to late language learners acquiring their FL initially in a classroom setting. Furthermore, neither the BIA-d nor the RHM specify how connections between NL and FL lexical representations and semantic representations are formed when a FL is acquired informally. To my knowledge, there are no FL word learning models that were designed to account for incidental FL word learning. Another issue with the current FL word learning models is that both models assume that the FL words are learnt in combination with their NL translations. Hence neither model can account for the connections that may result from FL words learnt in combinations with pictures, as may be the case in the context of a multi-modal situation. This will therefore be one of the aims of this thesis.

1.5. Current thesis

1.5.1. Overall topics of the thesis

As was reviewed in Section 1.3, much prior research on incidental vocabulary acquisition has been conducted in the context of unimodal and bimodal inputs of FL. However, these types of situations are only accessible to more advanced FL learners, as beginners do not have enough vocabulary knowledge to be able to derive the meaning of new words from the context (see Section 1.1). In contrast, multi-modal situations are accessible to all levels of learners, as the meaning of FL words can be derived from the pictures. Research using this type of situation to promote incidental vocabulary learning is still scarce however, and, as discussed in Section 1.3.4, often plagued with methodological issues. This thesis will therefore focus on incidental acquisition of vocabulary through multi-modal situations throughout, and the first question I will address in Chapters 2 to 6 is:

1) Can people learn FL vocabulary incidentally through exposure to multi-modal situations?

An important factor in incidental vocabulary acquisition is the amount of exposure necessary for learning to occur. In the studies reviewed in Section 1.3, researchers have measured learning effects following more than 20 encounters with lexical items (e.g., Pellicer-Sánchez & Schmitt, 2010) whilst others have only exposed participants to a few occurrences (e.g. Rott, 1999; see Tables 1.2 to 1.6). There is still a debate as to the added benefits of increasing the number of encounters with lexical items. Furthermore, it is still largely unknown from how many exposures early vocabulary acquisition can be detected, especially in the context of multi-modal situations. This thesis will therefore also investigate the following two questions:

2.1) Are there frequency of exposure effects in incidental FL acquisition in the context of multi-modal situation?

2.2) How many exposures are needed to detect incidental learning?

Chapters 3 to 6 will provide evidence to answer both questions, whilst the experiment reported in Chapter 4 was designed to investigate these questions specifically.

As discussed in Section 1.3.4, multi-modal situations involve the integration of multiple inputs of information. If these types of situation are to be used to promote FL vocabulary learning it is important to investigate the allocation of attention to the different elements of the multi-modal situation. Some studies reviewed in section 1.3.4 specifically investigated the allocation of attention to the written element of one type of multi-modal situation, namely the reading of subtitles in films. In the thesis I will investigate this more broadly and try to answer the following question:

3) How do people allocate their attention to verbal and visual aspects of multi-modal situations?

Chapter 2 will explore this question in the context of films with subtitles, whilst Chapter 6 will use a simpler multi-modal situation to answer this question.

As discussed in Section 1.3.4, the advantage of using multi-modal inputs of information rather than bimodal input comes from the additional pictorial element. It was mentioned in the same section that previous research in the field of memory suggests that there is a picture superiority effect. However, there are no studies yet exploring the role of pictorial information in incidental learning of FL vocabulary. Therefore Chapter 6 will focus on the question:

4) What is the importance of pictorial information in multi-modal incidental vocabulary acquisition?

An important theoretical issue is the type of connections established during incidental vocabulary learning. As reviewed in Section 1.4, both the RHM and the BIA-d models suggest that new FL word forms are firstly linked to their NL translation equivalents, and that semantic representations are accessed via the NL translation. However, neither model considered what might happen in the case of vocabulary acquisition during an incidental learning situation. Because in this case the focus is not the learning of meaning per se, comparatively to a classroom situation where learners are encouraged to learn translation equivalents, it is possible that learners are less reliant on FL-NL lexical links. In the case of incidental acquisition of vocabulary through multi-modal situations, it is also possible that the presence of the pictorial information encourages the acquisition of direct connections between FL word forms and meaning representations. Therefore an important potential benefit of using multi-modal situations is the likelihood that learners are less reliant on form-form links and that instead they acquire form-meaning connections early on. The final question that Chapter 5 of this thesis will try to answer is:

5) What are the implications of multi-modal incidental learning for current models of FL vocabulary learning?

The next section will describe the methods and paradigms used in the thesis to answer the five main research questions.

1.5.2. Methods and paradigm used in the thesis

1.5.2.1 Eye-tracking

As mentioned in the previous section, an important question investigated in this thesis is the allocation of attention during multi-modal situations. A reliable measure of overt attention, extensively used in psycholinguistic research is eye-tracking (Rayner, 2009). Using an eye-tracker, eye-movements are measured and parsed into blinks, saccades (the actual eye-movements) and fixations (when the eyes are not moving), from which a variety of dependent measures are derived. For example, the total fixation duration (also called gaze duration or dwell time) is calculated by adding the time spent with the eyes fixed within a pre-defined region (called region of interests, or 'ROI') at any time during a trial. Other dependent measures include first-pass reading time (the total duration of fixations occurring before the eyes first exit a ROI), number of regressions (number of times the eyes return to a ROI after the first-pass) number of fixations, average fixation duration, etc. Most psycholinguistics eye-tracking research uses fixations to derive dependent measures, as this is when the viewer acquires new information. When the eyes are moving across a visual display, no new information can be acquired, as the eyes are moving too fast, although it is thought that cognitive processes started during fixations continue during saccades (Rayner, 2009). Eye-tracking methodology will be used in this thesis to measure the allocation of attention whilst participants are engaged in multi-modal situations, namely a film with subtitles in Chapter 2, and a simple multi-modal situation with written and auditory words and pictures in Chapter 6. In both cases, the screen area will be divided into different ROIs, and these will be used to compare dependent measures across conditions. This will be explained in more detail in the Methods section of each chapter.

1.5.2.2 Savings Paradigm

A potential contributing factor to the modest amount of incidental vocabulary acquisition reported in studies using multi-modal situations as well as studies of reading, listening and reading-while-listening, is the measuring instruments used to detect the learning effects. The level of word knowledge measured, as well as whether this knowledge was measured at the receptive or the productive level also varies considerably across studies. Overall, researchers have used typical recognition and recall measures, but as the results of the EEG experiments reviewed in Section 1.3.1 suggested, it may be that more sensitive methods for detecting early learning effects are needed. Although a simple recognition measure was used to assess learning in Chapter 2, it became necessary following this initial study to develop a more sensitive measure of vocabulary acquisition. Therefore, a potentially more sensitive measure of early learning effects based on the savings paradigm (Ebbinghaus, 1964) will be introduced in Chapter 3. The savings paradigm has been shown to be more sensitive than typical recognition and recall tests (Groninger & Groninger, 1980; MacLeod, 1988; Nelson, 1978, 1985) and has been used successfully in recent studies of language attrition to detect traces of knowledge (Bowers, Mattys, & Gage, 2009; De Bot, Martens, & Stoessel, 2004; De Bot & Stoessel, 2000; Hansen, Umeda, & McKinney, 2002). The idea of the savings paradigm originally comes from Ebbinghaus who noticed that once something had been learnt, even after a long delay, a certain amount of residual knowledge remained in memory (referred to as the "forgetting curve"); this residual memory trace facilitated relearning by reducing the number of trials to criterion, a phenomena now known as “savings” (Ebbinghaus, 1964). Importantly, in contrast to studies on language attrition, the savings paradigm will be used in Chapters 3 to 6 to detect traces of new FL vocabulary knowledge that has not necessarily reached the threshold for explicit recognition or recall. It was thought that through mere exposure during an initial incidental learning phase, participants would start building some knowledge about FL words, and that this would help them reach the recognition threshold faster for these words then for completely new words during a subsequent explicit learning phase (see Figure 1.3). This paradigm will be explained in more detail in Chapter 3.

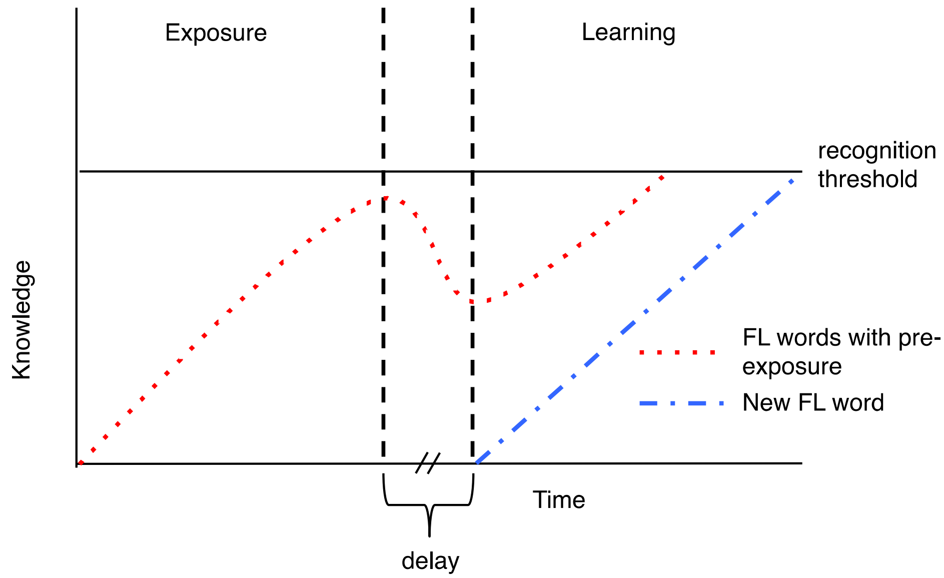

Figure 1.3: Illustration of the savings in learning due to pre-exposure.

1.5.2.3 Cross-modal priming

In order to investigate the lexical-semantic connections occurring from exposure to multi-modal situations (see Section 1.5.1), a measure that did not require explicit activation of lexical and semantic information was needed. I therefore used a priming paradigm to investigate this question. Furthermore, because of the multi-modal incidental learning situation with auditory, written and pictorial information, it was necessary to use a priming task that included multiple modalities. This type of priming task is called cross-modal priming, and has been used to investigate the time-course of interaction between the written and auditory modalities during NL word recognition (Kiyonaga, Grainger, Midgley, & Holcomb, 2007; Holcomb, Anderson, &

Grainger, 2005), as well as cross-modal semantic interactions (Holcomb & Anderson, 1993). Dobel et al. (2009) used cross-modal priming with spoken word form primes and picture targets whilst participants were engaged in a semantic categorization task to investigate the neural correlates of novel word learning (pseudowords in the NL). In Chapter 5, I used a similar paradigm as Dobel et al., with FL auditory primes from the incidental learning situation along with picture targets to explore the form-meaning links hypothesis. In addition, I included FL auditory primes with NL written word targets to explore the possible involvement of FL-NL lexical links. This auditory prime with written target combination is similar to Holcomb and Anderson (1993), except that as well as being cross-modal, the priming task used in this thesis included translation priming from FL to NL.

1.5.3. Overview of thesis

This thesis includes five experimental chapters, each reporting a large experiment. These experiments included multiple groups of participants, and involved data collection over several experimental sessions, and using both behavioral and eye-tracking methods. As mentioned in Section 1.5.1, some experiments were designed to answer a specific question, whilst others provided evidence for multiple questions. To summarize the content of the thesis, Chapter 2 addressed incidental vocabulary acquisition from watching a film with subtitles as well as attention allocation during a film with subtitles. Chapter 3 investigated whether exposure to simple multi-modal stimuli lead to incidental vocabulary acquisition using a sensitive method of detecting early learning effects. In Chapter 4, the role of frequency of exposure in multi-modal FL vocabulary learning was investigated, as well as the number of exposures necessary in order for incidental learning to occur. In Chapter 5, the type of connections established during incidental vocabulary learning was investigated using a priming paradigm. Chapter 6 addressed the allocation of attention to written and pictorial stimuli during a simple multi-modal situation using eye-tracking, as well as and its impact on subsequent explicit learning. A general discussion of the findings ensued in Chapter 7.

Chapter 2: Incidental vocabulary acquisition and attention allocation while watching a film with subtitles [3]

2.1 Introduction

As was mentioned in Chapter 1 (see Section 1.1), it is estimated that in order to achieve fluency in a FL, learners need to know approximately 8000 word families, which represents approximately 35 000 words (Nation, 2001; Schmitt 2008, 2010). There are therefore more words to learn than can be realistically taught through educational context (de Groot & van Hell, 2005; Horst, 2005), thus, learners are often encouraged to read in the FL in order to increase their incidental acquisition of FL vocabulary. Although more research has been conducted on incidental acquisition of FL vocabulary through reading (e.g., Brown, et al., 2008; Horst, 2005; Hulstijn et al., 1996; Kweon & Kim, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Rott, 1999; Vidal, 2011; Waring, & Takaki, 2003), learners can also acquire FL words incidentally through listening (Brown et al., 2008; Vidal, 2011), or through reading-while-listening (Brown et al., 2008; Horst et al., 1998; Webb et al., 2013). Many studies of incidental vocabulary learning from unimodal (reading or listening, see Section 1.3.1 and 1.3.2) and bimodal (see Section 1.3.3) inputs of information were reviewed in detail in Chapter 1. The results of

[3] The data presented in this chapter has been published in:
Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2012). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 1–20.

these studies showed that exposure to FL from written and spoken input does lead to incidental vocabulary acquisition.

Although reading, listening and reading-while-listening can be useful to increase one's exposure to FL, these type of incidental learning situations are only suitable for more advanced FL learners, as the meaning of new words must be derived from the context. In order to achieve this, learners must already know most of the words in a text (a figure of 95% coverage was put forward by Nation in 2001). In contrast, a multi-modal situation including both verbal and pictorial information can be beneficial even for complete beginners in a FL, as FL word meaning can be derived from the pictures. Furthermore, in a multi-modal situation, both auditory and written verbal information can be presented concurrently with the pictorial information. For example, in a film with subtitles, verbal information is presented through the soundtrack and the subtitles (either in a FL or a NL), whilst the film's images provide the pictorial information. In the case of standard subtitling, FL learning may occur by linking the FL phonological word forms from the soundtrack to the meaning representations accessed through the film's images, or by linking the FL phonological word forms to the NL translation equivalents provided in the subtitles. In the case of reversed subtitling, as well as using the pictorial information to access word meaning, the FL orthographic word forms may be linked to the NL words from the soundtrack. Finally, with intralingual subtitles, both FL phonological and orthographic word forms are presented through the soundtrack and subtitles respectively, and therefore word meaning can only be extracted from the film's visual images.

Although films with subtitles provide an information rich multi-modal situation with considerable FL exposure, it is unclear whether in such a situation, FL vocabulary learning does occur. Some prior research suggests that watching films with subtitles can be beneficial to learners, unfortunately, many studies were hampered by methodological issues, or used films with subtitles in an explicit learning situation (i.e., participants were aware of an upcoming vocabulary test or were told to pay particular attention to the words; see Section 1.3.4 and Tables 1.6 and 1.7 in Chapter 1 for a review). Putting these limitations aside, studies conducted with films with intralingual subtitles (Boras & Lafayette, 1994; Koskinen et al., 1996; Mitterer & McQueen, 2009; Neuman & Koskinen, 1992; Sydorenko, 2010; Winke et al., 2010) all support their usefulness for FL vocabulary learning. In contrast, the work of Danan (1992), d'Ydewalle and Pavakanun (1995, 1997), Lambert et al. (1981) and Lambert and Holobow (1984) provide evidence of FL vocabulary learning through reversed subtitling. Finally, studies conducted by Koolstra and Beentjes (1999) and d'Ydewalle and van de Poel (1999) suggest that standard subtitling leads to the acquisition of FL vocabulary. Although there is evidence in favor of each of the subtitling conditions, research to date does not indicate which subtitling condition is more favorable for incidental FL vocabulary acquisition. Furthermore, it is unclear how much FL vocabulary can be learnt incidentally from watching films with subtitles.

One aim of the current study was therefore to assess the impact of exposure to films with subtitles on the incidental acquisition of FL vocabulary. Participants were asked to watch a film in one of the subtitling conditions following which an unexpected auditory vocabulary test was administered to measure the acquisition of FL vocabulary.

Overall, the evidence from the studies of films with intralingual subtitling mentioned above, as well as research on bimodal input of information (Bird & Williams, 2002; Brown et al., 2008, Horst et al., 1998; Hu, 2008; Ricketts et al., 2009; Rosenthal & Ehri, 2008; Webb et al., 2013; also see Table 1.5 in Chapter 1) suggest that intralingual subtitling would be the most beneficial for vocabulary acquisition. In this subtitling condition, it is possible for learners to map the FL phonological word forms onto their orthographic word forms, and the subtitles may help learners segment the stream of auditory FL into word units. It was expected that standard and reversed subtitling would also lead to some vocabulary acquisition, however, as the presence of the NL may distract from the processing of the FL, it was expected that less vocabulary acquisition would occur than in the intralingual condition. In addition, all subtitling conditions should lead to more vocabulary acquisition than watching the film without subtitles. Finally, in order to ensure that the results of the vocabulary test can be attributed to incidental learning, an additional group of participants that did not watch the film also completed the vocabulary test.

Although one focus of the study was on the acquisition of FL vocabulary, another important question concerned the allocation of attention in the multi-modal situation. If films with subtitles are to be used to promote the acquisition of FL vocabulary, it is important to ascertain first to what extent the subtitles are read. Early studies on the processing of subtitles in films suggested that at least some of the subtitles were read in both standard subtitling and NL intralingual subtitling (d'Ydewalle et al., 1985; d'Ydewalle et al., 1991; d'Ydewalle & van Rensbergen, 1989; d'Ydewalle et al., 1987). Unfortunately, many methodological details are missing from these early studies, making them difficult to replicate (see Chapter 1, Section 1.3.4). More recent evidence showed that approximately 41% of the presentation time in standard subtitling and 26% of the presentation time in reversed subtitling conditions was spent reading the subtitles (D'Ydewalle & De Bruycker 2007). So far, there are no studies investigating the reading of intralingual FL subtitles. Furthermore, the results of d'Ydewalle and De Bruycker, may include time spent in the subtitle area because of the visual aspect of the movie taking place there rather than reading the subtitles. The second aim of the study was therefore to investigate learners viewing behavior whilst watching a film with subtitles, and to shed light on the amount of time devoted to looking at each type of information (verbal and visual) in each subtitling condition. In order to answer these questions, participant's eye-movements were recorded whilst they watched the film in one of the subtitling conditions. Furthermore, a control group watching the same film without subtitles was included in the study to exclude the time spent looking in the subtitle area as part of visual information of the movie taking place there. Based on the previous studies mentioned above, it was predicted that participants would read the subtitles regardless of the subtitling condition, but that reading behavior, as assessed by the number and duration of the fixations, would differ in each subtitling condition. More specifically, the duration and number of fixations should be higher in the standard condition as the subtitles are in the NL, followed by the intralingual condition where both subtitles and soundtrack are in a FL, and finally the reversed condition where the FL subtitles are superfluous and the FL is unknown to the participants.

Dutch was chosen as the FL in the film as most people in the United Kingdom have no knowledge of it. However, because Dutch comes from the same Germanic language family as English, the two languages have many similarities, which could aid the acquisition of Dutch vocabulary.

2.2 Method

2.2.1. Participants

Sixty-four participants took part in this experiment and they received £6 for their participation. All participants completed a self-reporting language questionnaire to verify their language background and to ensure that they were native English speakers without any knowledge of Dutch. Because of the information reported on the language questionnaire, ten participants had to be excluded for the following reasons: two participants had reading difficulties, three participants were not monolingual, three participants were not native English speakers and one participant had spent 74 days in a Dutch speaking country.

In total, 36 participants (mean age 24.6, 26 females) were included in the eye-movement analysis and 54 participants (mean age 24.2, 39 females) were included in the vocabulary test analysis. Even though 54 participants qualified for the experiment, 11 were in the "no movie" condition and as such no eye-movement data was collected for them. Of the 43 participants in one of the movie conditions, eye-tracking data was not recorded for 7 of them due to failed calibrations or software problems. When this was the case, participants still watched the film and took part in the vocabulary test.

2.2.2. Design

This experiment had a between-subject design with four conditions: 1) control condition: Dutch audio without subtitles (DA-NS), 2) intralingual subtitling: Dutch audio with Dutch subtitles (DA-DS), 3) standard subtitling: Dutch audio with English subtitles (DA-ES) and 4) reversed subtitling: English audio with Dutch subtitles (EA-DS). Another group of participants did not watch the movie and only took part in the vocabulary test (No-Movie).

2.2.3. Stimuli

The film excerpt lasted 25 minutes and consisted of four DVD chapters (chapters 2 to 5) of the movie SpongeBob Square Pants (Hillenburg, 2004). An animated film was chosen as the language of the soundtrack can easily be changed without affecting lip-synchronicity. Four different versions of the movie were created and reformatted into avi-files, which are compatible with the Experiment Builder software. The DA-NS version was created using the original movie chapters in Dutch. The DA-DS version was also created using the original movie chapters in Dutch to which a transcription of the Dutch audio soundtrack was added as subtitles. It was necessary to create my own Dutch subtitles (generated by a native-Dutch speaker) in order to have a true word-for-word transcription instead of the abbreviated subtitles which came with the movie DVD. I did, however, use the original subtitle timing information to add a transcription to the movie using the subtitling program Submerge. The DA-ES version was created by using the original English subtitles and reformatting them using Submerge in order to match the DA-DS version with regards to the font, size, color and display settings of the subtitles. The EA-DS version was created by adding the Dutch subtitles to the English soundtrack, as with the previous two versions.

The vocabulary test consisted of two lists of 78 items (see Appendix A). The items were one word audio extracts selected from the movie soundtrack in Dutch. They varied in frequencies of occurrence in the movie (from 1 to 115 in the Dutch version of the movie and from 2 to 132 in the English version of the movie) and type of words (30 nouns, 16 pronouns, 10 adverbs, 6 adjectives, 6 verbs and 10 other word types.) Each item was presented once with the correct English translations in the match trials (e.g., *koning*-king) and once with a foil in the mismatch trials (e.g., *koning*-coming). The foils were written English words that were phonologically similar to the Dutch ones and were taken from the movie's English subtitles wherever possible. In cases where no word in the English subtitles was phonologically similar, another English word, which was possible in the context of the movie, was chosen. This was done in order to keep the participants from relying solely on the phonological similarity or the context of the movie to complete the vocabulary test. The order of presentation of the lists was counterbalanced and each item appeared once in each list in a random order. Each list contained half match and half mismatch trials such that if an item was presented with their correct translation in List 1, it was presented with its foil in List 2. The allocation of match and mismatch trials for each list was pseudo-randomised.

2.3 Procedure

Participants were tested individually. They were informed that they would have to watch an animated film and they were shown the head-mounted eye-tracker that they would be wearing during the film viewing. They were told that the film could be either in English or in a FL and that there might be subtitles. They were asked to watch the film as they would normally do at home and they were told that there would be some questions to answer afterwards about the film and about themselves. Participants were not explicitly asked to read the subtitles nor were they told to pay attention to the FL. Furthermore, they were not told that there would be a FL vocabulary test after the film.

The eye-tracker was set up and calibrated using a nine-point calibration grid at the beginning of the session, and was re-calibrated after each chapter in the movie (4 times in total). The movie chapters were presented in the same sequential order for all participants in order to follow the storyline. After watching the animation, participants completed an auditory vocabulary test and a language questionnaire. In the vocabulary test, participants heard a one word extract from the Dutch movie soundtrack three times whilst viewing either the correct English translations on the screen or a foil. They then had to decide whether the word they heard in Dutch had the same meaning as the English word they saw. The entire experiment lasted approximately 1 hour.

Participant's eye-movements were recorded using an Eyelink 1 eye-tracker (SR Research, Canada). This is a head-mounted eye-tracker with a sampling rate of 250 Hz, equivalent to a temporal resolution of 4 ms. The eye-tracker algorithm parses the eye-movements into saccades (when movement of 0.5 degrees of visual angle or more is detected for 2 or more samples in a sequence), blinks (when the pupil data is missing for 3 or more samples in a sequence) and fixations (any period that is neither a saccade nor a blink), and records these events into an EDF file (Eyelink Data File) stored on the host computer. The movie chapters and vocabulary test were displayed on a PC using SR Research's Experiment Builder software and E-Prime respectively.

2.4 Results

The fixations were first split according to whether they occurred in the image area or the subtitle area. The subtitle area was taken as an area of 1024 pixels (the whole width of the screen) by 218 pixels, and started 50 pixels from the bottom of the display screen. This was deliberately larger than the actual area that displayed the subtitles to account for small vertical and horizontal inaccuracies in the recording of the eye-movements. The image area included an area of 1024 pixels by 450 pixels, and started 50 pixels from the top of the display screen. The subtitle's timing information was used to determine whether a fixation occurred during the presentation of a subtitle.

The eye-tracker output files were split into four chapters for each participant. Each movie chapter was then processed using the Eyetracker Output Utility program (van Heuven, 2010). The data from eight chapters (6% of the data) was excluded from the analysis because of excessive vertical drift, whereby all fixations are shifted up or down. Only chapters where this problem clearly occurred from the start were excluded. It was also necessary to discard some of the subtitles as they occurred while there was already some writing in English in the image or the subtitle area as part of the movie. This was the case for 10% of the Dutch subtitles and 8% of the English subtitles. Furthermore, the control condition (DA-NS) was used to exclude the subtitles for which other visual aspects of the movie co-occurred in the subtitle area. This involved a further 24% of Dutch subtitles and 20% of the English subtitles. In total, fixations occurring during 249 Dutch and 311 English subtitles were included in the analysis (see Table 2.1), which represents 66% and 71% of the original subtitles respectively. Furthermore, Dutch subtitles were presented during 47% of the 25-minute movie as opposed to 50% in the case of English subtitles. Subtitles comprised either one or two lines of text of different lengths and were presented on average for 3 seconds for the Dutch subtitles and 2.5 seconds for the English subtitles. An example of subtitle presentation and reading pattern for one participant in each subtitling condition is presented in Figure 2.1.

Table 2.1: Number of subtitles with exclusions, average subtitle duration in milliseconds, average number of words per subtitle and number of subtitles with one and two lines of text for each chapter of the movie for both Dutch and English subtitles.

	Dutch					English				
Chapter	1	2	3	4	total	1	2	3	4	total
All subtitles	108	76	102	89	375	128	89	121	98	436
Excluded										
English	13	7	4	12	36	13	6	4	13	36
action	45	15	19	11	90	40	16	20	13	89
N final	50	54	79	66	249	75	67	97	72	311
n 1 line of text	29	26	33	27	115	40	39	52	28	159
n 2 lines of text	21	28	46	39	134	35	28	45	44	152
M duration (ms)	2760	3093	3158	2759	2943	2410	2646	2568	2428	2513
M num. words	7.0	7.0	7.8	7.0	7.2	5.8	6.0	6.6	6.7	6.3

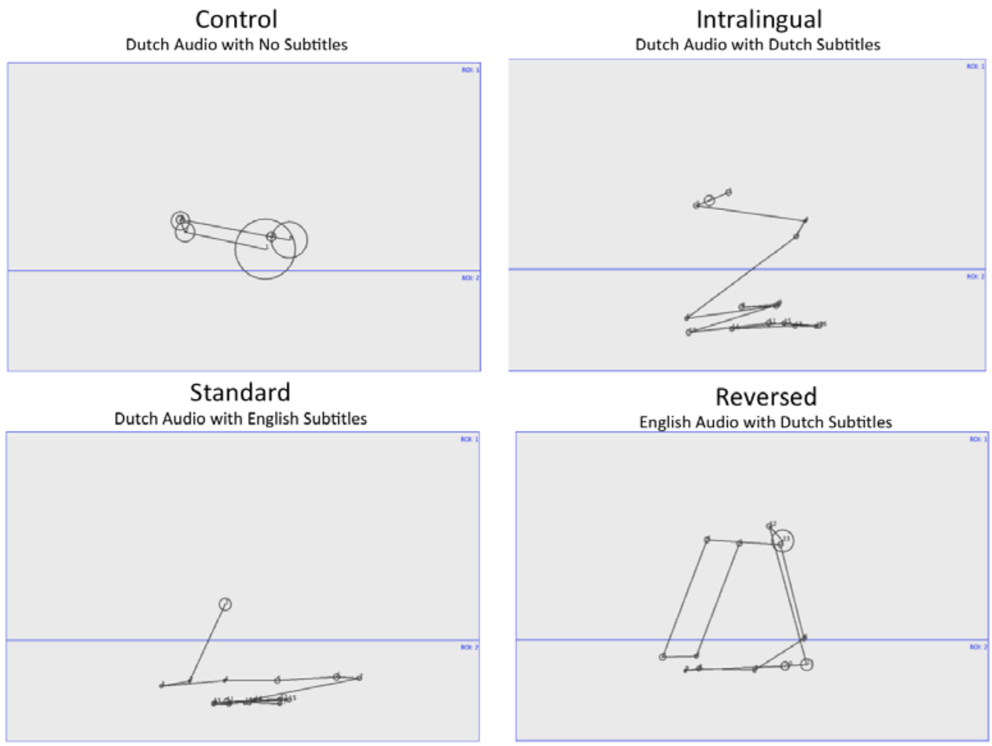

Figure 2.1: Example of fixation pattern (in black) for one participant in each subtitling condition. Fixations are represented by circles connected by lines (saccades), with larger circles indicating longer fixations. The blue lines separated the subtitle and image areas for the purpose of the analyses only.[4]

2.4.1. Subtitle area

The total fixation duration, the number of fixations and the average fixation duration were calculated for each subtitle and averaged for each participant. The number of skipped subtitles (subtitles that participants did not fixate upon) was also calculated for each participant. All of these variables were then averaged for each condition (see Table 2.2) and

[4] Background image removed to comply with copyright law.

submitted to a one-way ANOVA [5] with subtitle condition as a between-subject factor (4 levels: Dutch soundtrack with no subtitles (DA-NS), Dutch soundtrack with Dutch subtitles (DA-DS), Dutch soundtrack with English subtitles (DA-ES), and English soundtrack with Dutch subtitles (EA-DS)). Results revealed a significant main effect of subtitle condition for all four measurements: $F(3, 32) = 41.19$, $p < .001$, $\eta_p^2 = .79$ for the total fixation duration, $F(3, 32) = 49.98$, $p < .001$, $\eta_p^2 = .82$ for the number of fixations, $F(3, 32) = 5.49$, $p < .01$, $\eta_p^2 = .34$ for the average fixation duration and $F(3, 32) = 78.40$, $p < .001$, $\eta_p^2 = .88$ for the number of skipped subtitles. Tukey post-hoc analyses revealed that participants spent longer fixating in the subtitle area when they heard the Dutch soundtrack with either Dutch or English subtitles than when they heard the English soundtrack; $p < .001$ in both cases. The pattern was exactly the same for both the number of fixations and the number of skipped subtitles. However, there were no significant differences between the condition with the Dutch soundtrack with Dutch subtitles and the Dutch soundtrack with English subtitles (means for total fixation duration, number of fixations and skipped subtitles, all *ps* > .84). In addition, Tukey post-hoc analyses revealed that the average fixation duration was significantly longer for the condition

---

[5] Each dependent variable was first submitted to a mixed design ANOVA with condition as a between-subject factor (4 levels: Dutch soundtrack with Dutch subtitles (DA-DS), Dutch soundtrack with English subtitles (DA-ES), Dutch soundtrack with no subtitles (DA-NS), English soundtrack with Dutch subtitles (EA-DS)) and chapter as a within-subject factor (4 levels: chapter 1 to chapter 4) in order to rule out an effect of chapter or an interaction between chapter and condition. As only an interaction between chapter and condition was found for the number of skipped subtitles, and that this effect was driven by the control condition (Dutch audio with no subtitles) the dependent variables were collapsed across chapters for all other analyses.

without subtitles (DA-NS) than both conditions with subtitles and Dutch soundtrack (DA-DS, $p < .05$ and DA-ES, $p < .01$) but the difference was only approaching significance between the condition without subtitles and the condition with English soundtrack and Dutch subtitles (EA-DS, $p = .059$). The average fixation duration was not significantly different between the conditions with subtitles (DA-DS, DA-ES and EA-DS, all *p*s > .68). In order to confirm that the experimental groups were looking in the subtitle area because of the presence of the subtitles rather than because of other visual aspects of the movie taking place there, each subtitled condition (DA-DS, DA-ES and EA-DS) was compared with the no subtitles control group (DA-NS). Tukey post-hoc analyses showed that participants spent significantly longer fixating in the subtitle area for each experimental group than for the control group ($p < .001$ for each comparison). This significant difference between the experimental and control groups was also found for the number of fixations and the number of skipped subtitles (all *p*s < .001).

Table 2.2: Mean (*M*) and standard error (*SE*) for the total fixation duration (in milliseconds), the number of fixations and the average fixation duration in the subtitle area and in the image area, and mean (*M*) and standard error (*SE*) for the number of skipped subtitles for each subtitling condition.

		Subtitle Area				Image Area		
		Tot. fix dur.	Num. fix.	Num. skipped	Avg. fix. dur.	Tot. fix dur.	Num. fix.	Avg. fix. dur.
Condition	*n*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*
DA-NS (no subtitles)	8	81 (17)	0.28 (0.07)	186.88 (10.60)	283 (15)	2516 (54)	6.70 (0.32)	450 (36)
DA-DS (intralingual)	9	1285 (99)	5.65 (0.43)	20.78 (5.76)	240 (6)	1222 (88)	4.34 (0.33)	303 (18)
DA-ES (standard)	11	1284 (63)	5.90 (0.28)	10.82 (3.40)	227 (9)	832 (66)	3.16 (0.23)	274 (11)
EA-DS (reversed)	8	786 (135)	3.30 (0.54)	65.25 (15.24)	243 (10)	1738 (159)	5.56 (0.57)	350 (37)

Note. Tot. fix. dur. = total fixation duration; Num. fix. = number of fixations; Num. skipped = number of subtitles for which there was no fixations in the subtitle area; Avg. fix. dur. = average fixation duration; DA-NS = Dutch audio with no subtitles; DA-DS = Dutch audio with Dutch subtitles (intralingual subtitling); DA-ES = Dutch audio with English subtitles (standard subtitling); EA-DS = English audio with Dutch subtitles (reversed subtitling).

As the duration of the subtitles' presentation differed between the conditions with Dutch and English subtitles, normalised total fixation durations in the subtitle area were calculated by dividing the total fixation duration in the subtitle area by the duration of the presentation for each subtitle. This was then averaged for each participant and condition (see Table 2.3) and submitted to a one-way ANOVA. Results showed that there was a main effect of condition, $F(3,32) = 48.85$, $p < .001$, $\eta_p^2 = .82$. Tukey post-hoc analyses revealed the same pattern as above, with every condition being significantly different from each other (all ps < .01) except for the Dutch audio with Dutch subtitles condition and Dutch audio with English subtitles (DA-DS vs. DA-ES, $p = .26$).

Table 2.3: Mean normalised total fixation duration and mean normalised number of fixations in the subtitle area and normalised number of skipped subtitles with standard error (*SE*), minimum and maximum values for each subtitling condition.

		Normalised total duration			Normalised fixations			Normalised skipped subtitles		
Condition	*n*	*M (SE)*	*min*	*max*	*M (SE)*	*min*	*max*	*M (SE)*	*min*	*max*
DA-NS (no subtitles)	8	0.03 (0.01)	0.01	0.07	0.06 (0.01)	0.02	0.14	0.82 (0.03)	0.65	0.95
DA-DS (intralingual)	9	0.43 (0.03)	0.30	0.58	0.92 (0.06)	0.72	1.3	0.09 (0.03)	0	0.25
DA-ES (standard)	11	0.51 (0.03)	0.33	0.65	1.04 (0.04)	0.78	1.23	0.04 (0.01)	0	0.09
EA-DS (reversed)	8	0.27 (0.05)	0.11	0.46	0.59 (0.09)	0.26	1.03	0.26 (0.06)	0.02	0.47

Note. DA-NS = Dutch audio with no subtitles; DA-DS = Dutch audio with Dutch subtitles (intralingual subtitling); DA-ES = Dutch audio with English subtitles (standard subtitling); EA-DS = English audio with Dutch subtitles (reversed subtitling).

Similarly, because of the unequal number of words between the conditions with Dutch and English subtitles, normalised numbers of fixations were calculated by dividing the number of fixations by the number of words for each subtitle. This was then averaged for each participant and each condition (see Table 2.3) and submitted to a one-way ANOVA. Results indicated that there was a significant main effect of condition, $F(3,32) = 56.42$, $p < .001$, $\eta_p^2 = .84$. Tukey post-hoc analyses again revealed that differences between each condition were significant for the number of fixations even after controlling for the number of words in the subtitle (all *p*s < .01) except for the Dutch audio with Dutch subtitles condition and Dutch audio with English subtitles (DA-DS vs. DA-ES) which was again not significant, $p = .44$.

Furthermore, a normalised number of skipped subtitles was calculated by dividing the number of skipped subtitles by the total number of subtitles presented for each participant. This was then averaged for each condition and submitted to a one-way ANOVA. The results showed once more a main effect of condition, $F(3,32) = 107.59$, $p < .001$, $\eta_p^2 = .91$. Tukey post-hoc analyses continued to show the same pattern with the Dutch audio with Dutch subtitles and Dutch audio with English subtitles conditions (DA-DS vs. DA-ES) not being significantly different ($p = .59$). All other conditions were significantly different from each other (all *p*s < .01).

In an attempt to distinguish the reading behavior in the conditions with Dutch audio and either Dutch or English subtitles (DA-DS and DA-ES), the number of consecutive fixations in the subtitle area was calculated for each participant for each subtitle (these ranged from 2 to 20 consecutive fixations). The proportion of consecutive fixations was then calculated by adding up the number of consecutive fixations and dividing by the total number of fixations in the subtitle area. This process was repeated for each minimum number of consecutive fixations and averaged for each condition. For example, the proportion of fixations in the subtitle area for which there were 2 or more consecutive fixations was .91 for the Dutch subtitle group and .94 for the English subtitle group, but the proportions dropped to .41 and .49 for 7 or more consecutive fixations for the Dutch and English subtitle groups respectively (see Figure 2.2). Chapter 4 was chosen for this analysis as it is the chapter with the biggest sample size and the most closely matched in terms of average number of words per subtitle between the Dutch and English subtitling conditions. An independent sample *t*-test (two-tailed) for each minimum number of consecutive fixations was calculated and the results showed no significant differences between the two conditions for any minimum number of consecutive fixations (all $ps > .17$).

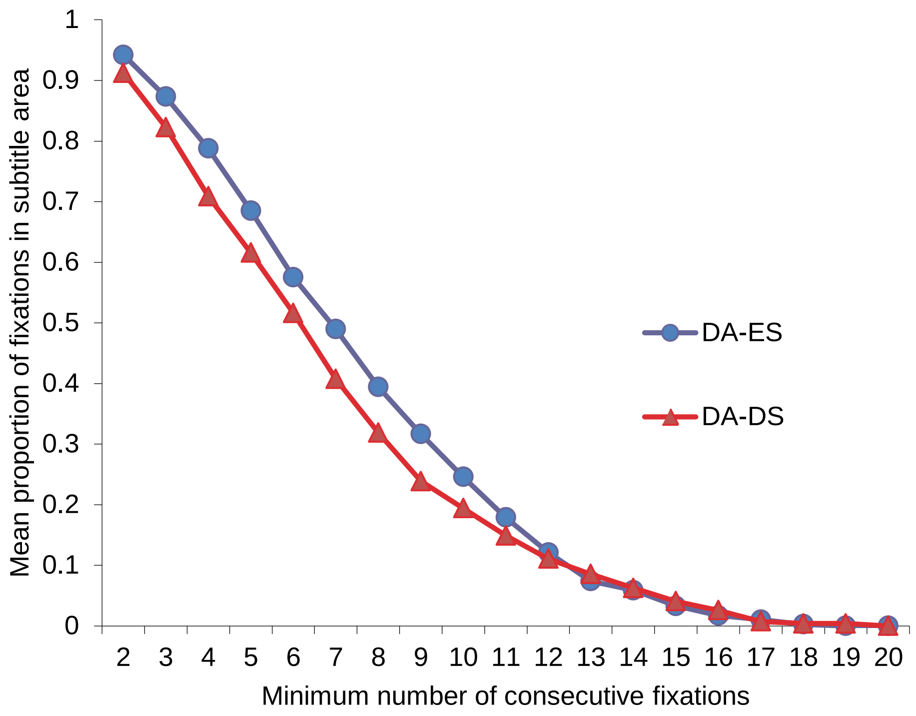

Figure 2.2: Mean proportion of fixations in the subtitle area which were consecutive for chapter 4 for the conditions with Dutch audio with English subtitles (DA-ES) and Dutch audio with Dutch subtitles (DA-DS).

One of the reasons why participants may read the subtitles in the conditions with Dutch subtitles is the presence of Dutch-English orthographically similar or identical words. For example, the Dutch word "*kind*" is orthographically identical to the English word "kind" and the Dutch word "*promotie*" is orthographically similar to the English word "promotion". Orthographically similar words can have different meanings if they are interlingual homographs (e.g., Dutch-English word "kind" meaning child in Dutch), or have the same meaning if they are cognates (e.g., "*promotie*" / "promotion"). To examine overlap in orthography regardless of meaning, the number of Dutch words that were either identical or similar to English words in each subtitle was calculated using the normalised Levenstein distance, which is a measure of orthographic similarity [6] (Schepens, Dijkstra, & Grootjen, 2012). This was then correlated with the normalised number of fixations and the normalised fixation duration in the subtitles while controlling for the number of words in the subtitles (partial correlation). The result of the partial correlations showed that there was no significant relationship between the number of identical or similar Dutch-English words and the normalised number of fixations for both the condition with Dutch soundtrack and Dutch subtitles (r = -.02, p = .76) and English soundtrack and Dutch subtitles (r = -.05, p = .42). Furthermore, there was no relationship between the number of identical or similar Dutch-English words and the normalised duration of fixations (r = .10, p = .09 for the condition with Dutch soundtrack and Dutch subtitles and r =.05, p =.40 for the condition with English soundtrack and Dutch subtitles).

2.4.2. Image area

The total fixation duration, the number of fixations and the average fixation duration in the image area were calculated during each subtitle presentation and averaged for each participant and each condition (see Table 2.2) and submitted to a one-way ANOVA with subtitling condition as a between-subject factor (4 levels). The results indicated a significant main effect of subtitling condition on the total fixation duration, $F(3, 32)$ = 58.49, $p < .001$, η_p^2= .85, the number of fixations, $F(3, 32)$ = 18.41, $p < .001$,

[6] The normalized Levenstein distance scores range from zero (orthographically identical) to one (completely different). The criteria used to classify the words in this study was < .3.

η_p^2 = .63, and the average fixation duration, $F(3,32) = 9.11$, $p <.001$, η_p^2= .46. Tukey post-hoc analyses revealed that participants watching the film without subtitles (DA-NS) spent more time fixating in the image area than the participants with the English soundtrack and Dutch subtitles condition (EA-DS, $p < .001$) and they in turn spent more time in the image area than participants with the Dutch soundtrack and Dutch subtitles condition (DA-DS, $p < .01$). However, the latter still spent more time in the image area than participants with the Dutch soundtrack and English subtitles (DA-ES, $p < .05$).

However, Tukey post-hoc tests revealed that the number of fixations in the image area was significantly higher for the condition without subtitles compared to both conditions with Dutch audio and either Dutch or English subtitles (DA-DS and DA-ES, both *p*s < .001), but this was not the case when comparing with the condition with English audio and Dutch subtitles (EA-DS, $p = .17$). There were also no significant differences between the conditions with Dutch audio and either Dutch or English subtitles (DA-DS vs. DA-ES, $p = .09$) or between the conditions with English audio or Dutch audio with Dutch subtitles (EA-DS vs. DA-DS, $p = .12$). However, the number of fixations in the image area was significantly higher in the condition with English soundtrack as opposed to English subtitles (EA-DS vs. DA-ES, $p < .001$).

In addition, Tukey post-hoc tests showed that the average fixation duration in the image area was significantly longer for the condition without subtitles (DA-NS) compared to the conditions with Dutch audio and Dutch or English subtitles (DA-DS and DA-ES, $p < .01$), but only marginally so for the condition with English audio and Dutch subtitles (EA-DS, $p = .059$). There were no significant differences between the conditions with subtitles (DA-DS, DA-ES and EA-DS, all *ps* > .16).

2.4.3. Vocabulary test

The percentage of correct responses on the vocabulary test for each participant was averaged for each condition (see Figure 2.3). The mean scores were then submitted to a one-way ANOVA which revealed no main effect of subtitling condition, $F < 1$. However, one-sample *t*-tests showed that the mean scores for each conditions were significantly higher than chance (all *ps* < .001).

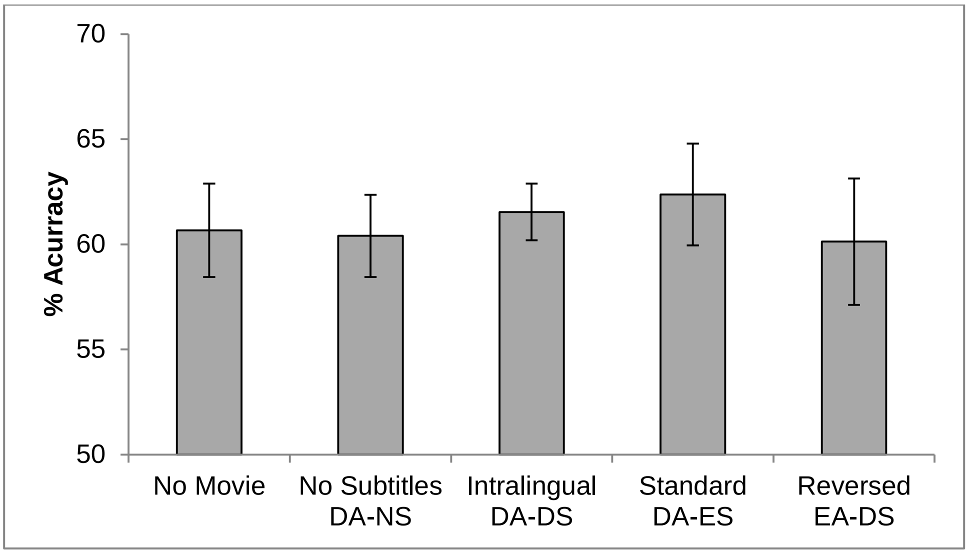

Figure 2.3: Percentage accuracy in the vocabulary test for each condition with error bars.

2.5 Discussion

One of the aims of the present study was to investigate the reading of subtitles while watching FL films with standard, reversed and intralingual (FL) subtitles using eye-tracking. One of the predictions was that when the soundtrack was in the FL, participants would spend more time reading the subtitles if they were in their NL (standard subtitling) as opposed to in the FL (intralingual subtitling), but no significant differences were found for the duration of the fixations nor for the number of fixations in the subtitle area between those two conditions. Furthermore, no differences were found in the number of skipped subtitles between these two conditions.

The average fixation durations for the standard, intralingual and reversed subtitling conditions (227 ms, 240 ms and 243 ms respectively) were in line with the average fixation duration (between 225 and 250 ms) for silent reading (Rayner, 1998, 2009). Fixations are normally slightly longer in the case of listening while reading (Rayner, 1998) which would have been expected at least with intralingual subtitles, as participants could have followed the FL words more closely in the subtitles as they heard them in the FL soundtrack. Another reason why longer average fixation durations might be expected in the case of the intralingual and reversed subtitling conditions is that participants were reading in an unknown FL. As this was the case, their average fixation durations were expected to be similar to the longer durations associated with low-frequency words, which normally reflect extra processing effort (Inhoff &

Rayner, 1986; Rayner, 1998; Rayner & Raney, 1996). However, even though the average fixation durations were higher in both the intralingual and reversed subtitling conditions, these differences were not significant.

A possible explanation as to why I did not find differences in the average fixation durations is the lack of familiarity with the subtitling situation for the participants in the current experiment. In their study with adults and children from a subtitling country [7], d'Ydewalle and De Bruycker (2007) observed shorter than average fixation durations in both the standard (179 ms) and reversed subtitling conditions (193 ms) for their adult participants. In the current study, fixation durations were similar to those obtained by the children in the d'Ydewalle and De Bruycker experiment (247 ms and 261 ms for the standard and reversed conditions respectively) who did not have as much experience reading subtitles as the adults in the same experiment. It is unlikely that the participants in the current study watched subtitled films regularly [8], which could explain why their average fixation durations are similar to those of children in d'Ydewalle and de Bruycker (2007).

[7] Subtitling countries in Europe include Belgium, Denmark, Finland, Greece, Luxembourg, The Netherlands, Portugal and Sweden, as opposed to dubbing countries such as Austria, France, Germany, Italy and Spain (Koolstra et al., 2002).

[8] Although I did not ask participants about their audio-visual watching habits, as most of them are living in the United Kingdom which does not often use subtitles, it is likely that they do not have as much experience with subtitles as someone living in a subtitling country.

In the current study, even though participants in both the standard and intralingual subtitling conditions read most of the words in the subtitles, as indicated by the normalised number of fixations scores of about 1, they did not use the full subtitle presentation time to read the subtitles; instead they returned to view the image. This is even more evident in the intralingual condition where they could have benefited from a longer presentation time (due to the fact that there were more words in the subtitles) to read the subtitles, but instead they used the extra available time to look at the image. This seems to support prior results indicating that the reading of subtitles still allows for the processing of the images (Perego, Del Missier, Porta, & Mosconi, 2010) instead of being just a reading exercise (Jensema et al., 2000).

The reading behavior in the standard and intralingual subtitling conditions was investigated further by looking at the consecutive fixations, and it was found that there were no significant differences between the two. This seems to indicate that participants were not merely attracted to specific words in the subtitles, but that once their gaze moved to the subtitle area they read in a normal, uninterrupted fashion instead of alternating between the image and the subtitle areas. Taken together, this seems to suggest that the reading behavior was similar across these two conditions. This is surprising since participants in the intralingual condition had no prior knowledge of the FL (Dutch) used in the subtitles. It is possible that participants were simply trying to use the available information in the subtitles to understand the movie since they could not rely on the FL soundtrack for understanding. On the other hand, participants in the reversed condition did not need the FL subtitles to help them understand the film as their NL was present in the soundtrack. However, they still spent a considerable amount of time in the subtitle area.

There may be different reasons why participants read subtitles. In view of the results of the current experiment, it seems unlikely that the appearance of subtitles on its own explains the reading behavior since the average normalized number of fixations was 0.92 for the intralingual condition (nearly 1 fixation per word) and 0.59 for the reversed condition (more than half of the words were fixated upon). The possibility that participants read the subtitles because of Dutch-English orthographically similar or identical words (cognates or interlingual homographs) was also ruled out, as there was no correlation between the normalized number or duration of the fixations and the number of orthographically similar words. However, it is possible that the dynamic nature of the subtitles, i.e., the appearance and disappearance of the subtitles on the screen, coupled with the fact that the subtitles contained words was enough to generate the reading behavior. Moreover, the automatic reading of words is well established (Laberge & Samuels, 1974; Samuels, 1994) and has an effect even in tasks such as the Stroop task in which reading is irrelevant (MacLeod, 1991; Stroop, 1935). However, it was surprising to find that this automatic reading behavior also seems to occur for FL words that were unknown to participants and that this occurred throughout the study.

Both Dutch and English use an alphabetical script and this may have helped triggering the reading behavior. Furthermore, participants might have applied English spelling-to-sound rules to read the Dutch subtitles because of the similarities of the orthographic patterns between the two languages, which could have contributed to the reading behavior. Studies using languages with minimal orthographic (e.g., alphabet vs. non-alphabetic scripts) and phonological similarities could be used in the future to investigate this further.

Another way of explaining the reading of the subtitles is their saliency. Saliency maps have been used to understand the deployment of visual attention in static scene perception (Itti & Koch, 2000, 2001) and more recently in the area of dynamic scene perception (Mital, Smith, Hill, & Henderson, 2010). According to such research, the most salient feature attracts the viewer's gaze which, in the case of dynamic scenes, is motion. Even when the goal of the viewer is oriented elsewhere, this salient feature will trigger an automatic saccade towards it. This supports the idea that participants may be attracted to the subtitle area firstly because of the dynamic nature of the subtitles. This might also explain why some subtitles were skipped: it is possible that subtitles are more likely to be skipped if their appearance on the screen co-occurs with a more salient feature in the image area. In addition, it has been shown that text in a visual scene is also salient and therefore likely to attract participant's gaze, even when it is not relevant to the task at hand (Cerf, Frady, & Koch, 2009), which could account for the reading of the subtitles even in the conditions where the subtitles are in FL. In future research, it would be interesting to use saliency maps with the different subtitling conditions in order to investigate the factors influencing the reading of the subtitles.

The reading of the subtitles might also be influenced by the redundancy of information in the image, the soundtrack and the subtitles. D'Ydewalle and de Bruycker (2007) observed that participants spent proportionally less time reading one-line subtitles compared to two-line subtitles in the standard subtitling condition. They argued that this is because some information in one-line subtitles is redundant as it is often already available in both the image and the soundtrack. As far as I know, the research on audio-visual redundancy has not yet looked at films with subtitles and future studies could therefore investigate the influence of redundancy on the reading of subtitles.

Finally, there is also the possibility that participants read the subtitles because they believed this was expected of them, although they were not explicitly told to read subtitles. If this were the case, they presumably would have all read them regardless of the experimental condition. The data presented in this chapter shows, however, that participants read most of the subtitles in the intralingual and standard conditions, but only read some of the subtitles in the reversed condition. Thus, the assumption that they simply read the subtitles because they believed it was expected of them cannot account for the pattern of results.

The other aim of the current study was to investigate the incidental acquisition of FL vocabulary using an auditory vocabulary test. The mean percentages of correct responses on the vocabulary test were higher than chance, but as this was the case even when the movie was not watched, there is no evidence of vocabulary acquisition. The lack of differences across the conditions might be due to the limited exposure to the FL (25 minutes). However, because the vocabulary test measured knowledge at the recognition level only, it is possible that the participants did acquire some vocabulary knowledge, but that it did not reach the recognition level. Future studies should use a more sensitive measure of vocabulary acquisition.

2.6 Conclusion

To summarize, the results indicate no significant differences between the standard and the intralingual subtitling conditions in terms of the fixation duration, the number of fixations, the number of skipped subtitles or the number of consecutive fixations in the subtitle area, while participants spent longer looking at the image in the intralingual condition. Participants spent less time reading the subtitles in the reversed subtitling condition compared to the standard and intralingual subtitling conditions, but they did spend some time reading the subtitles even though they were not required to. It is probable that participants are attracted to the subtitle area primarily because of the saliency of the subtitles: their dynamic nature as well as the text they contain makes them very salient features. In addition, the high normalised number of fixations, even in the conditions with FL subtitles, indicate that reading behavior does occur. The possibility of using English spelling-to-sound rules as well as the automatic reading of text are most likely responsible for triggering the reading behavior in the FL subtitling conditions. Because subtitles were processed to some extent in each condition, it would suggest that language acquisition using FL films with subtitles is possible. However, it will be necessary to develop more sensitive measures of vocabulary acquisition to answer this question. As language learners can also use the soundtrack to increase their knowledge of a FL, future studies will need to investigate to what extent the soundtrack is also processed when subtitles are available. The current study showed that both the subtitles and the images in the FL films are processed, but further investigations that more explicitly investigate all three input modalities are needed in order to explore their effects on language acquisition. I will return to this in Chapter 6.

As no vocabulary acquisition was detected in the experiment presented in this chapter, the next chapter uses a simpler multi-modal situation (one spoken and written foreign language word form along with a line drawing depicting the meaning of the word) to investigate incidental vocabulary acquisition. In addition, Chapter 3 uses a potentially more sensitive measure of early vocabulary learning based on the savings paradigm.

Chapter 3: Incidental acquisition of foreign language vocabulary through brief multi-modal exposure[9]

3.1 Introduction

In the previous chapter, the processing of subtitles and the incidental acquisition of FL vocabulary were investigated using a film with subtitles. Overall, it was found that people read both NL and FL subtitles, and that more regular reading occurred when the soundtrack was presented in the FL (standard and intralingual subtitling). As participants read the FL subtitles to some extent in both the intralingual and reversed subtitling conditions, some of the FL was processed. Unfortunately, no FL vocabulary acquisition was detected.

One possible reason for the lack of incidental learning in the previous chapter is that a film with subtitles might be too noisy a situation for the incidental acquisition of FL vocabulary to occur. There is a lot of information to process when watching a film with subtitles, as viewers need to integrate information from the subtitles, the soundtrack and the film's images in order to understand the film. This may leave little cognitive resources available for the incidental acquisition of vocabulary to occur. In order to assess the impact of a multi-modal situation on the acquisition of FL vocabulary, the current study therefore used a simpler

[9] The data reported in this chapter has been published in:
Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLoS ONE*, 8(4), e60912.

multi-modal situation (a single auditory and written FL word along with a simple line drawing depicting the meaning of the word). Furthermore, all words in the current study were presented in the FL, which is similar to intralingual subtitling.

As was discussed in Chapter 2 (see Section 2.5), the lack of incidental learning effect in the previous experiment was also possibly due to the measuring instrument not being sensitive enough to detect early vocabulary knowledge. In Chapter 2, a simple recognition test was used to assess the incidental acquisition of FL words, however, the results showed no acquisition. It may be that some knowledge was gained during the exposure to the FL, but that this knowledge did not reach the threshold for recognition. Furthermore, the need for sensitive measures of vocabulary knowledge has also been highlighted in other studies of incidental FL vocabulary learning (Brown et al., 2008; Kweon & Kim, 2008; Pellicer-Sánchez & Schmitt, 2010). One of the aims of the present chapter was therefore to assess the usefulness of a more sensitive methodology based on the savings paradigm (Ebbinghaus, 1964; see Chapter 1, Section 1.5.2.2) to measure the acquisition of FL vocabulary. The savings paradigm has been shown to be more sensitive than typical recognition measures as it captures knowledge below the recognition threshold (Groninger & Groninger, 1980; MacLeod, 1988; Nelson, 1978, 1985). Studies in the field of language attrition have used this paradigm successfully to detect traces of forgotten language (see Bowers et al., 2009; De Bot et al., 2004; De Bot & Stoessel, 2000; Hansen et al., 2002). Typically, participants in those studies relearn words they were exposed to as children and for which they think they have no remaining knowledge. The results however showed that they are able to relearn these words faster than completely new words, thus indicating that the savings paradigm detects residual knowledge.

The idea of the savings paradigm was applied slightly differently in the experiment presented in this chapter. It involved two phases, an incidental learning phase, followed by an explicit learning phase (Figure 3.1, A). Phase 1 of this experiment, the incidental learning phase, made use of multi-modal FL stimuli by presenting auditory and written FL words with a picture illustrating the meaning. Participants engaged in a letter-search task in order to provide an incidental learning situation. Importantly, participants did not know the FL and were unaware that their acquisition of FL vocabulary would be assessed later on. In order to complete the task, participants only needed to attend to the written word form: the auditory word form and the picture were irrelevant for the task. However, the meaning of the FL word could be inferred from the picture. In Phase 2, the explicit learning task, participants were asked to learn the meaning of FL words through a translation recognition task. Auditory FL word forms from Phase 1 (old words) as well as new auditory FL words not previously encountered were presented simultaneously with an English word that was either the correct or incorrect translation. As explained in Chapter 1 (Section 1.5.2.2), it was expected that in the incidental learning phase, participants would start building some knowledge about the old words, and that this would help them reach the translation recognition threshold faster for these words then for completely new words during the subsequent explicit learning phase (see Figure 3.1, B).

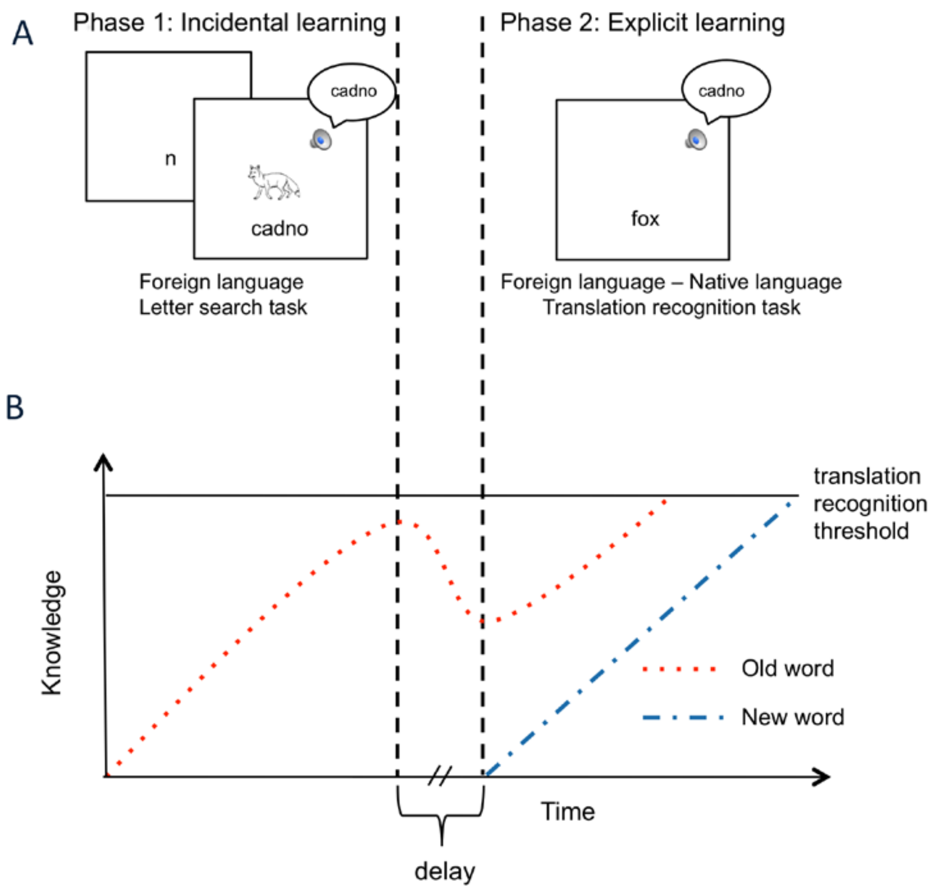

Figure 3.1: Experimental paradigm (panel A) and predicted effect on foreign language (FL) word knowledge (panel B).

To ensure that differences in performance for the old and the new words in the explicit learning task could be attributed to acquisition rather than to attentional arousal, in the incidental learning phase, a different group of participants (mismatched group) saw picture stimuli that did not match the correct meaning of the words. If attentional arousal leads to an advantage for the old words, the results for this group should not differ from the group where the pictures matched the meaning of the words, as both groups were exposed to the same FL word forms.

Another group of participants (multi-session group) took part in Phase 2 of the experiment the next day rather than immediately after Phase 1 and they completed the translation recognition task once again one week later. This multi-session group was used to explore whether the incidentally acquired form-meaning links were transitory or became embedded in memory after a relatively long retention interval.

3.2 Method

3.2.1. Participants

Sixty-six participants took part in the experiment and received payment for their participation. Participants were all native English speakers with no prior knowledge of Welsh. They were split into four groups of participants. Two groups of 16 participants completed Phase 1 and 2 of the study in a single-session: matched picture group (mean age 21.6, 11 females) and mismatched picture group (mean age 21.0, 15 females). A multi-session group of 18 participants (mean age 18.9, 15 females) completed Phase 1 on the first day of the study, Phase 2 the next day, and returned one week later to complete Phase 2 once more.[10] A

[10] One participant from this group did not return one week later and was therefore removed from the analyses.

further 16 participants (mean age 25.0, 12 females) were included as a control group and only completed Phase 2 of the study.

3.2.2. Design

A mixed design was used for the single-session groups with condition as a between-subject factor (the "matched" picture group which was pre-exposed to the written, auditory and pictorial forms of half of the words, and the "mismatched" picture group which was also pre-exposed to half of the words, but for which the pictorial stimuli represented the wrong meanings) and the type of word (old words, from the incidental learning phase, and new words) as a within-subject factor. A repeated-measures design was used for the multi-session group with type of word (old words and new words) and explicit learning phase (Phase 2 occurring one day after Phase 1, and Phase 3 occurring one week after Phase 2) as within-subject factors.

3.2.3. Stimuli

Welsh was chosen as the FL because it uses the same script as English but is sufficiently different from English so that participants could not simply guess the meaning of the words based on phonological or orthographic similarity. The stimuli consisted of 80 Welsh words, both the written and auditory forms (see Appendix B) and 80 pictures corresponding to these words (Snodgrass & Vanderwart, 1980). The words were split into two sets, and these were matched for category (Snodgrass & Vanderwart, 1980), word frequency in English (based on CELEX and British National Corpus) and word length in Welsh. None of the words were Welsh-English cognates. In Phase 1 of the experiment (incidental learning phase), participants were exposed to one set of words (counterbalanced across participants and groups) with their corresponding pictures, whilst in Phase 2 of the experiment (explicit learning phases), all words were used. For the mismatched picture group, the words were presented with a randomly assigned picture in the incidental learning phase (e.g. a picture of a dog presented with the auditory and written Welsh word “bwrdd” meaning “table”) and presented with the same picture for all the trials in the incidental learning phase. The words used in Phase 1 were labeled “old words” and the words participants were exposed to for the first time in Phase 2 were labeled “new words”. For the control group, one set of words was also classified as “old words” and the other as “new words” (counterbalanced across participants) to perform the analysis despite all of the words being presented for the first time for this group in Phase 2 of the experiment.

3.3 Procedure

3.3.1. Phase 1: Incidental learning

In Phase 1 (incidental learning), participants were asked to perform a letter-search task. In each trial, they were presented first with a letter and then a written Welsh word (see Figure 3.1A). Participants were told that the words would be in a FL, but they were not informed that the FL was Welsh. Their task was to indicate with a button press whether or not the word contained the letter. Each word was presented 4 times with a letter that was included in it and 4 times with a letter that was not (320 trials in total). Although irrelevant to the task, the corresponding auditory Welsh words and pictures were presented simultaneously with each written Welsh word. Since each trial was terminated as soon as participants pressed a button to complete the letter-search task, the auditory Welsh words may not have finished playing prior to the end of each trial.

3.3.2. Phase 2: Explicit learning

In Phase 2 (explicit learning), participants were presented with each auditory Welsh word and were asked to indicate with a button press whether the written English word presented simultaneously on the screen was the correct translation or not (see Figure 3.1 A). Each Welsh word was presented once with the correct translation and once with a foil in each block. The foils were chosen randomly from amongst the correct English translations and were different for each block. After each trial, participants received feedback on the screen ("correct" or "incorrect") and they were instructed to use this feedback to learn the correct translations. At the end of each block (160 trials), the percentage of correct answers was calculated and displayed on the screen, and the task continued until a criterion of 80% correct answers in one block was met or after a maximum of 4 blocks (this was reduced to a maximum of 3 blocks for the multi-session group). For this part of the experiment, participants were informed that they would be asked to learn some Welsh words, however they were not told that some of the words had already been presented in Phase 1.

3.4 Results

The results of the single and multi-session groups were analysed separately. For the single-session groups, the matched and mismatched picture groups were compared first to investigate a potential effect due to attentional arousal. Following this, the analyses compared the control group to the matched picture group to investigate a learning effect due to the incidental learning phase. Finally, the control group and the mismatched picture group were compared to investigate a potential effect due to exposure to the material (although the wrong meanings) during the incidental learning phase.

3.4.1. Single-session groups

3.4.1.1 Matched vs. mismatched picture groups

Accuracy. The overall error rate in the letter-search task of incidental learning phase (Phase 1) was low (5.8%).

The accuracy data of Phase 2 was analyzed using a mixed-design ANOVA with condition as a between-subject factor (matched and mismatched picture groups) and word type (new and old words) as a within-subject factor. The results showed significant main effects of word type, $F_1(1, 30) = 6.19$, $p < .05$, $\eta_p^2 = .17$, $F_2(1, 79) = 9.58$, $p < .01$, $\eta_p^2 = .11$ and condition, $F_1(1, 30) = 9.21$, $p < .01$, $\eta_p^2 = .24$, $F_2(1, 79) = 24.82$, $p < .001$, $\eta_p^2 = .24$ as well as a significant interaction between word type and condition, $F_1(1, 30) = 21.91$, $p < .001$, $\eta_p^2 = .42$, $F_2(1, 79) = 44.74$, $p < .001$,

$\eta_p^2 = .36$. This interaction occurred because the percentage of correct responses was significantly higher for old words ($M = 68.3$, $SE = 3.2$) than for new words ($M= 56.0$, $SE = 1.6$) in the matched picture group, $F_1(1, 30) = 25.70$, $p < .001$, $\eta_p^2 = .46$, $F_2(1, 79) = 36.17$, $p < .001$, $\eta_p^2 = .31$, whereas for the mismatched picture group the percentage of correct answers was only numerically lower for old words ($M = 53.2$, $SE = 1.5$) than for new words ($M = 57.0$, $SE = 1.3$), $F_1(1, 30) = 2.40$, $p = .13$, $\eta_p^2 = .04$, $F_2(1, 30) = 5.64$, $p < .05$, $\eta_p^2 = .07$ (see Figure 3.2).

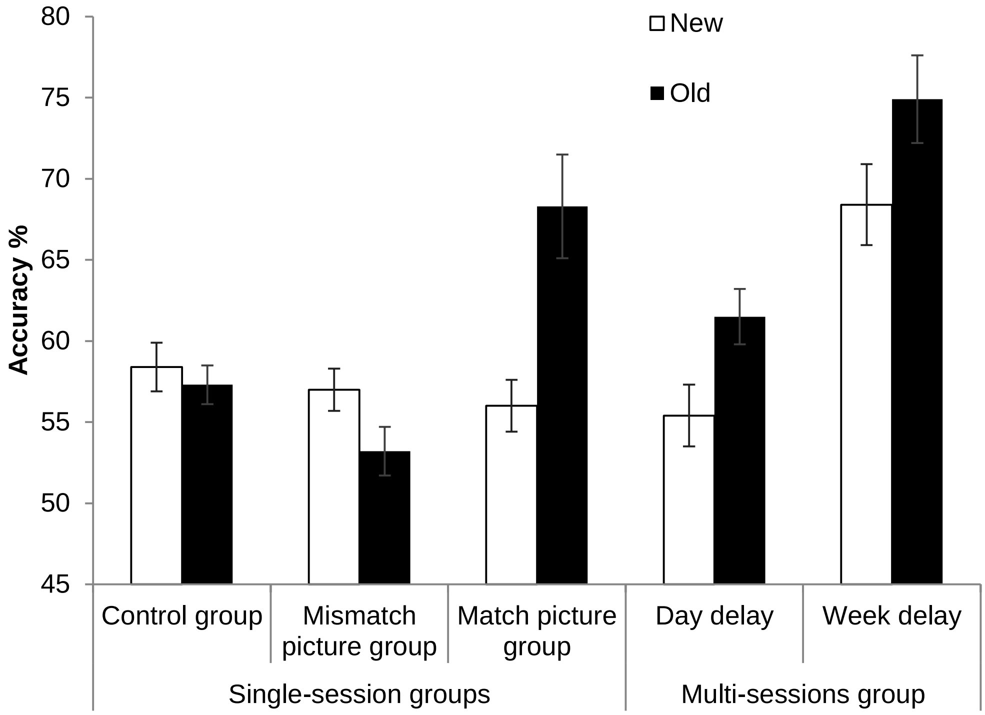

Figure 3.2: Percentage accuracy for old and new words in the translation recognition task (Phase 2: Explicit learning) in the first block of trials for each group with error bars.

Response Times (Phase 2). A mixed-design ANOVA with condition as a between-subject factor (matched and mismatched picture groups) and word type (new and old words) as a within-subject factor revealed a main effect of condition, $F_1(1, 30) = 4.48$, $p < .05$, $\eta_p^2 = .13$, $F_2(1, 79) = 122.59$, $p < .001$, $\eta_p^2 = .61$, but no main effect of word type, $F_1(1, 30) = 2.58$, $p = .12$, $F_2(1, 79) = 2.85$, $p = .10$. However, there was a trend towards an interaction between word type and condition, $F_1(1, 30) = 3.66$, $p = .07$, $\eta_p^2 = .11$, $F_2(1, 79) = 6.60$, $p < .05$, $\eta_p^2 = .08$, because the mismatched picture group were significantly slower at responding to the old words (*M*=1502 ms, *SE* = 58 ms) than the new words (*M* = 1437 ms, *SE* = 47 ms), $F_1(1, 30) = 6.19$, $p < .05$, $\eta_p^2 = .17$, $F_2(1, 79) = 9.29$, $p < .01$, $\eta_p^2 = .11$, whereas in the matched picture group there was no significant difference between the responses to old words (*M* = 1323 ms, *SE* = 44 ms) and new words (*M* = 1328 ms, *SE* = 49 ms), *F*s < 1 (see Figure 3.3).

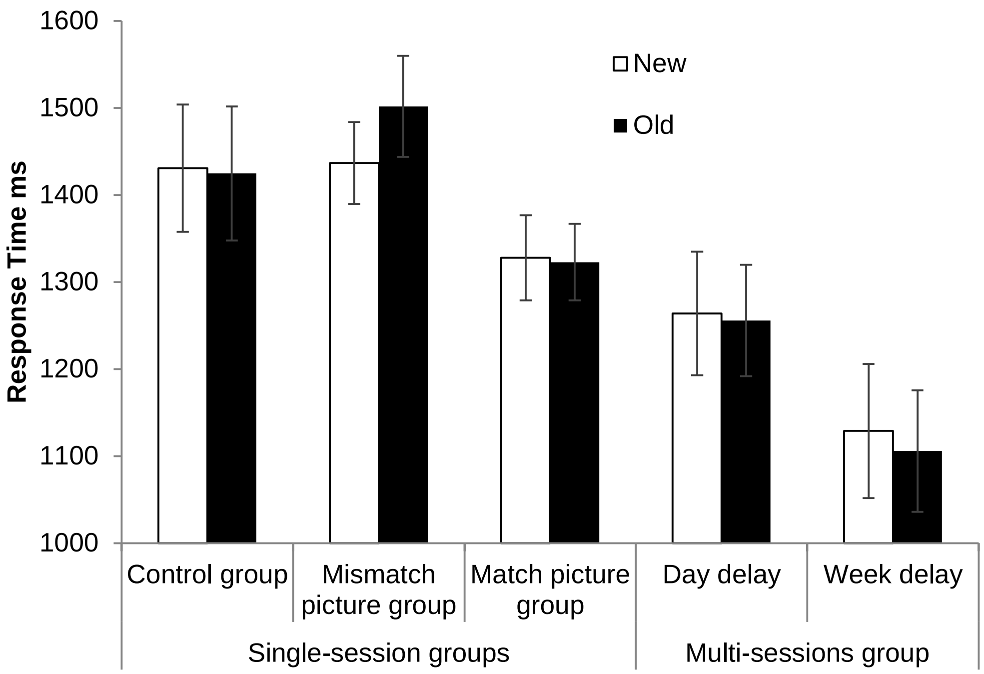

Figure 3.3: Response times (ms) for old and new words in the translation recognition task (Phase 2: Explicit learning) in the first block of trials for each group with error bars.

3.4.1.2 Control vs. Matched Picture groups

A mixed-design ANOVA with condition as a between subject factor (matched picture group and control group) and word type as a within-subject factor (old and new words) revealed a significant main effect of word type, $F_1(1, 30) = 11.53$, $p < .01$, $\eta_p^2 = .28$, $F_2(1, 79) = 17.36$, $p < .001$, $\eta_p^2 = .18$. The main effect of condition was only marginally significant in the subject analysis, however, it was significant in the item analysis, $F_1(1, 30) = 3.48$, $p = .07$, $\eta_p^2 = .10$, $F_2(1, 79) = 8.34$, $p < .01$, $\eta_p^2 = .10$. Crucially the interaction between word type and condition was significant, $F_1(1, 30) = 16.49$, $p < .001$, $\eta_p^2 = .36$, $F_2(1, 79) = 26.12$, $p < .001$, $\eta_p^2 = .25$, indicating that the matched picture group performed better on the old words than on the new words, however this was not the case in the control group (M = 57.3, SE = 1.2 vs. M = 58.4, SE = 1.5), Fs < 1 (see Figure 3.2). Furthermore, response times were not significantly different between the old and new words (M = 1425 ms, SE = 77 ms vs. M = 1431 ms, SE = 73 ms) in the control group, Fs < 1 (see Figure 3.3).

3.4.1.3 Control vs. Mismatched Picture groups

Accuracy. A mixed-design ANOVA with condition as a between subject factor (control group and mismatched picture group) and word type as a within-subject factor (old and new words) revealed a trend for a main effect of word type, $F_1(1, 30) = 3.80$, $p = .06$, $\eta_p^2 = .11$, $F_2(1, 79) = 3.87$, $p = .05$, $\eta_p^2 = .05$, and a trend for a main effect of condition, $F_1(1, 30) = 3.17$, $p = .09$, $\eta_p^2 = .10$, $F_2(1, 79) = 4.37$, $p < .05$, $\eta_p^2 = .05$. Importantly, the interaction between word type and condition was not significant, $F_1(1, 30) = 1.14$, $p = .29$, $\eta_p^2 = .04$, $F_2(1, 79) = 1.52$, $p = .22$, $\eta_p^2 = .02$ (see Figure 3.2).

Response Times. A mixed-design ANOVA with group as a between-subject factor (mismatched picture group and control group) and word type (new and old words) as a within-subject factor revealed neither a main effect of word type, $F_1(1, 30) = 4.14$, $p = .05$, $\eta_p^2 = .12$, $F_2(1, 79) = 2.61$, $p = .11$, $\eta_p^2 = .03$, nor a main effect of group, $F_1 < 1$, $F_2(1, 79) = 14.16$, $p < .001$, $\eta_p^2 = .15$. However, there was a strong trend for an interaction between group and word type, $F_1(1, 30) = 5.84$, $p < .05$, $\eta_p^2 = .16$, $F_2(1, 79) = 3.48$, $p = .07$, $\eta_p^2 = .04$, reflecting the slower responses to the old words in the mismatched picture group whereas response times were not significantly different between the old and new words ($M = 1425$ ms, $SE = 77$ ms vs. $M = 1431$ ms, $SE = 73$ ms) in the control group, Fs < 1 (see Figure 3.3).

3.4.2. Multi-session group

Accuracy. Error rates in the letter search task of Phase 1 were again low (6.4%).

The accuracy data of Phase 2 were submitted to a repeated-measures ANOVA with word type (new and old) and delay between phases (one day and one week) as within-subject factors. The results (see Figure 3.2) showed that accuracy was significantly higher for old words than for new words ($M = 68.2$, $SE = 1.9$ vs. $M = 61.9$, $SE = 1.9$), $F_1(1, 16) =$

13.76, $p < .01$, $\eta_p^2 = .46$, $F_2(1, 79) = 25.99$, $p < .001$, $\eta_p^2 = .25$. Furthermore, accuracy was overall higher one week later than the next day ($M = 71.7$, $SE = 2.4$ vs. $M = 58.5$, $SE = 1.5$), $F_1(1, 16) = 39.51$, $p < .001$, $\eta_p^2 = .71$, $F_2(1, 79) = 106.45$, $p < .001$, $\eta_p^2 = .57$.[11] Importantly, there was no interaction between word type and delay between phases, *F*s < 1, which indicates that participants scored significantly higher for the old words both the next day, $F_1(1, 16) = 9.16$, $p < .01$, $\eta_p^2 = .36$, $F_2(1, 79) = 10.48$, $p < .01$, $\eta_p^2 = .18$ and one week later, $F_1(1, 16) = 11.35$, $p < .01$, $\eta_p^2 = .42$, $F_2(1, 79) = 17.16$, $p < .001$, $\eta_p^2 = .18$.

Response Times. A repeated-measures ANOVA with type of word (old and new) and test time (next day and next week) showed that responses were faster one week later (M =1098 ms, SE = 9 ms) than the next day, (M = 1243 ms, SE = 10 ms), $F_1(1, 16) = 6.94$, $p < .05$, $\eta_p^2 = .30$, $F_2(1, 79) = 196.98$, $p < .001$, $\eta_p^2 = .71$.[3] However, there was no main effect of word type $F_1(1, 16) = 2.43$, $p = .14$, $F_2 < 1$, and no interaction between word type and test time, *F*s < 1 (see Figure 3.3).

3.5 Discussion

The results revealed incidental acquisition of FL vocabulary through a brief exposure to multi-modal stimuli. Being exposed to the written and auditory word forms of the FL words, as well as a picture

[11] Participants returned to complete the translation recognition task one week later, having already completed 2 or 3 (depending on when they reached the 80% criterion level) blocks of learning on this task the day after Phase 1. This explains the overall higher accuracy scores one week later relative to the first block after a day delay (Figure 2A).

illustrating the meaning of the word, resulted in incidental acquisition of FL vocabulary knowledge as shown by the higher scores for these words in the translation recognition task both immediately after the incidental learning task as well as the next day. In addition, the incidental learning effect remained one week later in the subsequent explicit learning task.

Participants in the mismatched picture group did not benefit from being exposed to the old words in the incidental learning phase, in fact, they suffered from being exposed to the wrong pictures as shown by significantly slower responses to the old words than the new words in the explicit learning phase. This disadvantage caused by the mismatched pictures in the incidental learning phase indicate that this group made form-meaning links that were incorrect. Thus, the higher scores for the words included in the incidental learning phase for the groups exposed to the correct pictures is due to the representation of form-meaning links rather than simple arousal.

An important question is what kind of learning best explains the results of both the matched and mismatched picture groups. Crucially, the observed acquisition of vocabulary reflects more than paired-associate learning between the auditory FL word form and the written NL word form, as this pairing was not presented in phase 1. Here, participants were exposed to the written FL word form (necessary to complete the letter-search task), the auditory FL word form and the meaning of the word via the picture. Written English translations were not presented in Phase 1. One explanation for the results is that participants linked the FL word forms with the semantic representation of the words activated by the pictures during Phase 1. Then, when the auditory FL word forms were presented in Phase 2, participants activated the meaning of the FL words (acquired via the pictures in Phase 1) and from there, they could accessed the written NL word form and reach a decision as to whether the translation was correct. Equally, translation recognition could have occurred if the written English word form activated its meaning which in turn was linked to the FL word form. Either way, participants relied on form-meaning links acquired during Phase 1 to complete the translation recognition task. This interpretation is compatible with Dobel et al. (2009) who also argued that form-meaning links were created during their statistical learning paradigm. In their study, participants were exposed to novel phonological word forms (pseudowords in the NL) in combination with pictures, with correct pairings occurring more frequently than incorrect ones. After completing 5 sessions over 5 consecutive days, participants achieved 90% accuracy in a translation test. The authors concluded that their results showed learning beyond mere stimulus-stimulus association, as the NL word forms used in the translation test were never presented during the statistical learning paradigm.

An alternative explanation for the observed incidental learning effect found here is based on the cascading activation model of speech production (see Kuipers & La Heij, 2009; Meyer & Damian 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005). This model predicts that even irrelevant pictures will automatically activate their conceptual representation, and that this in turn will cascade down to activate the lexical representations. Applying this model here would suggest that during the incidental learning phase, the presentation of the line drawings automatically activated the semantic representation for the concept and that this in turn activated the NL lexical representation of the word. As a consequence, it is possible that links were created between the latter and the FL lexical representations (phonological and/or orthographic). However, the task used in the current experiment did not involve naming, and it is less clear whether pictures that are irrelevant would activate lexical representations in a task that does not require explicit picture naming (but see Bles & Jansma, 2008). Crucially, even though the cascading model predicts the activation of the NL word form during the processing of the line drawing, the concept still needs to be activated first. Therefore, links could have been created between the FL word forms and BOTH the concept AND the NL word form, i.e., both form-meaning and form-form links. It is also important to remember that all this happened extremely rapidly and in parallel while participants were performing the letter-search task, which required attention to be focused on the FL written word form.

The data of the current experiment cannot rule out the second explanation for the locus of the incidental FL vocabulary learning. However, what is certain is that representations in the mental lexicon, either semantic and/or lexical, were automatically activated during the incidental learning phase, and that this in conjunction with the processing of the FL word forms was responsible for the learning. Furthermore, the current experiment was not able to reveal the neural mechanisms responsible for the creation and consolidation of form-meaning links, nor was it the aim of the experiment. However, these would likely involve working memory structures (for example the episodic buffer; Baddeley, 2000; 2002) with a rapid initial familiarization stage followed by a slower consolidation process as proposed by the complementary learning systems model of memory (Davis & Gaskell, 2009; see Chapter 1, Section 1.4.1).

Regardless of the precise locus of the form-meaning links, the acquisition of FL vocabulary occurred very rapidly in the incidental learning phase as FL words were only presented 8 times. This is much faster FL word learning with complete novices than found in previous studies. For example, McLaughlin et al. (2004) reported the first evidence of vocabulary learning (the learning of word forms) after 14 hours of exposure to French in a classroom setting. However, learners only became sensitive to the semantic properties of the FL words after approximately 60 hours of exposure. Another interesting study using informal exposure to a 7-minute Chinese weather report showed that some participants became sensitive to the spoken word forms included in the report as opposed to new FL word forms (Veroude et al., 2010). This study used a similar approach to ours, as the learning was incidental, and the words were presented between 2 to 8 times each in the weather report. However, the results revealed sensitivity to the word forms, which is an early stage of FL word learning, but not to the meaning of the words (also see Gulberg et al., 2012 using the same weather report, reviewed in Chapter 1, Section 1.3.4).

Another important aspect of the present data is the persistence of the incidental learning the next day as well as one week later, which highlights the long lasting impact of informal multi-modal FL exposure in vocabulary learning. Thus, this predicts that multi-modal FL exposure through activities such as games or watching FL films with subtitles could facilitate subsequent formal vocabulary learning even days later.

The new methodology to measure vocabulary acquisition in the current experiment was based on the savings paradigm. The results indicate that this paradigm is sensitive enough to detect differences in lexical knowledge between words presented in an incidental learning phase and completely new words. This provides an excellent method to measure vocabulary acquisition during incidental exposure and therefore the experiments in the following chapters will use the same paradigm.

3.6 Conclusion

Overall, the results of the current experiment revealed incidental vocabulary learning beyond the form level with complete beginners in the FL. Importantly, this learning persisted the next day as well as one week later. Learning and being able to use FL vocabulary fluently takes a long time, and the present findings show that incidental vocabulary acquisition through multi-modal exposure can play an important role in facilitating this process. The incidental acquisition of vocabulary happened very quickly in this study as participants were only exposed to each word 8 times. However, can incidental learning occur from even fewer exposures, and what is the role of the frequency of exposure to the words in the incidental learning phase? These questions will be addressed in the next chapter.

Chapter 4: The role of frequency of exposure in incidental multi-modal acquisition of foreign language vocabulary

4.1 Introduction

Many factors have been found to have an impact on the incidental acquisition of vocabulary, such as the type of word (Kweon & Kim, 2008), and the similarity of the FL with the NL (d'Ydewalle & van de Poel, 1999; Vidal, 2011). However, another potentially important factor is the frequency of exposure to the new words. Thus far, it remains largely unclear how many exposures are required before word knowledge is acquired. Crucially, while considerable research to date has focused on incidental learning from reading with advanced or intermediate learners of a FL (e.g. Hulstjin et al., 1996; Kweon & Kim, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Rott, 1999; Vidal, 2011; Waring & Takani, 2003), not much is known about the frequency of exposure necessary in other contexts such as multi-modal situations, or when the learners are complete beginners in a FL. In Chapter 3, it was found that within 8 exposures to new words in a multi-modal incidental learning situation, participants who had no prior knowledge acquired FL vocabulary. However, can vocabulary acquisition occur from even fewer exposures to the FL words? Furthermore, in this type of situation can we expect an increase in vocabulary acquisition commensurate with an increase in the frequency of exposures? The current chapter addresses these questions.

One of the first to study the effects of frequency of exposure on learning and memory was Ebbinghaus (1964). In his famous series of experiments, Ebbinghaus memorized series of nonsense syllables until he could recite them. Exposures to the series of syllables during the learning phase varied between 8 and 64, and learning was achieved through reading and reciting. After the learning phase, Ebbinghaus measured the savings in relearning as a function of the length of the series, the retention interval, and most relevant for this research, as a function of initial repetitions. Ebbinghaus found that increased repetitions during the learning phase led to a decrease in the time necessary to relearn the series of syllables the next day, which is essentially a frequency effect. Ebbinghaus's methods and ideas are still relevant today, and many recent studies in the field of incidental vocabulary acquisition have investigated the effects of repeated exposures on the learning and retention of vocabulary. These studies were reviewed in detail in Section 1.3, and Table 4.1 provides a summary of the findings in terms of frequency of exposure effect as well as the minimum number of exposures required before incidental learning was detected in the context of reading, listening and reading-while-listening.

Table 4.1: Summary of findings on frequency effects.

Type of study	Authors	Frequency bands	Frequency effect?		Min. num. exposures required	Additional info.
Reading	Brown et al., 2008	2-3, 7-9, 10-13, 15-20	✓	2-3 < 7-9 < 10-13 < 15-20	No stats.	-Multiple comp., unclear if Bonferroni corrected, no control group/items
	Hulstjin et al., 1996	1 or 3	✓	1 < 3	No stats.	-No control group/items +self-reported prior knowledge checked post treatment
	Kweon & Kim, 2008	1-6, 7-19, 20 or more	?	Frequency effect not tested statistically	1-6	-Self-reported knowledge scale with pre-test / post test comparison
	Pellicer-Sánchez & Schmitt, 2010	1, 2-4, 5-8, 10-17, 28 or more	✓	8 or less < 10 or more	No stats.	-Multiple comp., unclear if Bonferroni corrected, no control group/items
	Rott, 1999	2, 4, 6	✓	2 = 4 < 6	2	+Comp. with control items
	Vidal, 2011	1-6	✓	47% of variance explained by freq. predictor	No stats.	-Pre-test / post test comp.
Listening	Vidal, 2011	1-6	✓	24% of variance explained by freq. predictor	No stats.	-Pre-test / post test comp.
	Brown et al., 2008	2-3, 7-9, 10-13, 15-20	**x**		No stats.	-Multiple comp., unclear if Bonferroni corrected, no control group/items
Reading-While-Listening	Brown et al., 2008	2-3, 7-9, 10-13, 15-20	✓	2-3 < 7-9 < 10-13 = 15-20	No stats.	-Multiple comp., unclear if Bonferroni corrected, no control group/items
	Horst et al., 1998	2-17	?	.49 correlation between freq. and knowledge gain, unclear if statistically significant.	No stats.	-Pre-test / post test comp.
	Webb et al., 2013	1, 5, 10, 15	✓	1 = 5 = 10 < 15	5	+Comparison with control group

Note. Min. num. exposures required = minimum number of exposures require before incidental learning occurred; No stats. = not tested statistically; freq. = frequency; comp. = comparisons.

The studies summarized in Table 4.1 all took place with either intermediate or advanced FL learners. Furthermore, these studies used reading, listening and reading-while-listening as an incidental learning situation, where learners had to derive the meaning of novel words from the language context. With complete beginners in a FL, it would be impossible to study incidental acquisition in such situations, as learners do not possess enough knowledge to be able to infer the meaning of the words from the provided context. Nation (2001) and Schmitt (2010) both estimated that at least 95% of the words in a text must be known in order for learners to do this. However, a multi-modal situation with pictorial information could be used with complete beginners as in this case, learners have access to the meaning of the words through the pictorial information. Since this field of research is relatively new, little is known about the number of exposures required before beginner learners can extract knowledge about the form and meaning of novel words. However, in Chapter 3, it was found that participants were able to learn the meaning of FL words in a multi-modal incidental learning situation within 8 exposures to an auditory and written FL word form presented along with a picture, even though they had no prior knowledge of the FL. An incidental learning effect was found both immediately after the exposure to the FL words, as well as the following day. Furthermore, following further explicit learning of the FL words one week later, an incidental learning advantage was still found for the words presented in the multi-modal incidental learning situation.

Another recent study presented a 7-minute weather report in Chinese to participants without prior knowledge of this FL (Gullberg et al., 2012). After watching the weather report either once or twice, participants completed an auditory word recognition test asking them to indicate whether or not Chinese words had occurred in the weather report. The frequency of occurrence was found to be a significant predictor of recognition scores, with words occurring more frequently in the report (8-16 exposures) recognized with significantly more accuracy than words occurring infrequently (2-4 exposures). Unfortunately, it is not clear whether the accuracy scores for each of the frequencies were significantly above chance. Furthermore, there were no control items. Therefore it is difficult to determine whether the higher accuracy scores for the more frequent items were due to the frequency of occurrence of the items, rather than to the items themselves being easier to recognize.

Overall the results of previous studies suggest that there are frequency of exposure effects in the incidental learning of vocabulary. However, how many exposures are required to acquire new words for different proficiency learners and how this might be modulated in different learning situations remain open questions. An important factor to consider when measuring word acquisition is the method used to assess the vocabulary learning. In the FL literature, most studies have used traditional recognition and recall tests. However, such tests might not tap into the very earliest stages of word learning. Very early effects have been detected using EEG, albeit with native speakers encountering pseudowords (Batterink & Neville, 2011; Borovsky et al., 2010; Mestres-Missé et al., 2007). In the previous chapter (Chapter 3), a variation of the savings paradigm was used to detect traces of knowledge that had not necessarily reached the threshold for explicit recognition with beginner FL learners (see Chapter 1, Section 1.5.2.2 and Chapter 3, Section 3.1 for further details about this method). Participants were exposed to novel words in an incidental learning phase before being asked to learn words explicitly. As half of the words in the explicit learning phase had been presented in the incidental learning phase and half of the words were completely new, it was possible to determine whether simple exposure (incidental learning) impacted later overt/explicit vocabulary learning. Crucially, participants performed reliably better in the explicit learning phase for FL words they had been exposed to during the incidental learning phase compared to new FL words. The same methodology for detecting early word knowledge was used in the study presented in the current chapter.

There are some methodological concerns with previous research, as mentioned in Chapter 1 (Section 1.3), that makes it problematic to draw strong conclusions about incidental vocabulary learning (also see Table 4.1). For example, many earlier studies have not made use of a control group and/or a set of control items in order to confirm that the learning and frequency effects are due to the their experimental manipulation (e.g., Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010). In some studies, learning is measured by comparing pre- and post-tests (e.g. Rott, 1999;

Vidal, 2011), and this can be problematic as it increases participant's awareness of the word learning aspect of the experiment and draws participant's attention to specific words. Some studies have used different numbers of words in their frequency bands (e.g. between 2 and 10 words per frequency band in Pellicer-Sánchez and Schmitt, 2010 and between 3 and 10 words per frequency band in Vidal, 2011) whilst others used only a few items overall (e.g. 6 target words in Rott, 1999).

Most importantly, there is a lack of research on the incidental learning of vocabulary both in multi-modal situations, and with complete beginners in a FL. The purpose of the current study was to investigate the influence of frequency of exposure on vocabulary learning with beginners, using a methodology sensitive to early vocabulary gains as in Chapter 3. Participants were exposed to the FL (Welsh) words in a multi-modal incidental learning phase. Subsequently in an explicit learning phase, they were asked to learn FL words. Unbeknownst to participants, half of the words had been presented during the incidental learning phase, and half of the words were completely new. The frequency of exposure to the FL words in the incidental learning phase varied within-subjects in four frequency bands: 2, 4, 6 and 8 exposures, with 10 words in each frequency band and 2 different sets of items. One set was used in the incidental learning phase (counterbalanced across participants) and the other set was used as a control comparison in the explicit learning phase. Based on the findings of Chapter 3, it was predicted that participants would perform better on the words they had been exposed to during the incidental learning phase compared to the new words. Furthermore based on prior research on frequency of exposure effects, it was expected that performance in the explicit learning phase would increase with the frequency of exposure in the incidental learning phase.

4.2 Method

4.2.1. Participants

Seventy-eight participants took part in this experiment (mean age 20.7, 60 females) and received payment or course credit for their participation.[12] Participants were all students at the University of Nottingham, and they all completed a self-reporting language background questionnaire to ensure that they were native English speakers and that they had no prior knowledge of the FL (Welsh) used in the study. Following this, 10 participants were excluded from the analyses for the following reasons: 3 participants were bilingual, 1 participant was not a native speaker of English, 4 participants reported having prior knowledge of Welsh (although very minimal), 1 participant reported living close to Wales and visiting the region frequently and 1 participant did not answer the question related to prior knowledge of Welsh and was therefore excluded as a precaution. Finally, one further participant was excluded because of a technical problem during Phase 2 of the experiment.

[12] Three participants did not provide age information and two participants did not indicate their gender.

4.2.2. Design

A repeated-measures design was used for this experiment with type of word (old and new) and frequency of exposure (4 frequency bands: 2, 4, 6 and 8 exposures) as within-subject factors.

4.2.3. Stimuli

The stimuli used for this experiment included the auditory and written word form of 80 Welsh words (same as Chapter 3; see Appendix B), as well as line drawings corresponding to the meaning of the words (Snodgrass & Vanderwart, 1980). The words were split into two sets, with one set of words used in Phase 1 of the experiment (counterbalanced across participants) and both sets of words used in Phase 2 of the experiment. The words presented during Phase 1, the incidental learning phase, were considered “old words” and the words seen for the first time during Phase 2, the explicit learning phase, were considered “new words”. Within each set of words, the items were split into 4 frequency bands such that words were presented either twice, 4 times, 6 times or 8 times during the incidental learning phase. In order to control for the fact that some words might be easier to learn than others, the words in each set were ranked according to the average percentage of accuracy score achieved across all participants in the control group of Chapter 3. This was used to match word difficulty across the frequency bands.

4.3 Procedure

4.3.1. Phase 1: Incidental learning

The procedure for Phases 1 and 2 of the experiment was very similar to the one used in Chapter 3. In Phase 1 of the experiment, participants were exposed to the words in an incidental learning situation in which they were asked to complete a letter-search task. Participants were first presented briefly with a letter (500 ms) prior to the appearance of a written word. Their task was to indicate with a button-press whether the letter they saw was present in the written word that would appear on the screen. For half of the trials, the letter appeared in the word ("Yes" responses) and for half of the words it did not ("No" responses). Although irrelevant for the letter-search task, participants heard the auditory form of the Welsh words and saw a line drawing depicting the meaning of the words simultaneously with the presentation of the written word forms. Participants were not told that the FL was Welsh and they were not asked to learn the words: they were simply instructed to complete the letter-search task. Participants completed 8 practice trials with feedback prior to the main letter-search task where they completed 200 trials.

4.3.2. Phase 2: Explicit learning

In Phase 2 of the experiment, participants completed a translation recognition task, in which they were explicitly asked to learn the meaning of Welsh words. Both sets of words (old and new) were used in Phase 2 of the experiment, however, participants were not told that they had previously been exposed to half of the words. Participants were presented with the auditory word form of Welsh words while concurrently viewing a possible English translation. Their task was to indicate with a button-press whether the English word was the correct translation for the auditory Welsh word. Participants received feedback after each answer ("correct" or "incorrect") and they were told to use this feedback to learn the correct translations. Each auditory Welsh word was presented once with its correct English translation and once with a foil. The foils were English translations pseudorandomly assigned to an auditory Welsh word and were different for each block of trials. At the end of each block, participants viewed their percentage accuracy for this block prior to continuing to the next block. Participants were informed that the experiment would continue for a maximum of 3 blocks or until they achieved 80% accuracy in one block.

4.4 Results

Accuracy was high in the letter-search task (M = 95%, SE = 0.8%) indicating that participants were attending to the stimuli. However, one participant only achieved 58% accuracy and was removed from further analyses.

Only the results of blocks 1 and 2 of the translation recognition task are reported. The results of block 3 were not analyzed as some participants reached criterion in block 2 and therefore did not proceed to block 3. For each participant, the percentage of accuracy was calculated for the old and new words for each frequency band during the explicit learning phase (see Table 4.2). These were submitted to repeated-measures ANOVAs with word type (old vs. new) and frequency of exposure (2, 4, 6 and 8) as within-subject factors for each block separately. Results revealed a main effect of word type for both block 1, $F_1(1, 65) = 41.35$, $p < .001$, $\eta_p^2 = .39$, $F_2(1, 76) = 51.43$, p , .001, $\eta_p^2 = .40$, and block 2, $F_1(1, 65) = 23.09$, $p < .001$, $\eta_p^2 = .26$, $F_2(1, 76) = 28.94$, $p < .001$, $\eta_p^2 = .28$. Furthermore, for block 1, simple effects analyses revealed that accuracy scores were significantly higher for the old words in each frequency band, $F_1(1, 65) = 9.05$, $p < .01$, $\eta_p^2 = .12$, $F_2(1, 76) = 8.10$, $p <.01$, $\eta_p^2 = .09$, for frequency band 2, $F_1(1, 65) = 15.97$, $p < .001$, $\eta_p^2 = .20$, $F_2(1, 76) = 12.28$, $p <.01$, $\eta_p^2 = .15$, for frequency band 4, $F_1(1, 65) = 10.99$, $p < .01$, $\eta_p^2 = .15$, $F_2(1, 76) = 9.62$, $p <.01$, $\eta_p^2 = .12$, for frequency band 6, and $F_1(1, 65) = 27.81$, $p < .001$, $\eta_p^2 = .30$, $F_2(1, 76) = 23.92$, $p <.001$, $\eta_p^2 = .30$ for frequency band 8. Furthermore, the advantage gained from exposure to the words in the incidental learning phase was still significant in block 2 following further explicit learning for the words with 2 exposures in the incidental learning phase, $F_1(1, 65) = 5.13$, $p < .05$, $\eta p^2 = .07$, $F_2(1, 76) = 7.42$, $p <.01$, $\eta p^2 = .10$, 6 exposures, $F_1(1, 65) = 9.70$, $p < .01$, $\eta p^2 = .13$, $F_2(1, 76) = 8.18$, $p <.01$, $\eta p^2 = .10$, and 8 exposures, $F_1(1, 65) = 12.83$, $p < .01$, $\eta p^2 = .17$, $F_2(1, 76) = 13.52$, $p <.001$, $\eta p^2 = .18$. For the words with 4 exposures in the incidental learning phase, it was no longer the case, $F_1(1, 65) = 2.13$, $p = .15$, $\eta p^2 = .03$, $F_2(1, 76) = 2.24$, $p = .14$, $\eta p^2 = .05$.

Table 4.2: Mean (*SE*) percentage accuracy for each word type and frequency of exposure in Blocks 1 and 2.

	Block 1		Block 2	
	% accuracy		% accuracy	
	Old	New (control)	Old	New (control)
Freq. 2	61.2 (1.7)	55.3 (1.5)	76.9 (1.7)	72.3 (1.7)
Freq. 4	64.9 (1.6)	57.7 (1.3)	74.1 (1.7)	71.6 (1.6)
Freq. 6	66.4 (1.6)	59.9 (1.5)	77.0 (1.4)	72.2 (1.4)
Freq. 8	68.9 (1.3)	58.8 (1.4)	75.9 (1.5)	69.8 (1.2)

Note. Freq. = frequency band

In order to investigate the frequency of exposure effect, a difference score between old and new words was calculated, referred hereafter as the "incidental learning effect" (see Figure 4.1). The frequency of exposure effect was investigated by computing linear contrasts using a repeated-measures ANOVA in the participant analysis and a one-way ANOVA in the item analysis. Results revealed no significant linear contrasts neither for block 1, $F_1(1, 65) = 2.31$, $p = .13$, $\eta p^2 = .03$, $F_2(3, 76) = 1.64$, $p = .20$, $\eta p^2 = .02$, nor block 2, Fs < 1. To further investigate whether an increase in the frequency of exposure led to an increase in incidental learning effect, the two extreme frequency bands were compared, i.e., frequency of exposures 2 and 8 using a paired sample t-test and an independent sample t-test in the participant and item analyses respectively (one-tailed). For block 1, results revealed a significant increase in the incidental learning effect from 2 to 8 exposures in the participant analysis, $t_1(65) = 1.71$, $p <.05$, $d = 0.21$, and a strong trend in the item analysis, $t_2(38) = 1.55$, $p =.06$, $d = 0.49$. However no significant differences were found in block 2, *t*s < 1.

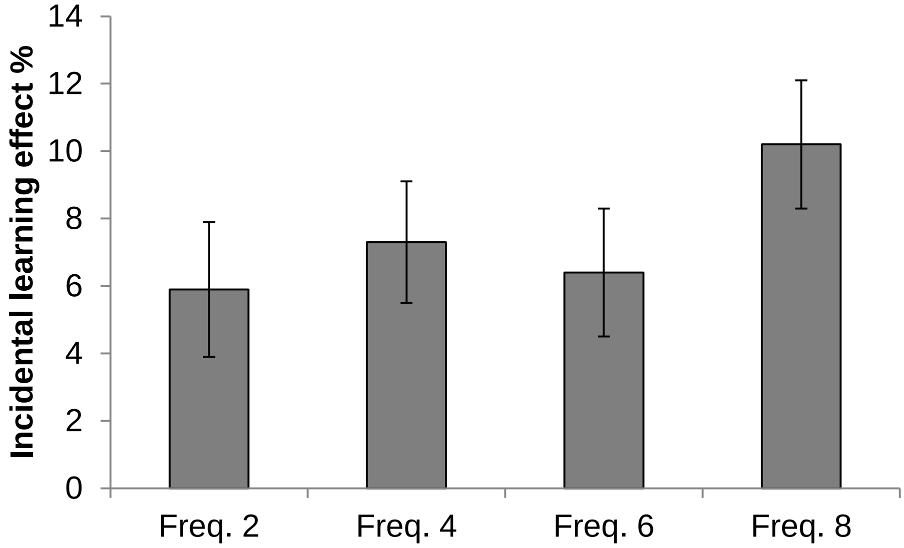

Figure 4.1: Incidental learning effect (difference score between old and new words) for each frequency band in block 1 of the explicit learning task.

It is also possible that the knowledge gained for each exposure to the words is not constant across frequency bands. For example, for the words with 2 exposures in the incidental learning phase, accuracy scores were about 6 % higher for old words compared to new words, which is equivalent to a 3% increase in accuracy per exposure to the words. In contrast, for the words with 8 exposures in the incidental learning phase, the accuracy scores were approximately 10% higher for the old words, which is equivalent to a 1.25% increase per exposure. Therefore the incidental learning effect was normalized by dividing the incidental learning effect by the number of exposures in each frequency band (see Figure 4.2). The normalized incidental learning effects were then investigated using ANOVAs which revealed a small but significant negative linear trend F_1 (1, 65) = 3.98, p = .05, ηp^2 = .06, F_2 (1, 76) = 4.62, p < .05, ηp^2 = .05, indicating that the knowledge gained from each exposure decreased as the number of exposure increased.[13]

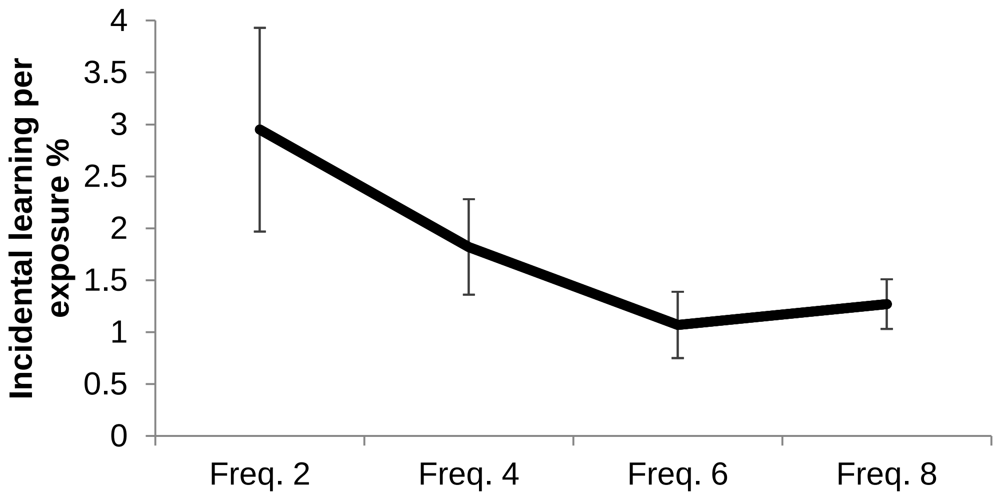

Figure 4.2: Normalized incidental learning effect in each frequency band (incidental learning effect divided by the number of exposures in each frequency band).

4.5 Discussion

This chapter investigated the effect of frequency of exposure on the incidental acquisition of FL vocabulary through a multi-modal situation. The results replicated the findings of Chapter 3, demonstrating an incidental learning effect on the performance in the translation recognition task with participants without prior knowledge of the FL. Furthermore, the incidental learning effect persisted following further

[13] As this data was not normally distributed, the Pages for F_1 and Jonckheere for F_2 analyses of trend tests were also computed. The results revealed significant linear trends in both cases, L = 1692, z = 2.03, p < .05, r = .25, J = 994, z = 1.78, p < .05, r = .20.

explicit learning in block 1 into block 2. Crucially, the current results showed that even when participants are only exposed to words twice in the incidental learning phase their performance in the explicit learning phase improved compared to new words. This indicates that some knowledge about the lexical items was acquired after as little as 2 exposures in an incidental multi-modal situation. Our early learning effect is consistent with studies showing that 2 exposures to novel words while reading led to a learning effect (Rott, 1999), as well as to previous EEG studies which tapped into early word learning (see Batterink & Neville, 2011; Borovsky et al., 2010; Mestres-Missé et al., 2007).

Importantly, the results suggests that more exposures in the incidental learning phase led to better performance on the translation recognition task. However, this was only the case when comparing the scores from the two extreme frequencies of occurrences (2 versus 8 exposures). Although an overall frequency of exposure effect was not found, further analyses showed that this was likely due to the fact that the impact of each encounter with a word in the incidental learning phase was not constant across frequency of exposures. Crucially, as the number of exposures increased, additional encounters with a word had less of an impact on incidental learning.

It is important to consider that in the current study, as well as in the experiment reported in Chapter 3, the meaning of the words could easily be ascertained from the pictorial stimuli: the pictures used were simple static line drawings, and the FL words were presented as isolated words. Further, the words were always presented with the same line drawing, making it likely that meaning was extracted from these representations early on. This could explain why further encounters had less of an impact. It is also likely that form-meaning links are easier to establish when both form and meanings are presented simultaneously or contiguously. Thus, there may be an advantage to being exposed to FL words through a multi-modal situation with pictorial information because in this type of situation, word meaning can be presented clearly and at the same time as the word forms. In a more complex multi-modal situation, for example a film with subtitles, where the meaning of the words is less transparent and where the words are presented in sentences rather than isolated words, it is likely that more repetitions will be necessary for knowledge to be acquired. Therefore, although learning effects were found from 2 exposures, this may not be enough in other more complex contexts or when the meaning of the novel words is less clear.

Another factor that may have influenced the unequal impact of additional exposures to novel FL in the current study is the incidental learning phase itself. Here participants had to complete a letter-search task, and this did not require them to pay attention to the FL auditory word forms or to the pictures; only the FL written word forms were important. Thus initially, participants may have been interested in all the stimuli because of their novelty and saliency. However after the first few exposures, they were no longer interested in attending to this ‘extra’ information. It would therefore be important in future research to use eye-tracking to investigate the allocation of attention to the pictorial and written stimuli, and more specifically, to explore the viewing behavior across repetitions of the stimuli to determine how this impacts learning.

An overall effect of frequency of exposure has been found in other contexts such as reading (Brown et al., 2008; Hulstjin et al., 1996; Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Vidal, 2011), listening (Vidal, 2011) and reading-while-listening (Brown et al., 2008; Horst et al., 1998; Webb et al., 2013), however, the effect is difficult to compare across studies, as varied frequency bands were used. A normalized incidental learning effect for each frequency band, as was used in the current study, might therefore be useful to compare across studies, and to ascertain whether the unequal impact of additional exposure found in the current study is generalizable to other contexts. Furthermore, not many studies reported the number of exposures required in order to detect a significant incidental learning effect. To my knowledge, the two exceptions are Webb et al., (2013) who found significant incidental learning from 5 exposures compared to a control group that was not exposed to the words, and Rott (1999) who found an earlier effect, with 2 exposures leading to significant incidental learning compared to a control set of items without prior exposure. It would therefore be useful in the future that studies on frequency of exposure also include comparisons with a control group or a set of control items in order to explore early incidental learning effects.

4.6 Conclusion

The results of the current study indicate that the incidental acquisition of vocabulary can happen extremely fast even with complete beginners in a FL. As little as 2 exposures to new words in a multi-modal incidental learning situation was enough for knowledge about the new words to be acquired. In addition, the data showed that the impact of exposure was not constant across frequency bands, but rather decreased following the initial encounters. I will return to the unequal impact of additional exposures to FL words and pictures in Chapter 6 where eye-tracking will be used to explore the allocation of attention to the pictures and words across the duration of the incidental learning phase. Overall, the findings of the current study suggest that very few exposures to new words are required before learning starts to occur, and the variation on the savings paradigm used in the experiment was sensitive enough to detect learning from 2 exposures to FL words. However, this measure is not able to detect knowledge past the recognition of the initial form-meaning link. It would therefore be useful in the future to assess further learning using, for example, a recall measure. This will also be addressed in Chapter 6. Finally, in both the current experiment as well as in the previous chapter, rapid FL word learning occurred during an incidental learning phase where pictures were used to derive meaning. As was mentioned in Section 1.4 (see Chapter 1), the current models of word learning were not designed to account for this type of informal learning. Furthermore, they do not consider the type of links, lexical of lexical-semantic that may result from acquiring FL word meanings through exposure to pictorial stimuli. In the next chapter I will attempt to clarify the type of learning that results from incidental multi-modal FL word learning.

Chapter 5: Investigating lexical-semantic links in multi-modal incidental foreign language vocabulary learning

5.1 Introduction

The data reported in previous chapters showed that people learnt new FL words rapidly, from minimal exposures to FL words in an incidental learning situation. As the learning paradigm used in the previous chapters involved FL word forms and pictures, participants could have created form-meaning links between the auditory and written FL words forms and the semantic representations accessed from the pictures. However, as discussed in Section 3.5 (Chapter 3), it is also possible, that links were created between FL word forms and NL translation equivalents (lexical links). Even though NL translations were not presented during the incidental learning situation, they may have become activated automatically during the presentations of the pictures, as predicted by the cascading model of activation (see Kuipers & La Heij, 2009; Meyer & Damian 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005). As the previous experiments did not clarify which of these two possible explanations (form-meaning vs. form-form) best explained the learning, the aim of the present chapter was to shed light on the locus of the incidental learning effect.

In Chapter 1 (Section 1.4), two models of FL word learning were reviewed, i.e., the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994) and the developmental Bilingual Interactive Activation model (BIA-d; Grainger et al., 2010). In short, the RHM was initially designed to explain the links between native and FL lexicons,[14] as well as the links between each language's lexical and conceptual representations. The model includes stronger links between lexical representations than between FL word forms and semantic representations, as well as stronger links from FL words representations to NL words representations than vice-versa (see Figure 1.1 and Section 1.4.2 in Chapter 1 for an overview of this model). It was hypothesized that initially, learners relied on NL translations to access FL word meaning, and that with increased fluency, direct semantic links gradually became stronger.

Following on from the RHM, the BIA-d was designed specifically for FL learning which occurs in a classroom situation. It is assumed that in such a situation, learners are typically presented with FL word forms in combination with NL translations (through a vocabulary list, for example). Although more of a connectionist model, the BIA-d nevertheless makes similar predictions to the RHM, as it assumes that learners first establish lexical links between FL and NL word representations, and that direct semantic links develop later on, with increased fluency. In contrast to the RHM, the BIA-d includes detailed information about the development of the lexical and semantic links at the level of word form representations (see Figure 1.2). Furthermore, the BIA-d makes an important distinction between supervised and unsupervised learning, and highlights that the predicted stronger lexical links would develop during the initial

[14] Kroll and Stewart use L1 and L2 rather than native and foreign languages.

supervised learning situation (a classroom learning situation for example), and that direct semantic links would then be created between the FL word form representations and the semantic representations during an unsupervised learning period.

Importantly, neither the RHM nor the BIA-d makes predictions about the lexical and semantic links that would accrue from an *initial unsupervised* learning situation such as the incidental learning paradigm used in this thesis. Furthermore, the models do not make predictions for FL words learnt using pictorial information, as in for example, activities using flash-cards or involving multimedia, rather than through the presentation of translation equivalents. In the incidental learning situation used throughout this thesis, FL words are presented in combination with pictures, and this may encourage the formation of form-meaning links early on. The aim of the current experiment was therefore to investigate the lexical-semantic links occurring from learning vocabulary incidentally through a multi-modal learning situation involving pictorial information and FL words forms only, i.e., without explicitly providing NL translations.

In an initial phase, participants were exposed to FL words in combination with pictures in an incidental learning situation as in Chapters 3 and 4. Following this, in order to investigate whether the incidental learning situation led to the creation of links between the FL word forms and the semantic representations (form-meaning links), or to the creation of FL-NL lexical links (form-form links), it was important to find a task that tapped into the lexical-semantic network, without requiring explicit translations. This was achieved using a cross-modal semantic priming task, similar to the one used in Dobel et al. (2009), in which auditory FL word forms were used a primes, and participants had to complete a semantic categorization task based on pictures. In the current study however, auditory FL words from the incidental learning phase were used as primes for this task, and targets were either pictures (cross-modal semantic priming) or written NL translations (cross-modal translation priming). If form-meaning links were responsible for the incidental learning effect, it was expected that the categorization of picture targets following a FL prime matching the target in meaning would lead to a priming effect, however, the categorization of NL words would not. In contrast, if form-form links were responsible for the learning effect, it was expected that the categorization of NL words following the presentation of translation equivalent FL primes would led to a priming effect, but that the categorization of picture targets would not (see Figure 5.1). In order to make sure that the task was sensitive to priming, and to provide a baseline, participants also completed a priming task using English (NL) words as primes. Following the priming task, all participants completed the explicit learning task as in Chapters 3 and 4.

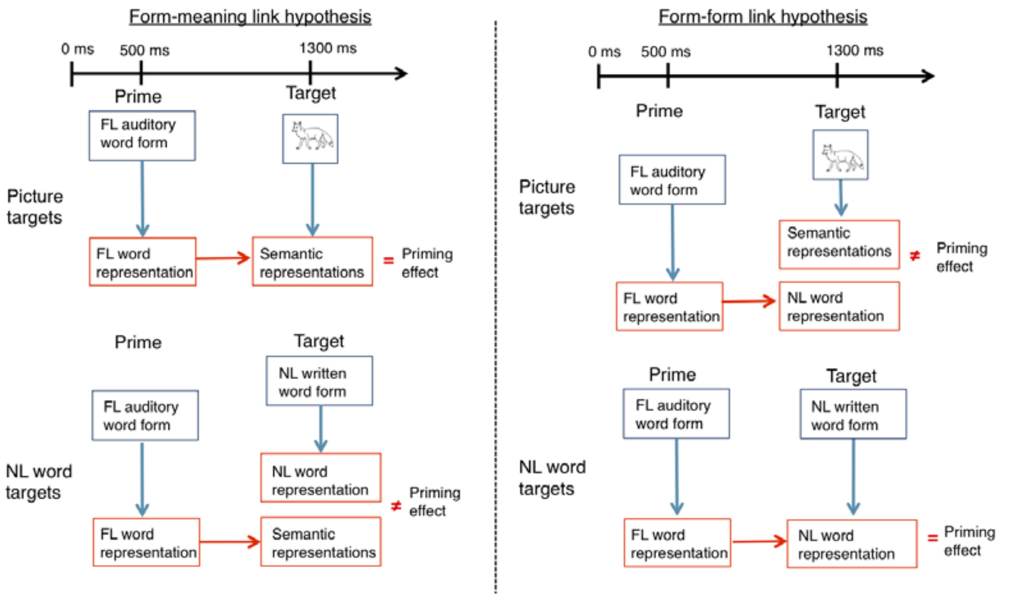

Figure 5.1: Schematic illustration of priming task showing predicted lexical-semantic activation and priming effect. Blue boxes indicate physical stimuli with blue arrows indicating predicted activation. Red boxes and arrows show predicted activation of mental representations.

Because of the important changes in the storage of new lexical information predicted by the Complementary Learning Systems (CLS; Davis & Gaskell, 2009; Lindsay & Gaskell, 2010) account of word learning (see Chapter 1, Section 1.4.1 for an overview of this model), it was important to assess the lexical-semantic links both immediately following the learning phase, as well as following an overnight consolidation period. Therefore, two groups of participants were included in the current study. One group of participants completed the priming task immediately following the incidental learning task (same-day group). The second group of participants completed the priming task one day following the incidental learning task (next-day group). Following the logic of the CLS model, it was expected that a stronger priming effect would emerge following an overnight consolidation period.

5.2 Method

5.2.1. Participants

Sixty-four participants took part in the experiment and received course credits or payment for their participation. Each participant completed a self-reporting language questionnaire at the end of the experiment to ensure that they were native English speakers and that they did not have any knowledge of the FL (Welsh) prior to the experiment. Accordingly, two participants were excluded because they reported having prior knowledge of Welsh (although very minimal) and one participant was excluded because they were a non-native speaker of English. The final sample therefore included 32 participants (mean age = 20.4, 28 females) in the same-day group, and 29 participants (mean age = 22.1, 22 females) in the next-day group.

5.2.2. Design

For both groups of participants a 2 x 2 x 2 repeated-measures design was used for the priming task with language of the prime (Welsh or English), type of target (pictures or written NL words) and type of trial (prime-target relationship: matched or unrelated) as within-subject factors. For the explicit learning phase, a repeated-measures design was used with type of word (old vs. new) as a within-subject factor.

5.2.3. Stimuli

The stimuli for Phase 1, the incidental learning phase, consisted of two sets of 30 auditory and written Welsh words (see Appendix C) with their corresponding line drawings (Snodgrass & Vanderwart, 1980). These were a subset of the words used in Chapters 3 and 4 that could be categorized as manmade or natural (half of the words in each set belonged to each category). The two sets of words were counterbalanced across participants for this phase of the experiment and became the old words thereafter. The other set of words was then used as the new words for Phase 3 of the experiment, i.e., the explicit learning phase. Each set of words was matched for word length in English and Welsh, category (Snodgrass & Vanderwart, 1980) and word frequency in English (Celex and BNC).

In Phase 2 of the experiment (the priming task), the same 30 auditory Welsh words from Phase 1 were used as primes, and the written English translations and pictures depicting the meaning of the words were used as targets. The pictures for this part of the experiment came from Moreno-Martínez and Montoro (2012), Brodeur, Dionne-Dostie, Montreuil and Lepage (2010) and from multiple websites (see Appendix C). Importantly, pictures for this part of this experiment were not the same as the pictures used in Phase 1.

In one third of trials, targets were a written translation of the prime or they were a picture depicting its meaning (matched trials). In one third of the trials (unrelated trials) the targets were unrelated to the prime and came from the opposite semantic category. For example, ‘horse’ (natural) served as prime for ‘necklace’ (manmade). The final third of trials served as filler trials and in this case, prime and target came from the same semantic category, i.e., both were manmade or natural. For example, 'arrow' served as prime for 'glass' and both are manmade, whereas 'fly' served as prime for 'mushroom', and both are natural. The filler trials ensured that participants did not develop a response switching strategy when targets were unrelated to the primes. The presentation of the targets within each block was pseudo-randomized with at least 15 trials occurring in between each repetition of the same target. The type of target (picture or written word) was blocked and the order of the presentation of the blocks was counterbalanced across participants.

Following the 2 blocks of Welsh priming (one block with written English words as targets and one block with picture targets), all participants also completed 2 blocks using English primes to investigate priming effects in the NL. The English priming task was identical to the Welsh priming task.

In Phase 3 of the experiment (the explicit learning phase) the auditory Welsh words from Phase 1 and 2, as well as auditory Welsh words that were new to the participants were used in combination with their written English translations as in Chapters 3 and 4.

5.3 Procedure

In Phase 1 of the experiment, participants completed a letter-search task following the same procedure as Chapters 3 and 4. However, each word was repeated 10 times, half of the presentations with a letter that was included in the written word and half of the presentations with a letter that was not.

In Phase 2 of the experiment, participants were told that they would hear FL words and English words and that they would then see either a written word or a picture. They were asked to perform a semantic categorization task on the written word or on the picture. The words or pictures had to be categorized as manmade or natural, and they were given examples of each. Participants then received specific written instructions as well as 6 practice trials at the beginning of each block. Trials started with the presentation of a fixation cross for 500 ms, followed by the presentation of the auditory prime (average prime duration = 762 ms, range 423 ms to 1077 ms). The target appeared 800 ms following the onset of the auditory prime, and remained on the screen until the participant responded.

The procedure for Phase 3 was the same as the explicit learning phase in Chapters 2 to 4, except that there were only 30 old and 30 new words to learn, and that participants only completed 2 blocks of learning. The number of blocks was reduced to 2 in order to keep the duration of the experiment within 1 hour. Furthermore, in previous chapters, it was found that the difference in accuracy scores between old and new words already emerged in block 1, and that some participants reached criterion of 80% accuracy in block 2 and therefore did not proceed to block 3.

In the next-day group, participants were asked to complete the priming task once more following the explicit learning task, and this followed the same procedure as per Phase 2. This additional priming task was completed in order to explore the priming effect following both incidental and explicit learning. Since FL words are presented in combination with translation equivalents in the explicit learning phase, it was expected that this type of learning would lead to a priming effect for the word targets.

5.4 Results

5.4.1. Phase 1: Incidental learning

The accuracy during the letter-search task was high for both groups, indicating that participants were attending to the written words (95.3 % accuracy for the same-day group and 94.4% accuracy for the next-day group).

5.4.2. Phase 2: Priming task

Filler trials, incorrect responses on the semantic categorization task (same-day group = 4.5% of data, next-day group = 6.2% data), and outlier values (responses faster than 250 ms and slower than 2000 ms;[15] same-day group = 1.1% of data, next-day group = 2.1 % data) were

[15] There does not seem to be a consensus as to what is considered an outlier value in the literature. For example, Holcomb and Henderson (1993) and Guo et al., (2012) chose to eliminate values < 200 ms and > 2000 ms, whilst Zhang, van Heuven and Conklin (2011) considered as outliers values < 300 ms and > 2500 ms. Here I chose values which fall within the range used in prior priming studies.

removed before analyses. All remaining trials were used for the English primes. For the Welsh primes, as priming was only expected for the FL words for which learning occurred during the incidental learning phase, the answers in block 1 of the explicit learning task were used to identify the Welsh words for each participant where both hits and correct rejections were obtained (49.0% data points remained for the same-day group, and 46.8% data for the next-day group). Response times during the priming task were averaged for each prime language (Welsh and English), each target type (picture and NL word) and each trial type (match vs. unrelated). These were then log transformed to reduce skewness, and submitted to a 2 x 2 x 2 repeated-measures ANOVA.[16]

5.4.2.1 Same-day group

For the same-day group, results revealed significant main effects for language, $F_1(1, 31) = 72.81$, $p < .001$, $\eta_p^2 = .70$, $F_2(1, 57) = 73.61$, $p < .001$, $\eta_p^2 = .56$, for target type, $F_1(1, 31) = 28.62$, $p < .001$, $\eta_p^2 = .48$, $F_2(1, 57) = 141.07$, $p < .001$, $\eta_p^2 = .71$, and trial type, $F_1(1, 31) = 20.27$, $p < .001$, $\eta_p^2 = .40$, $F_2(1, 57) = 12.58$, $p < .01$, $\eta_p^2 = .18$. The interaction between language and trial type was significant, $F_1(1, 31) = 71.59$, $p < .001$, $\eta_p^2 = .70$, $F_2(1, 57) = 61.92$, $p < .001$, $\eta_p^2 = .52$. However, neither the interaction of target by trial type nor of language by target type were significant, $F_1(1, 31) = 1.97$, $p = .17$, $\eta_p^2 = .06$, $F_2(1, 57) = 2.66$, $p = .11$, $\eta_p^2 = .05$ and Fs < 1 respectively. The three-way interaction was significant in the analysis by

[16] Prior to the analyses by item for the same-day group, 3 items were removed because of very few data points contributing to the mean in each cell (< 4) and because they skewed the distribution.

participant, $F_1(1, 31) = 4.90$, $p < .05$, $\eta_p^2 = .14$, however not in the analysis by item, $F_2(1, 57) = 2.74$, $p < .10$, $\eta_p^2 = .05$ (see Table 5.1).

Table 5.1: Response times in millisecond for both groups of participants for English and Welsh primes for each target and trial type with standard error.

		Auditory Prime	Picture targets			Written English targets		
Group	Phase	Prime Language	Trial type		Priming effect	Trial type		Priming effect
			Unrelated	Match		Unrelated	Match	
Same-day	Phase 2	English	510 (15)	453 (15)	57*	579 (19)	517 (17)	62*
		Welsh	523 (20)	560 (24)	-37*	616 (22)	612 (20)	4
Next-day	Phase 2	English	486 (11)	455 (13)	31*	550 (15)	502 (20)	48*
		Welsh	541 (23)	540 (22)	1	591 (22)	604 (20)	-13
	Phase 4	English	491 (13)	447 (13)	44*	567 (15)	506 (15)	61*
		Welsh	479 (14)	461 (12)	18*	567 (17)	549 (16)	18*

Note. Priming = Unrelated trials – Match trials; * = $p < .05$

The significant language by trial type interaction was investigated further by computing repeated-measures ANOVA on each language separately. For the English primes, the results showed significant main effects of both target type $F_1(1, 31) = 38.05$, $p < .001$, $\eta_p^2 = .55$, $F_2(1, 57) = 109.97$, $p < .001$, $\eta_p^2 = .66$ and trial type, $F_1(1, 31) = 91.14$, $p < .001$, $\eta_p^2 = .75$, $F_2(1, 57) = 110.51$, $p < .001$, $\eta_p^2 = .66$. There was no significant interaction, *F*s < 1.

For the Welsh primes, results showed a significant main effect of target type, $F_1(1, 31) = 12.64$, $p < .01$, $\eta_p^2 = .29$, $F_2(1, 57) = 55.71$, $p < .001$, $\eta_p^2 = .49$. The main effect of trial type was significant in the analysis by participant but not in the analysis by item, $F_1(1, 31) = 4.18$, $p < .05$, $\eta_p^2 = .12$, $F_2(1, 57) = 2.24$, $p = .14$, $\eta_p^2 = .04$. There was a strong trend for an interaction between target type and trial type, $F_1(1, 31) = 3.90$, $p = .06$, $\eta_p^2 = .11$, $F_2(1, 57) = 3.76$, $p = .06$, $\eta_p^2 = .06$. As this was the case, matched and unrelated trials were compared using paired-sample *t*-tests (two-tailed, Bonferroni corrected) for both picture and written word targets. Results revealed that participants responded significantly slower for matched compared to unrelated trials when the target was a picture, $t_1(31) = 2.48$, $p < .05$, $d = 0.44$, $t_2(57) = 2.32$, $p < .05$, $d = 0.30$, but not when the target was a written word, *t*s < 1.

5.4.2.2 Next-day group

As per the same-day group, results revealed significant main effects for language, $F_1(1, 28) = 40.48$, $p < .001$, $\eta_p^2 = .59$, $F_2(1, 59) = 105.84$, *p*

$< .001$, $\eta_p^2 = .64$, target type, $F_1(1, 28) = 28.64$, $p < .001$, $\eta_p^2 = .51$, $F_2(1, 59) = 75.91$, $p < .001$, $\eta_p^2 = .56$, and trial type, $F_1(1, 28) = 8.01$, $p < .01$, $\eta_p^2 = .22$, $F_2(1, 57) = 15.09$, $p < .001$, $\eta_p^2 = .20$. There was a significant interaction between language and trial type, $F_1(1, 28) = 11.70$, $p < .01$, $\eta_p^2 = .30$, $F_2(1, 59) = 50.41$, $p < .001$, $\eta_p^2 = .46$. However, there was neither an interaction between language and target type, nor an interaction between target and trial types, *F*s < 1. Furthermore, results revealed no significant three-way interaction, $F_1(1, 28) = 1.97$, $p = .17$, $\eta_p^2 = .07$, $F_2(1, 59) = 1.02$, $p = .31$, $\eta_p^2 = .02$.

The significant language by trial type interaction was investigated further by computing repeated-measures ANOVA on each language separately. For the English primes, results revealed main effects of target and trial types, $F_1(1, 28) = 31.98$, $p < .001$, $\eta_p^2 = .53$, $F_2(1, 59) = 70.82$, $p < .001$, $\eta_p^2 = .55$ and $F_1(1, 28) = 13.04$, $p < .01$, $\eta_p^2 = .32$, $F_2(1, 59) = 116.03$, $p < .001$, $\eta_p^2 = .66$ respectively. In contrast, although there was a main effect of target for the Welsh primes, $F_1(1, 28) = 8.84$, $p < .01$, $\eta_p^2 = .24$, $F_2(1, 59) = 30.06$, $p < .001$, $\eta_p^2 = .34$, there was no main effect of trial type, $F_1 < 1$, $F_2(1, 59) = 2.93$, $p = .09$, $\eta_p^2 = .05$. There was no interaction between target and trial type for either language, *F*s < 1. The results therefore indicated that picture targets were responded to faster than word targets, and that this was the case following both English and Welsh primes. Importantly however, a priming effect (faster responses to match vs.

unrelated trials) was found only following the English primes (see Table 5.1).

5.4.3. Phase 3: Explicit learning

For each participant, the percentage of accuracy in the translation recognition task was calculated for old and new words and submitted to repeated-measures ANOVAs for each block separately. For the same-day group, results revealed a main effect of word type for each block, indicating that participants achieved higher accuracy on the old words than on the new words both in block 1, $F_1(1, 31) = 41.76$, $p < .001$, $\eta_p^2 = .57$, $F_2(1, 57) = 42.46$, $p < .001$, $\eta_p^2 = .24$, and in block 2, $F_1(1, 31) = 11.68$, $p < .001$, $\eta_p^2 = .27$, $F_2(1, 57) = 19.05$, $p < .001$, $\eta_p^2 = .24$. The same results were obtained in the next-day group, with accuracy being higher for the old words than for the new words in both block 1, $F_1(1, 28) = 35.53$, $p < .001$, $\eta_p^2 = .56$, $F_2(1, 59) = 35.44$, $p < .001$, $\eta_p^2 = .38$ and block 2, $F_1(1, 28) = 7.33$, $p < .05$, $\eta_p^2 = .21$, $F_2(1, 59) = 10.46$, $p < .01$, $\eta_p^2 = .15$ (see Figure 5.2).

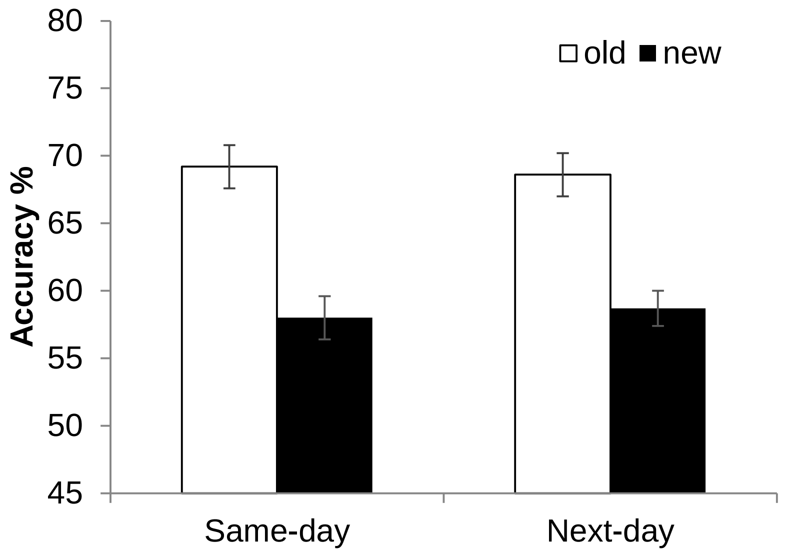

Figure 5.2: Percentage accuracy for old and new words in the translation recognition task (explicit learning phase) in the first block of trials for each group with error bars.

5.4.4. Phase 4: Priming task

Participants in the next-day group completed the priming task once more following the explicit learning task. Filler trials, incorrect responses on the semantic categorization task (7% of data), and outlier values (responses faster than 250 ms and slower than 2000 ms;1.5% of data) were removed before analyses. As in the Phase 2 analyses, all remaining trials were used for the English primes. However, for the Welsh primes, the answers in block 2 of the explicit learning task were used to identify the Welsh words for each participant where both hits and correct rejections were obtained (59.9% data points remained). Response times during the priming task were averaged for each language of the primes (Welsh and English), each target type (picture and NL word) and trial type (matched vs. unrelated). These were then log transformed to reduce skewness, and submitted to a 2 x 2 x 2 repeated-measures ANOVA.

Similar to the analysis of Phase 2, results revealed significant main effects of language, $F_1(1, 28) = 5.93$, $p < .05$, $\eta_p^2 = .18$, $F_2(1, 59) = 6.30$, $p < .05$, $\eta_p^2 = .10$, target type, $F_1(1, 28) = 107.67$, $p < .001$, $\eta_p^2 = .70$, $F_2(1, 59) = 253.56$, $p < .001$, $\eta_p^2 = .81$, and trial type, $F_1(1, 28) = 71.32$, $p < .001$, $\eta_p^2 = .72$, $F_2(1, 59) = 65.65$, $p < .001$, $\eta_p^2 = .53$. The interaction between language of the prime and trial type was significant, $F_1(1, 28) = 4.58$, $p < .05$, $\eta_p^2 = .14$, $F_2(1, 59) = 4.54$, $p < .05$, $\eta_p^2 = .07$, as well as the interaction between target type and trial type, $F_1(1, 28) = 20.83$, $p < .001$, $\eta_p^2 = .43$, $F_2(1, 59) = 12.18$, $p < .01$, $\eta_p^2 = .17$. The three-way interaction was not significant, *F*s < 1.

The significant two-way interactions were investigated further by submitting each language to separate repeated-measures ANOVAs. For both the English and the Welsh primes, the main effects of target, $F_1(1, 28) = 73.49$, $p < .001$, $\eta_p^2 = .72$, $F_2(1, 59) = 129.29$, $p < .001$, $\eta_p^2 = .69$, $F_1(1, 28) = 86.23$, $p < .001$, $\eta_p^2 = .76$, $F_2(1, 59) = 124.63$, $p < .001$, $\eta_p^2 = .68$ and of trial type $F_1(1, 28) = 118.99$, $p < .001$, $\eta_p^2 = .81$, $F_2(1, 59) = 77.41$, $p < .001$, $\eta_p^2 = .57$, $F_1(1, 28) = 7.48$, $p < .05$, $\eta_p^2 = .21$, $F_2(1, 59) = 8.50$, $p < .01$, $\eta_p^2 = .13$ were significant. However, there was no interaction, *F*s < 1. Response times following Welsh and English primes were then compared across languages using paired-sample *t*-tests (Bonferroni corrected). Results revealed that participants responded to word targets faster when they were preceded by English primes than when they were preceded by Welsh primes, $t_1(28) = -2.76$, $p < .05$, $d = -0.51$, $t_2(59) = -3.24$, $p < .01$, $d = -0.42$, however there was no significant difference between the two languages in response to picture targets, *t*s < 1 (collapsed across trial types). Furthermore, responses were significantly faster for matched trials following English primes compared to Welsh primes, $t_1(28) = -4.07$, $p < .001$, $d = -0.76$, $t_2(59) = -4.31$, $p < .001$, $d = -0.56$, however, there was no significant differences between the two languages for unrelated trials, $t_1(28) = 1.42$, $p = .33$, $d = 0.26$, $t_2 < 1$ (collapsed across targets). Finally, a comparison of response times for matched trials with word targets following Welsh and English primes revealed a larger priming effect when the primes were cross-modal repetitions of the targets (auditory English primes with written English targets) than when they were cross-modal translations (auditory Welsh primes with written English translations), $t_1(28) = -3.52$, $p < .01$, $d = -0.65$, $t_2(59) = -3.84$, $p < .001$, $d = -0.50$.

5.5 Discussion

The aim of this experiment was to explore the lexical-semantic links between FL and semantic representations, as well as between FL and NL translation equivalents following an incidental learning situation involving pictorial information. Two hypotheses were put forward to explain the incidental learning effect, form-meaning links and form-form links. In view of the results however, it is not possible to conclude either way, as there is evidence in support of both.

For the group of participants completing the priming task immediately following the incidental learning phase, a significant reversed priming effect in response to picture targets was found. In other words, participants responded slower overall to pictures when FL primes matched them in meaning compared to unrelated primes. However, no priming effect emerged for word targets. This suggests a link between the FL word representations and the semantic representations, however, in view of the nature of the priming effect, this link must still be weak at this stage of the learning process. In contrast, for the group of participants completing the priming task the day after the incidental learning phase, no priming effect emerged in response to picture targets. Furthermore, although participants responded slower to word targets following a translation prime compared to an unrelated prime, this difference was not significant.

For one group of participants (next-day group), the priming task was repeated following the explicit learning task, and clear priming effects emerged both for picture and word targets. Unfortunately, the results of the priming task did not clarify whether the explicit learning led to form-meaning links, form-form links, or to both. In this study, the stimulus onset asynchrony (SOA) between the presentation of prime and target was quite long as the target appeared 800 ms following the onset of the auditory prime. This long SOA may have allowed activation to spread from either semantic to lexical representations or vice-versa and therefore we cannot conclude whether form-meaning or form-form links are responsible for the effect. In other words, when participants heard the FL word forms, this may have activated their semantic representations (in the case of form-meaning link), which in turn may have activated the NL word representations and lead to a priming effect for word targets. However, it is also possible that on hearing the FL prime, the NL translation equivalent became activated (in the case of form-form links), and in turn activated the semantic representations which led to a priming effect for picture targets. However, if either were true, presumably one would lead to a bigger priming effect than the other. For example, if form-meaning links were responsible for the effect, a bigger priming effect would be expected for picture targets, as semantic representations would become activated prior to NL translations. In contrast, if form-form links were responsible for the effect a bigger priming effect should have occurred for word targets, as NL representations would have been activated prior to semantic representations. However, the priming effect was of the same magnitude both for picture and word targets. It is therefore possible that activation of both semantic representations and NL translation equivalents occurred in parallel, and therefore led to priming for both pictures and words. This seems plausible since the presentation of pictures in combination with FL word forms during the incidental learning phase would have encouraged form-meaning links, whilst the presentation of FL word forms in combination with NL translation equivalents during the explicit learning phase would have encouraged the creation of form-form links. Therefore, as this group of participants completed both the incidental and explicit learning phase prior to completing the priming task a second time, both types of links are expected. Interestingly however, for this group of participants, priming for picture targets did not occur when they first completed the priming task the day after the incidental learning phase. In contrast, priming for picture targets occurred following the explicit learning phase. It seems therefore that the incidental learning phase on its own was not enough to lead to the creation of form-meaning links, or that the priming task was not sensitive enough to detect an early learning effect. However, accuracy results for the first block of trials in the explicit learning phase were higher for the words presented during the incidental learning phase. Furthermore for the group of participants completing the priming task on the same day as the incidental learning phase, a reversed priming effect was found. Taken together, this suggests that incidental learning of the FL words did occur during the exposure to the FL words in combination with pictures. Therefore, it will be necessary to find a more sensitive method in order to explore the time-course of lexical-semantic activation further.

One method that has been shown to be sensitive to both lexical and semantic manipulation is EEG. For example, using EEG, Guo, Misra, Tam and Kroll (2012) investigated the time-course of lexical and semantic activation whilst fluent bilinguals completed a translation recognition task. In their first experiment, they found a reduction in the N400 (indexing semantic integration) in response to incorrect translations that were related in meaning (semantic distractors) compared to unrelated word pairs, indicating that participants accessed the meaning of the words to complete the task. However, they also found a reduction in the P200 (indexing lexical processing) in response to incorrect translations that were similar in form (lexical distractors) compared to unrelated distractors, suggesting that participants also accessed lexical representations. Importantly, the authors highlighted that their use of a long SOA (750 ms) may have allowed participants to access both, and it was not possible to tell whether the activation of semantic representation or NL lexical representation occurred first following the presentation of the FL word forms. In order to solve this issue, a second experiment was conducted, this time with a shorter SOA (300 ms), and results revealed an effect on the N400 component, but no effect on the P200, indicating that participants accessed the semantic representation first. Furthermore, an effect on the late positive component (LPC) suggested that NL lexical representations were activated, but that this occurred following the activation of the semantic representations. As participants were fluent bilinguals, it is not surprising that they activated semantic representations upon reading second language words. Importantly however, this study highlighted the sensitivity of EEG to both semantic and lexical distractors during the processing of FL words. Furthermore, the behavioral results were similar for both types of distractors, and differences only emerged in the EEG data. Finally, a short SOA duration was crucial in revealing differences in the time-course of activation, and it may therefore be the case that in the current study a shorter SOA should have been used.

The Guo et al. experiment mentioned above used an explicit translation recognition task with both FL and NL word forms presented in writing, and even at short SOA (300 ms) participants had enough time to recognize the FL word form prior to the appearance of the NL word form. In contrast, in the current study FL primes were presented auditorily. Furthermore, the priming task involved a FL auditory prime and either pictures or written word targets, and as such, involved cross-modal priming. Studies using cross-modal repetition priming with auditory primes and written targets in the NL, suggest that priming effects do occur across modalities and at different SOAs. For example, Holcomb and Anderson (1993) found cross-modal repetition priming effects at 0 ms, 200 ms and 800 ms SOAs. However, Holcomb, Anderson and Grainger (2005) found significantly bigger repetition priming effects at longer SOAs (800 ms) compared to both 0 and 200 ms SOAs. Furthermore, in both studies, EEGs were also recorded, and the results in both cases showed a robust N400 effect only with an 800 ms SOA. Taken together, the results of these two studies suggest that bigger and more robust cross-modal repetition priming occurs at longer SOA (800 ms). Furthermore, Dobel et al. (2009) used a cross-modal semantic priming task (with NL pseudoword auditory primes and picture targets) with targets appearing 300 ms post prime offset with an average prime duration of 400 ms, and therefore their SOA was approximately 700 ms. The choice of SOA in the current study was based on both the cross-modal repetition priming and cross-modal semantic priming studies mentioned above, hence the long 800 ms SOA. Furthermore, it was thought that the longer SOA would be best, as most auditory primes would have offset before the onset of the target (average prime duration = 762 ms, range 423 ms to 1077 ms). Furthermore, because of their lack of experience with the FL, it was thought unlikely that participants would be able to recognize the FL words much before their offset. However, the disadvantage of a long SOA, as discussed earlier, is that it allowed cascading activation of lexical and/or semantic representations, and therefore the type of learning, form-meaning or form-form that best explains the learning in the current study remains unclear. It is likely that EEG would provide crucial additional information, and as such, a further study should be conducted with a shorter SOA as well as EEG recording.

5.6 Conclusion

In the current study, participants were exposed to FL word forms and pictures during an incidental learning phase, and it was thought that this would lead to the creation of form-meaning links. Evidence in support of this was found with the group of participants completing the priming task immediately following the incidental learning task, although the effect was in the opposite direction, i.e., their responses were slower. With the group of participants completing the priming task the day after the incidental learning task, no priming effect was found to support the form-meaning link hypothesis. Following the explicit learning task, priming was found in support of both form-meaning and form-form links, and therefore future studies using shorter SOAs and EEG recordings will be necessary to investigate whether the presentation of pictorial information during incidental learning encourages the creation of form-meaning links. In the next chapter, the role of pictorial information in incidental word learning will be explored further.

Chapter 6: The role of verbal and pictorial information in multi-modal incidental acquisition of foreign language vocabulary

6.1 Introduction

The results of Chapters 3 to 5 showed that being presented with FL word forms, both auditory and written, as well as a picture in an incidental learning situation led to higher accuracy scores for those words in the explicit learning task. This incidental learning advantage was found immediately after the incidental learning situation, the next day, as well as one week later (Chapter 3). Furthermore, the benefits of having been exposed to FL words in an incidental learning situation was found following as little as two encounters with the FL words, and results showed that the first few encounters with the FL words led to more gains in learning than subsequent encounters (Chapter 4). Finally, in Chapter 5, it was found that learners showed a reversed semantic priming effect when presented with a picture following an auditory FL word form during a priming task conducted on the same day as the incidental learning phase.

Importantly, in all the above-mentioned chapters, FL acquisition occurred when both the auditory and written word forms presented during the incidental learning phase were in the FL. In order to derive meaning from the FL word forms, participants had to process the pictures. This is similar to the intralingual subtitle situation presented in Chapter 2, in which both the soundtrack and subtitles are presented in the FL. The advantage of this type of situation is that even complete beginners can benefit from it, as the pictures provide meaning information, and therefore the meaning of the FL words does not have to be derived from the context as in, for example, reading. However, in a multi-modal situation, such as a film with subtitles, it is also possible to provide meaning information in one of the verbal streams, i.e., through either the audio or written information. In fact, most subtitled films are presented with FL soundtrack and NL subtitles (standard subtitling). With standard subtitles, one can enjoy the visual aspects of the film, and understand the story through the NL subtitles. Although it is clear that people read NL subtitles (as was found in Chapter 2 and in d'Ydewalle and de Bruycker, 2007), it is less certain whether the FL words in the soundtrack are also processed. Since the subtitles provide a NL translation of the dialogue, there is no need to attend to the FL words. However, in order to learn FL vocabulary from watching a FL film with NL subtitles, it is essential that the FL word forms in the soundtrack are processed. In the literature, a few incidental learning studies using FL films with NL subtitles concluded that learning occurred (Koolstra & Beentjes, 1999; d'Ydewalle & Pavakanun, 1995; d'Ydewalle & Pavakanun, 1997; d'Ydewalle & van de Poel, 1999), which suggests that the FL words in the soundtrack were also processed. Furthermore, Saffran (1997) provided further evidence for the processing of irrelevant auditory information during an incidental learning task involving a statistical learning paradigm. Results of this study showed that even background auditory information that was not relevant for a task was recognised by participants. However, the auditory information consisted of only 6 nonwords repeated 300 times, and therefore it is difficult to compare this situation to a film soundtrack where words are not repeated so much.

In the multi-modal situation used in the experiments reported in Chapters 3 to 5, words were repeated from 2 to 8 times only, and this was sufficient for FL vocabulary acquisition to occur. Participants were able to recognise the correct translation equivalent of the FL auditory words in the test phase, suggesting that even irrelevant auditory information was processed during the incidental learning phase. However, because the written FL word forms were also provided, it is also possible that this was responsible for the learning, or at least contributed to it, especially since the letter-search task required participants to search the written FL word. Participants may have linked both the written and auditory FL word form to the meaning representation accessed from the pictures during the letter-search task. Having had access to both written and auditory FL word forms may have facilitated learning (see Bird & Williams, 2002; Hu, 2008; Ricketts, Bishop, & Nation, 2009; Rosenthal & Ehri, 2008). Since accuracy scores on the letter-search task were high, this indicates that the written FL word forms were indeed processed, and as such may have contributed to the incidental learning effect. In contrast, it is less clear to what extent the FL auditory word forms were processed and how much they contributed to the learning effect. The first aim of the current chapter was therefore to assess the incidental acquisition of FL vocabulary using a situation similar to a film with standard subtitles. The simple multi-modal situation used in Chapters 3 to 5 was repeated here, however, written translations of the FL auditory word forms were provided. Therefore, as the written word forms necessary to complete the letter-search task are presented in the NL, if learning occurs, this indicates that the FL auditory word forms were processed.

Another interesting question, which arises from using a situation similar to standard subtitling, is whether there is an added benefit of having access to pictures in addition to written translations (see Chapter 1, Section 1.5.1). Since participants can now access the meaning of the FL words through the NL translations, the meaning information provided by the pictures becomes redundant. Furthermore, the pictures are not necessary to complete the letter-search task, whilst the written translations are. As mentioned in Section 1.3.4 (see Chapter 1), previous work on free recall of information predicts that having access to a picture during encoding will be beneficial for recall (picture superiority effect), as a picture can be encoded both as verbal and nonverbal information (Paivio & Csapo, 1973). This dual-coding theory suggests that both verbal and nonverbal information can then serve as a cue at retrieval (Paivio & Csapo, 1973). Furthermore, Nelson et al. (1977) suggested that this picture superiority effect is due to the distinctiveness of pictorial information. These authors explain that pictures include more varied features than words, and this is beneficial for retrieval. Although little is known about the impact of pictorial information in an incidental learning paradigm involving a FL, the picture superiority effect predicts that the use of multimodal (visual and verbal) input will benefit learning. The second aim of this chapter was therefore to assess the impact of pictorial information on the incidental acquisition of FL vocabulary.

In order to address this question, a similar letter-search task was used to provide an incidental learning situation as in the previous chapters. However, the type of information presented for each FL word varied within participants. Participants were presented with three different types of trials: auditory FL word forms only, auditory FL word forms with written NL translations, and both auditory FL word forms with written NL translations presented with simple line drawings depicting the meaning of the words (see Figure 6.1).

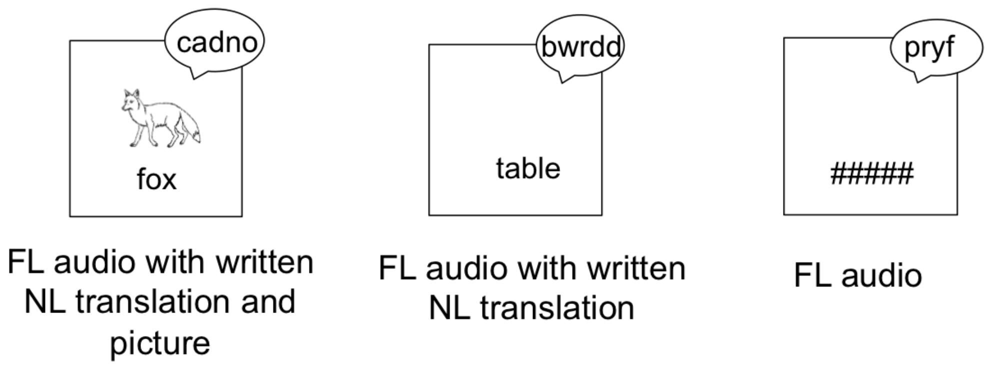

Figure 6.1: Example of the three types of word presentations during the incidental learning phase (letter-search task)

In the previous chapters, incidental vocabulary acquisition was assessed through an explicit learning task using translation recognition with feedback. In the present chapter the same task was used to assess learning the day following the incidental learning phase. In contrast to previous chapters, participants returned also one week later to complete a recall test and a translation recognition test (without feedback).

It was predicted that having access to all three types of information (auditory FL word forms, written NL translations and pictures) would be beneficial for learning as assessed by accuracy scores on the explicit learning task the next day, as well as on the recall and recognition tests one week later. Furthermore, the accuracy scores on both word conditions presented with meaning, whether written translations or both written translations and pictures, should be higher than for the words presented with auditory FL word forms only. Furthermore, a control group who did not take part in the incidental learning phase completed both the explicit learning task and the delayed recall and recognition tests. It was predicted that the experimental group would outperform the control group for words presented with meaning in the incidental learning phase. Furthermore, even though the control group was not presented with the different word conditions (this group learnt to link all auditory FL word forms and NL translations during the explicit learning phase) their accuracy scores were still split into the three word conditions (FL auditory word forms only; FL auditory word forms with written translations; FL auditory word forms with written translations and pictures) for the purpose of the analyses.

As was explained earlier, because the current experiment presented the written information in the NL, it was possible to assess whether the irrelevant FL auditory word forms were processed because this was the only way participants could learn the FL word form. However, as meaning information was presented through both NL translations and pictorial information, the learning of FL word meaning could occur whether either or both were processed. The third aim of this chapter was to use eye-tracking to assess the allocation of attention to the different elements of the multi-modal situation (See Chapter 1, Section 1.5.1), as well as their impact on learning. It was predicted that participants would spend a considerable amount of time looking at the written word because of the letter-search task. However, the processing of the pictures was less certain because they were irrelevant for the letter-search task. In order to increase the likelihood that participants would look at the pictures, these were presented slightly earlier than the onset of the written words (300 ms). This timing of the presentation of the different elements of the multi-modal situation also mimicked a film with subtitles: Film viewers are normally looking at the images in a film when subtitles appear on screen. It was important that participants looked at the pictures since the eye-tracking data will also serve to investigate the impact of the pictorial information on the learning of the FL words. Based on the picture superiority effect found in the memory literature mentioned above, it was predicted that the time spent looking at the pictures would predict scores on the explicit learning task and the recall and recognition test. It was also predicted that the time spent looking at the NL word would predict the learning outcomes, but that the time spent looking at the pictures would be a stronger predictor.

Finally, as it was found in Chapter 4 that the first few exposures to FL words and pictures led to bigger learning gains than subsequent exposure, the eye-tracking data was also used to investigate the allocation of attention to the different elements of the multi-modal situation *across the duration of the incidental learning situation*. The incidental learning phase was therefore split into 6 blocks of trials for the purpose of the analysis. The prediction was that the time spent looking at the pictures would decrease across the duration of the incidental learning phase. In contrast, as the written words were necessary to complete the letter-search task, it was predicted that participants' viewing behaviour would be more constant throughout the experiment.

6.2 Method

6.2.1. Participants

Sixty-six students from the University of Nottingham took part in the experiment and received course credits or payment for their participation. All participants completed a self-reporting language questionnaire at the end of the experiment to ensure that they were native English speakers and that they had no prior knowledge of the foreign language (Welsh) used in the experiment. Accordingly, ten participants were excluded on the basis that they were either non-native speakers of English (4 participants) or they had prior knowledge of Welsh or a related language (6 participants). A further two participants were excluded from the analyses as they were unable to complete all parts of the experiment, and one further participant was excluded because of technical difficulties with the eye-tracker. The final sample therefore included 28 participants in the experimental group (mean age = 23.1, 21 females) and 25 participants in the control group (mean age = 22.8, 18 females).

6.2.2. Design

A mixed-design was used in this experiment with group as a between-subject factor (2 levels: control group and experimental group) and type of stimuli presented during the incidental learning phase (3 levels: auditory FL word form only [A], auditory FL word form and written NL translation [AW], auditory FL word form and written NL translation with a picture illustrating the meaning [AWP]) as a within-subject factor.

6.2.3. Stimuli

The stimuli consisted of seventy-eight auditory Welsh words and their written English translations (see Appendix D) and pictures illustrating the meaning of the words. The words and pictures were the same as those used in the experiments reported in Chapters 3 to 5, except that 2 words were removed in order to create 3 sets of 26 words. Three lists of stimuli including 26 words from each word condition (A, AW and AWP) were then created to allow for counterbalancing across participants.

6.3 Procedure

There were three phases to this experiment: Phase 1 was the incidental learning phase, Phase 2 was the explicit learning phase and Phase 3 was the recall and recognition phase. The experimental group completed all the phases, whilst the control group started with Phase 2.

Phase 1 was completed on the first day of the experiment, Phase 2 the next day, and finally, Phase 3 one week later.

6.3.1. Phase 1: Incidental learning

In Phase 1 of the experiment, participants in the experimental group completed a letter-search task whilst their eye-movements were recorded with an Eyelink 1000 (SR Research, Canada) desktop eye-tracker (sampling rate of 1000 Hz). A chin rest was used to immobilize participant's head.

The procedure was similar to the one used in Chapter 3 to 5, except that each stimulus was presented 6 times only: 3 times with a letter that was present in it, and 3 times with a letter that was not. All 78 words were used in the incidental learning phase, resulting in a total of 468 trials. The trials were split into 6 blocks of 78 trials, with each word appearing once in each block. For the stimuli that included auditory FL word forms only, participants saw a series of hash symbols (#####) instead of the written NL translation appearing in the other two word conditions (see Figure 6.1). Participants were instructed to respond 'no' for these trials, i.e., the letter was not contained in the written word as no written word was presented. All written stimuli were displayed using Courier New font, size 20 in bold letters. Importantly, the written word forms were presented in the NL of the participants (English) unlike the experiments reported in Chapters 3 to 5.

The session started with the set up and calibration of the eye-tracker using a nine-point calibration grid. Following this, a series of 8 practice trials with feedback preceded the start of the main experiment. Each trial in the experiment started with the presentation of the to-be-searched letter at the top of the screen for 500 ms. Then the picture or a blank screen was displayed in the middle of the screen for 300 ms before the presentation of the written word (or hash symbols) at the bottom of the screen. The auditory FL word onset was simultaneous with the onset of the written word. The picture and written word stayed on screen until the end of the trial. The termination of each trial occurred when participants made a response, unless the auditory FL word was still playing. In those cases, the trial ended with the offset of the auditory FL word. A short break was included after each block allowing for the re-calibration of the eye-tracker. For this part of the experiment, the stimuli were displayed using SR Research's Experiment Builder software.

6.3.2. Phase 2: Explicit learning

Both groups of participants completed the explicit learning phase in which they were asked to complete the translation recognition task. The procedure for this phase was identical to the explicit learning phase used in the experiments reported in Chapters 3 to 5.

6.3.3. Phase 3: Recall and translation recognition test

Participants in the experimental and control groups came back one week after Phase 2 to complete a recall and translation recognition test. In the recall test participants were presented with the auditory Welsh words, and were asked to type their English translations. Participants were encouraged to enter a translation even if they were not sure it was correct. They also had the option of simply pressing the 'enter' key to proceed to the next trial without entering an answer. No feedback was provided during this recall test. The translation recognition test was the same one as used in Phase 2, except that no feedback was provided, and participants only completed one block of 156 trials (2 x 78: each target word was presented once with its correct translation and once with a foil). As in Phase 2, E-Prime was used for this part of the experiment to present the stimuli and record the responses. In both tasks the ordering of the trials was randomized for each participant.

6.4 Results

6.4.1. Phase 1: Incidental learning and eye-tracking

The accuracy scores on the letter-search task were high (M = 98.5%, SE = 0.2%) indicating that participants attended to both the letter and the written word.

For the purpose of the eye-tracking analyses, the screen display area was segmented into 4 regions of interest: the letter area, the image area, the word area, and the remaining display area (see Figure 6.2). Except the latter region, all regions of interest were centered along the x-axis. The letter area consisted of a region of 55 by 54 pixel starting 224 pixels from the top of the display area. The image area started on average 30 pixels below the letter area and varied in size according to the picture (average size 181 by 153 pixels). The word area started on average 97 pixels below the image area and also varied in size according to the length of the word (average size 137 by 76 pixel).

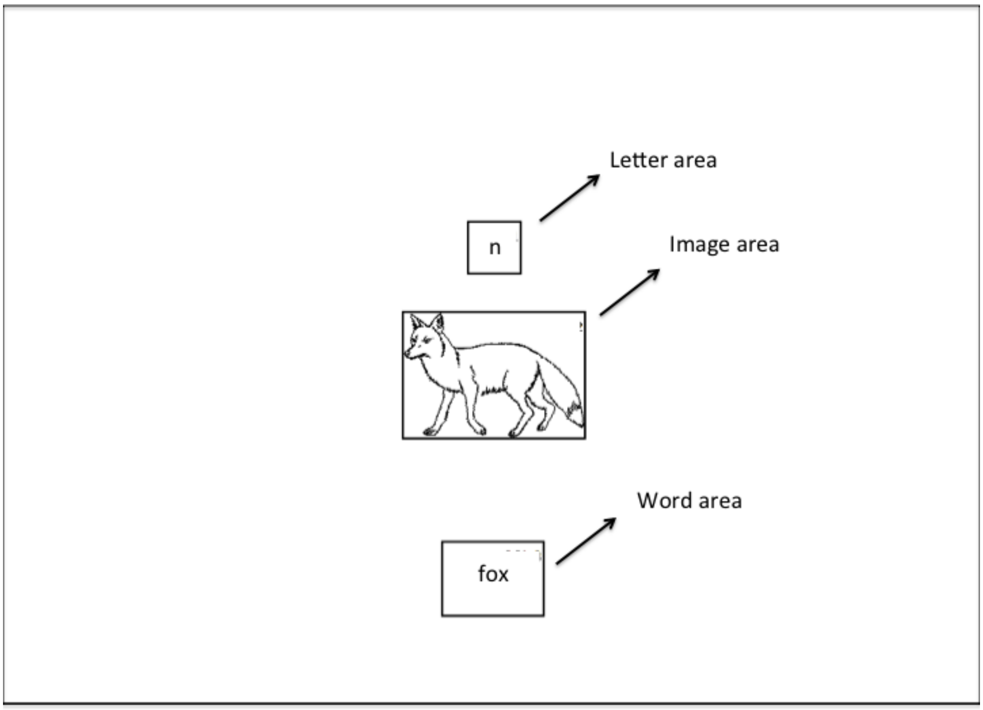

Figure 6.2: Example of a display screen showing the location of the letter area, the image area, and the word area. Fixations occurring elsewhere on the screen were defined as occurring in the "other" area (not actual size of stimuli).

The mean dwell time (total duration of all the fixations) in each region of interest was calculated for each participant and each block and averaged for each word condition (see Table 6.1).

Table 6.1: Mean (standard error) dwell time in millisecond for each word condition and region of interest per trial.

	Audio	Audio-written	Audio-written-picture
Area	Dwell time (ms)	Dwell time (ms)	Dwell time (ms)
Letter	278 (24)	275 (23)	252 (20)
Image	226 (30)	202 (26)	576 (46)
Word	974 (56)	1175 (54)	800 (56)
Other	452 (31)	417 (30)	385 (29)
All areas total	1931 (35)	2070 (34)	2014 (33)
Mean RT	634 (23)	773 (29)	717 (30)

Note. Response time is calculated from the moment the word appeared on the screen in contrast to the dwell time, which was based on the whole trial duration.

6.4.1.1 Dwell time in image area

As a picture was not presented in the condition with auditory FL and written information (AW) or auditory FL only (A) (both these word conditions included a blank screen instead) the dwell time in the image area was only investigated in the condition with auditory, written and pictorial information (AWP). The dwell time in the image area was investigated using a repeated-measures ANOVA with block as a within-subject factor (6 levels: blocks 1 to 6). Results revealed a significant linear trend, $F(1, 27) = 32.59$, $p <. 001$, $\eta_p^2 = .55$, indicating that participants spend less time fixating in the image area as they progressed through the experiment (Figure 6.3).

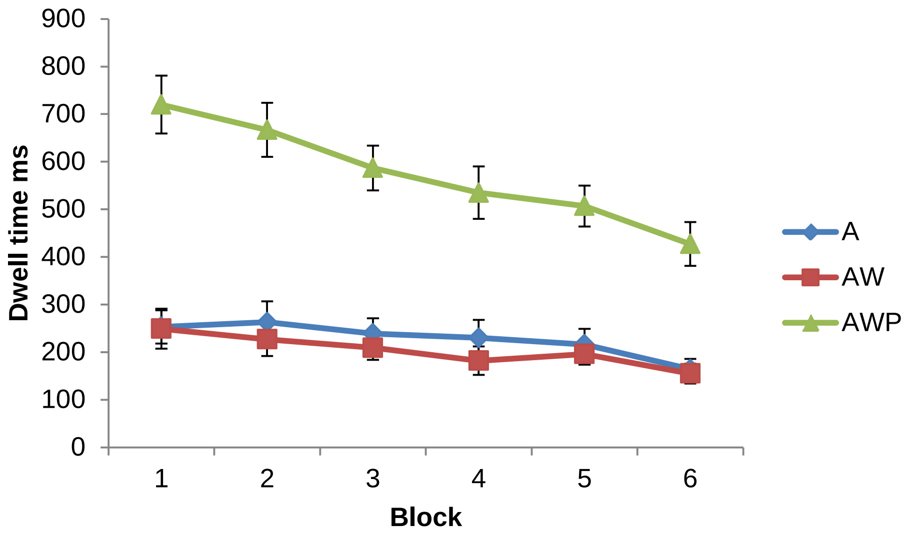

Figure 6.3: Dwell time in the image area in milliseconds (ms) across the incidental learning phase for each word condition with error bars (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

6.4.1.2 Dwell time in word area

For the word area, the mean dwell times were submitted to a 6 x 3 repeated-measures ANOVA with block number (6 levels: block 1 to block 6) and word condition (3 levels: A, AW, AWP) as within-subject factors. The results revealed a main effect of block, $F(3.20, 86.48) = 5.80$, $p <. 01$, $\eta_p^2 = .18$ and a main effect of word condition, $F(1.37, 36.99) = 72.58$, $p <. 001$, $\eta_p^2 = .73$, as well as a significant interaction, $F(5.16, 139.37) = 3.06$, $p <. 05$, $\eta_p^2 = .10$. Pairwise comparisons (Bonferroni corrected) were conducted to investigate the main effect of word condition further. These revealed that participants spent significantly longer looking in the word area in the condition with auditory FL and written NL translation compared to both the condition with auditory FL word form only (where a series of hash symbols was presented), and the condition with auditory FL, written translation and picture, all *ps* < .001 (see Table 6.1). Interestingly, participants spent longer fixating on the hash symbols in the auditory FL condition than on the word in the condition with auditory FL, written translation and picture, $p < .001$. In other words, participants spent the least time looking in the word area when a picture was presented. To break down the interaction between word condition and block, separate repeated-measures ANOVAs were computed for each word condition. These revealed significant linear trends for both the condition with auditory FL and written translations and the condition with auditory FL only, $F(1, 27) = 14.15$, $p <. 01$, $\eta_p^2 = .34$ and $F(1, 27) = 13.24$, $p <. 01$, $\eta_p^2 = .33$ respectively. For the condition that also included a picture (AWP), there was a significant quadratic trend, $F(1, 27) = 9.75$, $p <. 01$, $\eta_p^2 = .27$ (see Figure 6.4).

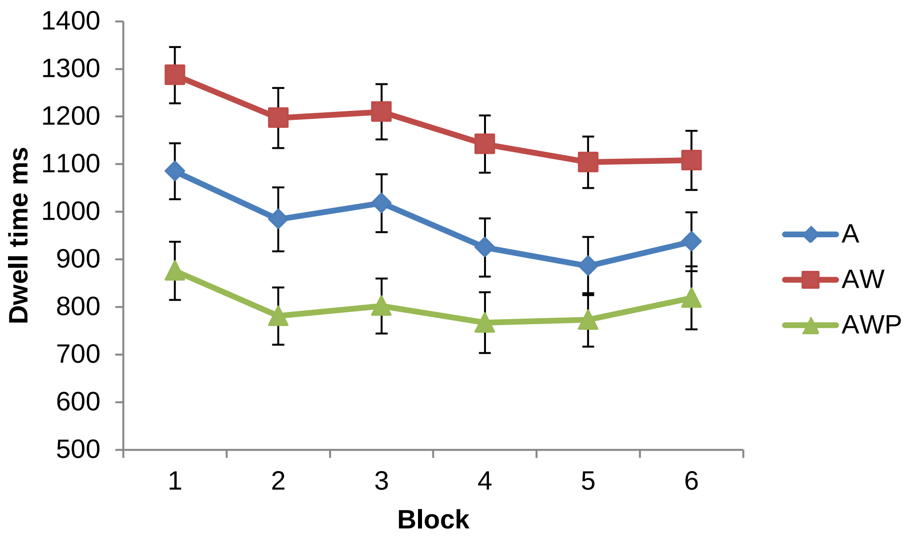

Figure 6.4: Dwell time in milliseconds (ms) in the word area across the incidental learning phase for each word condition with error bars (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

6.4.1.3 Dwell time in letter area

The mean dwell times in the letter area were also submitted to a 6 x 3 repeated measures ANOVA with the same factors as above. The results revealed a main effect of condition, $F(2, 52) = 13.04$, $p <. 001$, $\eta_p^2 = .33$, however neither the main effect of block nor the interaction between word condition and block were significant $F(5, 135) = 1.67$, $p = .33$, $\eta_p^2 = .04$ and $F < 1$ respectively (see Figure 6.5). To investigate the main effect of condition further, pairwise comparisons (Bonferroni corrected) were conducted. These revealed that participants spent longer fixating in the letter area for both the condition with auditory FL word only and the conditions with auditory FL word with written translation compared to the condition that also included a picture (AWP), $p < .001$ and $p < .01$ respectively (see Table 6.1). However there was no significant difference between both conditions that did not include a picture (A vs. AW, $p = 1$).

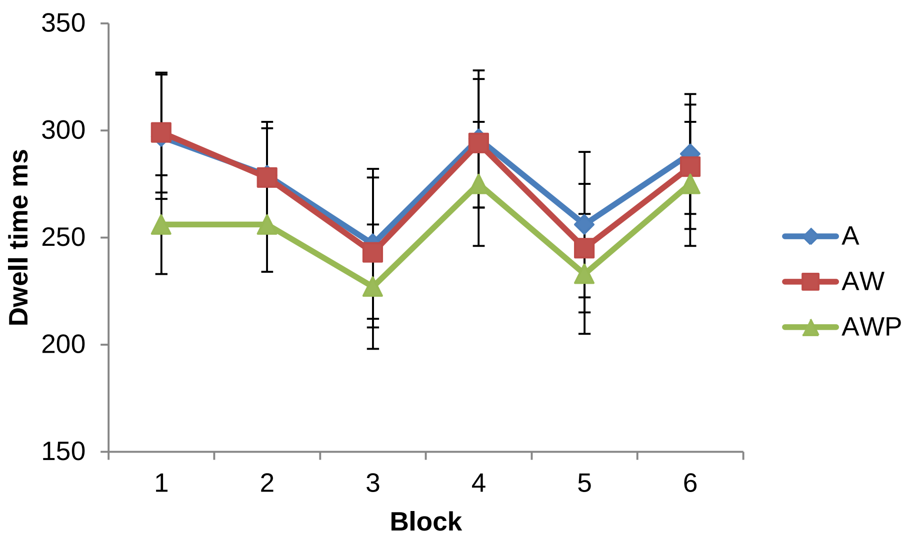

Figure 6.5: Dwell time in milliseconds (ms) in the letter area across the incidental learning phase for each word condition with error bars (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

6.4.1.4 Response time

Incorrect responses (1.5% of data), and outliers (responses faster than 250 ms and slower than 2000 ms, 1% of data) were removed before the analysis. The mean response time was calculated for each participant and word condition and averaged for each block.[17] These were submitted to a 6 x 3 repeated-measures ANOVA with the same factors and levels as above. The results revealed a main effect of block, $F_1(2.31, 62.37) = 37.22$, $p < .001$, $\eta_p^2 = .58$, $F_2(5, 385) = 153.52$, $p < .001$, $\eta_p^2 = .67$, with a significant linear trend, $F_1(1, 27) = 65.52$, $p < .001$, $\eta_p^2 = .71$, $F_2(1, 77) = 695.99$, $p < .001$, $\eta_p^2 = .90$, indicating that participants responded faster as they progressed through the incidental learning phase. In addition, results revealed a main effect of word condition, $F_1(1.65, 44.45) = 95.13$, $p < .001$, $\eta_p^2 = .78$, $F_2(2, 154) = 211.79$, $p < .001$, $\eta_p^2 = .73$. Pairwise comparisons (Bonferroni corrected) revealed that participants responded faster when the trials included auditorily FL only (A) than when they included auditory FL, written translation and pictorial information (AWP), but they responded faster in the latter case than when words were presented with auditory FL and written translation (AW), all *p*s < .001. Finally, there was a significant interaction between block and word condition, $F_1(10, 270) = 7.70$, $p < .001$, $\eta_p^2 = .22$, $F_2(10, 770) = 5.87$, $p < .001$, $\eta_p^2 = .07$. The interaction seems due to response times in block 1 not being significantly different for the words presented with auditory and written information (AW) and the words presented with auditory, written and pictorial

[17] Response times in the analyses by item were log transformed to reduce kurtosis.

information (AWP), *ts* < 1 (see Figure 6.5). Response times in all other blocks are significantly different between all three word conditions, all *ps* < .05, except in block 6 for the comparison of words presented auditorily only (A) and presented with auditory, written and pictorial information (AWP) in the analysis by item which showed a strong trend, *p* = .057.

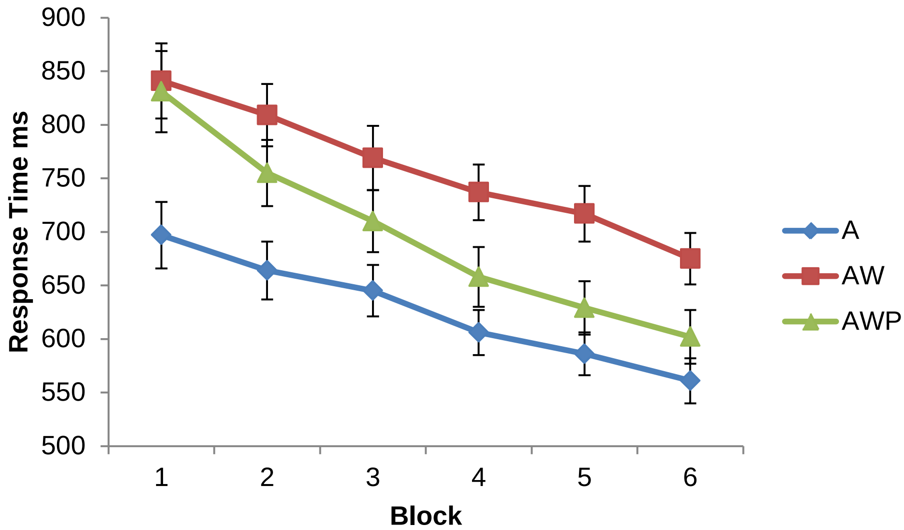

Figure 6.6: Response times in milliseconds for each word condition across the incidental learning phase with error bars (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

6.4.2. Phase 2: Explicit learning

The mean accuracy scores were calculated for each participant for each block and each word type and averaged for each group. As participants had reached criterion in block 2 and therefore did not proceed to block 3, I did not analyze the results of block 3.

For blocks 1 and 2, the mean accuracy scores were submitted to a mixed-design ANOVA with group as a between-subject factor (2 levels: control and experimental groups) and word condition in the incidental learning phase as a within-subject factor (3 levels, auditory FL word only [A], auditory FL word with written translation [AW], and auditory FL word with written translation and picture [AWP]). For the control group, the within-subject factor of word condition was still used even though this group of participants did not take part in the incidental learning phase.

For block 1, the analysis revealed a main effect of group, $F_1(1, 51) = 9.22$, $p < .01$, $\eta_p^2 = .15$, $F_2(1, 77) = 12.11$, $p < .01$, $\eta_p^2 = .14$, indicating that the experimental group was overall more accurate than the control group (M = 59.9 %, SE = 0.8% and M = 56.3% SE = 0.9% respectively). The main effect of word condition was not significant, $F_1(2, 102) = 1.03$, $p = .36$, $\eta_p^2 = .02$, $F_2 < 1$, however, there was a trend for an interaction between word condition and group, $F_1(2, 102) = 2.40$, $p = .096$, $\eta_p^2 = .05$, $F_2(2, 154) = 2.53$, $p = .08$, $\eta_p^2 = .03$ (see Figure 6.7). Simple effects analysis revealed that this was due to a significant effect of word condition in the experimental group, $F_1(2, 102) = 3.49$, $p < .05$, $\eta_p^2 =.07$, $F_2(2, 76) = 3.01$, $p = .05$, $\eta_p^2 =.07$, and not in the control group, Fs < 1. Furthermore, simple effects analysis revealed that participants in the experimental group were more accurate on the words presented with auditory FL and written translation (AW) compared to the control group, $F_1(1, 51) = 7.58$, $p < .01$, $\eta_p^2 = .13$, $F_2(1, 77) = 9.03$, $p < .01$, $\eta_p^2 = .11$. The same result was found for the words presented with auditory FL, written translation and pictorial information (AWP), $F_1(1, 51) = 5.95$, $p < .05$, $\eta_p^2 = .10$, $F_2(1, 77) = 10.17$, $p < .01$, $\eta_p^2 = .12$, but not for the words presented auditorily only (A), *F*s < 1.

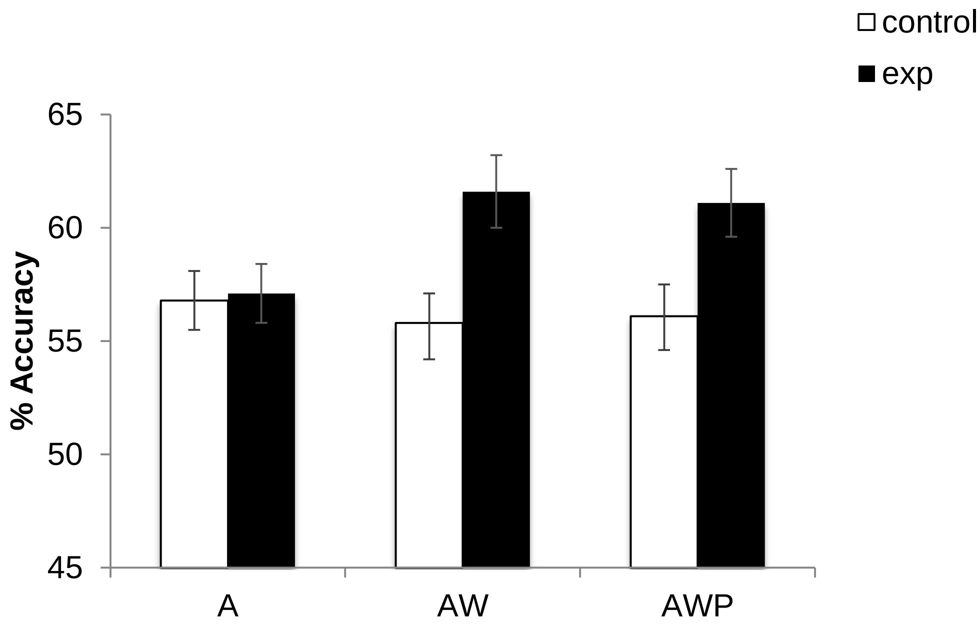

Figure 6.7: Percentage accuracy in block 1 of the translation recognition task for each group and word condition with error bars (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

As only the experimental group had been presented with the three word types, an effect of word type was only expected in the experimental group. Hence the percentages of accuracy in the translation recognition task for the experimental group were submitted to a repeated-measures ANOVA. The results showed a trend for a main effect of word type, $F_1(2, 54) = 2.94$, $p = .06$, $\eta_p^2 = .10$, $F_2(2, 154) = 2.86$, $p = .06$, $\eta_p^2 = .04$. Importantly, contrasts revealed that participants performed significantly better on the words presented with auditory FL and written translation (AW) than on the words presented with auditory FL only (A), $F_1(1, 27) = 5.26$, $p < .05$, $\eta_p^2 = .16$, $F_2(1, 77) = 4.93$, $p < .05$, $\eta_p^2 = .06$. Furthermore, there was a strong trend for participants to perform better on the words presented with audio, written and pictorial information (AWP) than auditorily only (A) $F_1(1, 27) = 4.11$, $p = .05$, $\eta_p^2 = .13$, $F_2(1, 77) = 3.89$, $p = .05$, $\eta_p^2 = .05$. However, whether participants were also presented with the picture did not have any additional impact on their performance on the translation recognition task (AW vs. AWP, *F*s < 1).

The analysis of block 2 also revealed a main effect of group that was significant by item, and a trend by participant, $F_1(1, 51) = 2.90$, $p = .095$, $\eta_p^2 = .05$, $F_2(1, 77) = 13.98$, $p < .001$, $\eta_p^2 = .15$. Neither the main effect of word condition $F_1(2, 102) = 1.22$, $p = .30$, $\eta_p^2 = .02$, $F_2(2, 154) = 1.74$, $p = .18$, $\eta_p^2 = .02$, nor the interaction between group and word condition were significant, *F*s < 1 (see Figure 6.8).

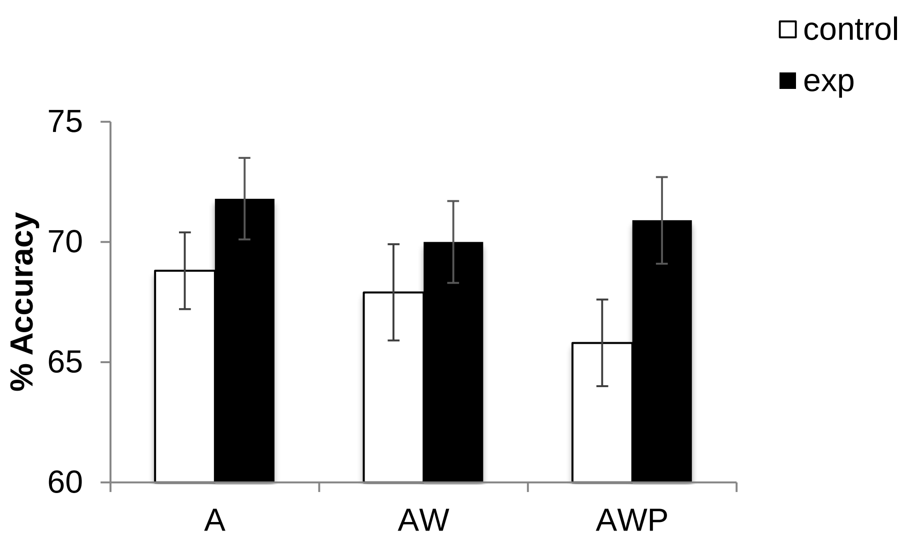

Figure 6.8: Percentage of accuracy in block 2 of the translation recognition task for each group and word condition with error bars (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

6.4.3. Phase 3: Recall and translation recognition

6.4.3.1 Recall

The percentages of accuracy on the recall task were calculated for each participant in each group for each word condition. These were analyzed by participant only, because many items were not recalled by any participants (accuracy = 0%). The percentages of accuracy were submitted to a mixed-design ANOVA with group as a between-subject factor (2 levels: control and experimental groups) and word condition in the incidental learning phase as a within-subject factor (3 levels, A, AW, AWP). Results revealed a significant main effect of group, $F(1, 51) = 8.81$, $p < .001$, $\eta_p^2 = .15$, as well as a trend for a main effect of word condition $F(2, 102) = 2.60$, $p = .08$, $\eta_p^2 = .05$. Importantly, the interaction between group and word condition was significant $F(2, 102) = 3.40$, $p < .05$, $\eta_p^2 = .06$ (see Figure 6.9). This interaction occurred because the effect of word condition was significant in the experimental group, but not in the control group, $F(2, 102) = 6.26$, $p < .01$, $\eta_p^2 = .12$ and $F < 1$ respectively. Simple effects analysis revealed that participants in the experimental group outperformed participants in the control group for the words presented originally with auditory, written and pictorial information (AWP), $F(1, 51) = 14.72$, $p < .001$, $\eta_p^2 = .29$, and there was also a trend for the experimental group to perform better than the control group for the words presented originally with auditory and written information (AW), $F(1, 51) = 3.32$, $p = .07$, $\eta_p^2 = .07$. There was no significant differences between the groups for the words presented auditorily only (A) during the incidental learning phase, $F(1, 51) = 1.25$, $p = .25$, $\eta_p^2 = .03$.

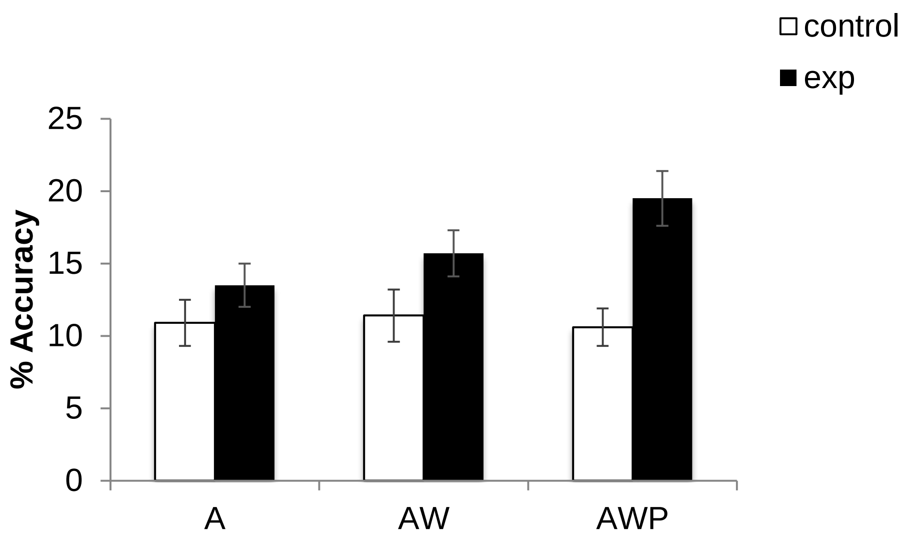

Figure 6.9: Percentage of accuracy in the recall task for each group and word condition in Phase 3 of the experiment (A = auditory FL only; AW = auditory FL with native written language translation; AWP = auditory FL with written NL translation and picture).

In order to explore the effect of exposure to the different word conditions further, a repeated-measures ANOVA was computed for the experimental group only. This revealed a significant main effect of word condition, $F(2, 54) = 5.44$, $p < .01$, $\eta_p^2 = .17$ as well as a significant linear contrast $F(1, 27) = 9.99$, $p < .01$, $\eta_p^2 = .27$. Contrasts revealed that participants recalled significantly more FL words if they had originally been presented auditorily with written translation and pictorial information (AWP) than if they had been presented auditorily with written translation only (AW), $F(1, 27) = 4.40$, $p < .05$, $\eta_p^2 = .14$ or auditorily only (A), $F(1, 27) = 9.99$, $p < .01$, $\eta_p^2 = .27$. However there was no significant difference between the FL words presented auditorily with written translation (AW) and auditorily only (A), $F(1, 27) = 91.46$, $p = .24$, $\eta_p^2 = .05$.

6.4.3.2 Recognition

The mean percentages of accuracy on the recognition test were calculated for each participant in each group for each word condition and submitted to a mixed-ANOVA with group as a between-subject factor (2 levels: control and experimental groups) and word condition in the incidental learning phase as a within-subject factor (3 levels: A, AW, AWP). Results revealed a significant main effect of group by items, $F_2(1, 77) = 4.17$, $p < .05$, $\eta_p^2 = .05$, but not by participants $F_1 < 1$. The main effect of word condition and the interaction between group and word condition were not significant, $F_1(2, 102) = 1.18$, $p = .31$, $\eta_p^2 = .02$, $F_2(2, 154) = 1.21$, $p = .30$ and Fs < 1 respectively (see Figure 6.10).

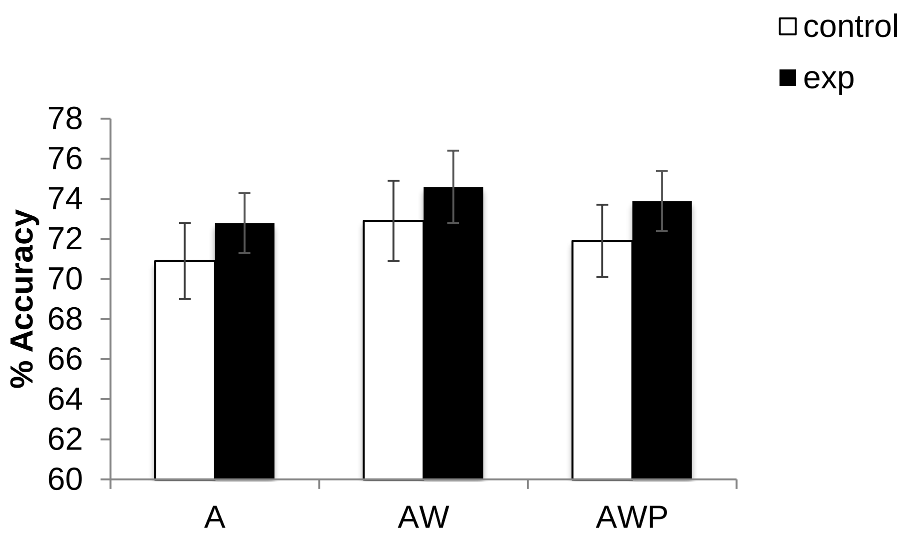

Figure 6.10: Percentage of accuracy in the translation recognition task for each condition and word type in Phase 3 of the experiment (A = auditory FL only; AW = auditory FL with written NL translation; AWP = auditory FL with written NL translation and picture).

6.4.4. The relationship between dwell time (eye-data) and learning

The impact of the dwell time in the image and word areas on the learning of the FL words presented auditorily with written translation (AW) as well as auditorily with written translation and pictorial information (AWP) was investigated using a generalized linear model with a logit. Each word condition was investigated separately.

6.4.4.1 Word type: AWP

For the FL words presented auditorily with written translation and pictorial information (AWP) during the incidental learning phase, both the dwell time in the image and word areas were investigated as potential predictors of the probability of obtaining a correct answer on the recall and recognition measures. The results showed that the dwell time in the image area was a significant predictor of the probability of obtaining a correct answer on both recall and recognition tests in phase 3, and it was a strong trend for the recognition scores in phase 2 for the second block of explicit learning (see Table 6.2). However dwell time in the image area was not a significant predictor for the scores in block 1 of the explicit learning phase, $p = .66$. Furthermore, adding the dwell time in the word area as a predictor did not significantly improve any of the models, *p*s $> .61$.

6.4.4.2 Word type: AW

For the FL words presented auditorily with written translation (AW), analyses revealed that the impact of the dwell time on the word was not a significant predictor of learning, neither in Phase 2 nor in Phase 3, all $ps > .31$.

Table 6.2: Predictors of learning for FL words presented with written translations and pictures (AWP).

Phase	Learning outcome	Predictors	*B (SE)*	*OR*	95% CI	χ^2	Pseudo R^2 H.L.	Pseudo R^2 C.S.	Pseudo R^2 N.
Phase 2	Recognition Block 2					3.44†	0.06	0.12	0.13
		Constant	0.63 (0.15)						
		DT picture	0.45 (0.24)†	1.57	[0.98, 2.53]				
Phase 3	Recall					4.47 *	0.09	0.15	0.18
		Constant	-1.91 (0.26)						
		DT picture	0.83 (0.39) *	2.29	[1.06, 4.96]				
Phase 3	Recognition					7.34**	0.16	0.23	0.29
		Constant	0.65 (0.15)						
		DT picture	0.68 (0.25)**	1.98	[1.21, 3.26]				

Note. DT = Dwell time; *OR* = odds ratios; CI = confidence interval; H.L. = Hosmer-Lemeshow; C.S. = Cox-Snell; N. = Nagelkerde.
The values of pseudo R^2 for the Phase 3 Recall model go up to 0.13, 0.15, 0.23 respectively once the two participants with 0% correct recall are removed.
† p = .06, * $p < .05$, ** $p < .01$

6.5 Discussion

One of the aims of this chapter was to investigate the allocation of attention to the different elements of the multi-modal situation. The eye-tracking data confirmed that both the letter and the written information were processed in all word conditions. However, the time spent looking in the word area differed across conditions. When only the written NL translation was presented on screen, participants spent about half of the trial duration fixating in the word area. However, this decreased significantly when the word condition also included pictorial information: participants spent less time looking in the word area when a picture was first presented. Furthermore, response times were faster in the condition including a picture than in the condition with written NL only. Because the picture onset occurred 300 ms prior to the onset of the written word, the results suggest that the processing of the picture primed the NL orthographic word form. In other words, the processing of the pictorial information activated semantic representations and lexical representations. When participants then searched the word to complete the letter-search task, the pre-activated lexical representations would have facilitated the processing of the written word forms. Finally, as participants spent on average 576 ms fixating in the image area, it is likely that they also returned to the image area once the letter-search task had been solved.

The second reason for using eye-tracking was to investigate the allocation of attention across the duration of the incidental learning phase. Based on the results of Chapter 4, it was expected that more attention would be allocated to the pictorial stimuli at the beginning of the incidental learning phase because of its novelty. Furthermore, since the pictures were irrelevant for the letter-search task, it was expected that as the experiment progressed and the pictures became less salient, less time would be devoted to looking at them. The results confirmed this prediction, as the time spent fixating in the image area decreased significantly across blocks of trials. In fact, the dwell time in the image area for the condition including a picture, decreased by about half between the beginning and the end of the incidental learning task, suggesting that the 'irrelevant pictorial information' did lose some of its appeal. Participants also spent less time looking in the word area for both conditions with auditory and written information, and auditory information only, presumably because they could solve the letter-search task faster as the experiment progressed, due to the repetition of the same words.

In the present experiment FL input was only presented auditorily and in order for the acquisition of FL vocabulary to occur, the FL auditory word forms had to be processed. In other words, if there is evidence of FL vocabulary acquisition through exposure to the multi-modal incidental learning task, I can conclude that the FL auditory word forms have been processed. The results of the explicit learning task completed one day following the incidental learning phase confirmed that exposure to the FL words led to the acquisition of FL vocabulary as participants in the experimental group achieved higher accuracy overall than participants in control group. Furthermore, an incidental learning effect was found in the results of the recall task one week later. Taken together, the results showed that the auditory FL words were processed, even though they were not relevant for the letter-search task.

Interestingly, being exposed to auditory FL word forms only was not enough to lead to an advantage in the explicit learning task. The important aspect that led to gains in vocabulary knowledge was the presence of both FL auditory word form and meaning information. However, having access to the meaning of the words through the NL translation only or both NL translation and pictorial information did not influence accuracy scores differently in the explicit learning task. However, in the recall task one week later, participants achieved significantly higher recall for the words presented with pictures and NL translations compared to the words presented with NL translation only.

In contrast to the results of Chapters 3 to 5, there was no incidental learning advantage in the second block of explicit learning, as the control group performed similarly to the experimental group. Furthermore, there were no significant differences between the groups in the recognition task completed one week later. It would seem that through explicit learning of translation equivalents, the control group quickly caught up with the experimental group.

The last aim of the study was to assess the impact of the pictorial information on incidental learning. The results showed that having pictorial information during the incidental learning phase was beneficial for recall one week later. Whether the pictorial information helps recall in general, or whether it is helpful only in the case of delayed recall is unclear however as the recall task was not completed the day after the incidental learning phase. However, as the goal of language learning is to be able to use the acquired vocabulary after a time delay, the results showed that pictorial information can play an important part in incidental vocabulary learning. Furthermore, the advantage gained from the pictures during the incidental learning phase occurred after only 6 presentations of the pictorial information, even though the pictures were irrelevant for the task, and the meaning of the FL words could already be accessed through the written translations. The results therefore suggest a special role for pictorial information. This was supported by the finding that the dwell time on the pictures during the incidental learning phase was a significant predictor of recall scores, thereby highlighting the usefulness of pictorial information for vocabulary learning. Although results did not show an advantage of having been exposed to pictorial information on the delayed recognition test (possibly due to a ceiling effect), the dwell time on the picture was also a predictor of the recognition scores one week later. Furthermore, there was a strong trend for the dwell time on the picture to predict the recognition scores in the second block of explicit learning the day following the incidental learning phase. Taken together, the results clearly show that having access to pictorial information during an incidental learning task is beneficial for FL vocabulary learning.

In this experiment, as participants also had access to the written NL translation during the incidental learning phase, it is not possible to evaluate the impact of the picture alone. It is possible that it was the combination of written and pictorial information that was beneficial. However, the time spent looking in the word area was not a predictor of learning, and therefore this suggests that the picture played a crucial role.

How could pictures support vocabulary learning? In memory research, superior recall for pictures has generally been found and many potential explanations have been put forward. For example, dual-coding theory states that this is because pictorial information can be encoded in memory both as nonverbal information and as verbal information, by generating a lexical label (Paivio & Csapo, 1973). The cascading activation model of speech production supports the idea of automatic activation of lexical information during picture processing, even when the pictures are irrelevant for a task (see Kuipers & La Heij, 2009; Meyer & Damian 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005). In the experiment presented in this chapter, both pictorial and lexical information were available during the incidental learning phase, therefore the encoding of both types of information would have been encouraged. However, how is the processing of the picture, notwithstanding the lexical label, beneficial for vocabulary acquisition? The semantic–sensory model posits that access to semantic representations is faster and more direct for pictures than for words, and that as pictures are more distinctive and varied in their mnemonic features they are more easily recalled (Nelson et al., 1977). In Chapter 5, it was found that participants were able to categorize pictures faster than words, which supports a direct semantic route for pictures as put forward by Nelson et al., (1977). Although the results of the current experiment do not allow to pinpoint why pictures are better recalled than words (nor was it the aim of this chapter), both explanations, i.e., dual-coding and direct semantic access are plausible.

In the field of FL learning, the findings with regards to picture superiority effects are not as clear as those obtained in memory research. For instance, Lotto and de Groot (1998) found no superiority effect for pictures in explicit learning of FL words. In fact, their results showed better FL word learning when their participants were presented with FL and NL word pairs during a learning phase than when they were presented with FL words in combination with pictures. Similarly, Carpenter and Olson (2011) found no advantage for using pictures and FL word pairs during a learning phase, which they explained by participant's overconfidence in their ability to recall FL words from pictures. Once they eliminated this bias, however, a picture superiority effect did emerge. In spite of the FL learning context of these two studies, an important factor that may have contributed to the lack of picture superiority effect is the learning paradigm used, i.e., explicit learning in the above two studies, compared to an incidental learning paradigm in the current experiment.

Furthermore, the benefits of having access to pictures in the current study emerged one week following the learning phase, and neither of the two studies mentioned above included a delayed test. Therefore, the results found here are important in showing that in contrast to what has been found in previous studies of FL word learning, there is a benefit of having access to pictures for FL vocabulary learning. It remains to be seen whether this is the case only for incidental learning paradigms, and whether learning benefits need more time to emerge.

6.6 Conclusion

The results of the current study revealed an important role for pictorial information in FL vocabulary learning. Not only did the pictures alter the viewing behavior during the incidental learning phase, but it also helped participants retrieve the correct translations of FL words one week later. The results are in contrast to those of prior studies of explicit FL vocabulary learning because a picture superiority effect was found. This highlights the importance of using delayed testing before concluding on the usefulness of a training method in FL teaching and learning. The FL words were presented auditorily in the current study, and combining this with both NL translations and pictures in an incidental learning situation, was an effective method of learning FL vocabulary.

Chapter 7 : General discussion

In the introduction chapter (Chapter 1), I outlined the 5 main research questions that this thesis set out to investigate (see Section 1.5.1). In the chapters that followed (Chapters 2 to 6), I reported the results of five experiments that were designed to answer these questions. In the current chapter, I will return to each of the main research questions in turn, and discuss these, drawing from the results obtained in each experimental chapters. A brief section on limitations and future studies will ensue before concluding the thesis.

7.1 Research question 1: Can people learn FL vocabulary incidentally through exposure to multi-modal situations?

This thesis set out to investigate whether people can learn FL vocabulary incidentally through exposure to multi-modal situations. Prior research on the incidental acquisition of FL vocabulary had been conducted using films with subtitles, however, as discussed in Chapter 1 (Section 1.3.4), the findings were not convincing because of methodological issues. For example, few items were used on the vocabulary tests (10 items in d'Ydewalle and Van de Poel, 1999) and the analyses were not adequate (Koolstra & Beentjes, 1999), or not reported (d'Ydewalle & Pavakanun, 1995). Furthermore, the incidental learning gains were very small. Therefore the aim of the first experiment (see Chapter 2) was to provide evidence of incidental acquisition of vocabulary through exposure to a multi-modal situation. In order to address this question, the experiment reported in Chapter 2 used a film with subtitles as a multi-modal situation. Participants watched the film extract in one of the subtitling conditions (standard, intralingual or reversed), and following this, their incidental vocabulary acquisition was measured using a translation recognition test. Unfortunately, all groups of participants performed similarly in the vocabulary test, even those in a control group that had not seen the film. It was concluded that either no vocabulary acquisition occurred, or the vocabulary learning measure used in the study was not sensitive enough to detect the small vocabulary gains likely to occur.

Following this disappointing result, and in order to address the question of incidental vocabulary acquisition, the multi-modal situation was simplified in Chapter 3 such that participants were presented with auditory and written FL word forms along with simple line drawings depicting the meaning of the words. Furthermore, a new measure based on the savings paradigm was used to assess incidental vocabulary acquisition. The results reported in Chapter 3 clearly showed incidental learning effects, as participants performed better in an explicit learning phase for words they had been exposed to during the incidental learning phase than for completely new words. This incidental learning effect was found for a group of participants completing the explicit learning phase immediately following the incidental learning phase, as well as for a group of participants with a one-day delay in between each phase. Furthermore, an incidental learning effect remained one week later following further explicit learning.

Following this, a similar incidental learning paradigm coupled with the same sensitive measure of early learning effects was used in the subsequent experiments reported in Chapters 4 to 6. In Chapter 4, an incidental learning effect was found even though the frequency of exposure to the words during the incidental learning phase varied (see Section 7.2 for further discussion of frequency effect) such that overall participants had less exposure to the words than in Chapter 3. Further evidence for an incidental learning effect was found in the experiment reported in Chapter 5. In this experiment two groups of participants were initially exposed to FL words and pictures in an incidental learning phase, as in Chapters 3 and 4, although fewer words were used overall. Following this, they either immediately completed a semantic categorization task and explicit learning task (same-day group), or returned the next day to complete both tasks. Importantly, for both groups of participants, the results revealed once more, a clear incidental learning effect.

Finally, in Chapter 6, participants completed incidental and explicit learning phases with one day in between, and returned one week later to complete a recall and a translation recognition test. Contrary to Chapter 3, there was no feedback provided when participants returned one-week later, and therefore no further learning could occur. An additional difference between this experiment and the experiments reported in

Chapters 3 to 5 is that the amount of multimodal information was manipulated during the letter search task. Although the same line drawings were presented as in Chapters 3 to 5, FL words were presented in auditory form only, and NL translations were presented concurrently in writing. For some of the FL words, there were no pictures provided; only auditory FL word forms and NL translations were presented, whilst some of the FL words were presented auditorily only with no meaning information. In addition, each word was presented 6 times only in Chapter 6, compared to 8 times in Chapter 3. Notwithstanding these differences during the incidental learning phase, the results of Chapter 6 showed an incidental learning advantage the next day during the explicit learning phase similarly to Chapter 3. However, in contrast to the results of Chapter 3, after a one-week delay, there was no incidental learning advantage in the translation recognition test. However, as mentioned earlier, the task used in Chapter 6 was slightly different to the one used in Chapter 3, as there was no feedback provided and therefore participants were not able to learn the words any further. This may explain the difference in the results between these two experiments after a one-week delay. In addition, the FL words were presented auditorily only during the incidental learning phase for the group of participants in the experiment reported in Chapter 6, whilst for Chapters 3 to 5, both auditory and written FL word forms were presented. Having access to the orthographic form of the FL might have boosted the learning of the auditory word forms found in Chapters 3 to 5 (see Bird & Williams, 2002; Hu, 2008; Ricketts, Bishop, & Nation, 2009; Rosenthal & Ehri, 2008). Interestingly, even though in

Chapter 6 no incidental learning effect was found in the translation recognition test after a one-week delay, a significant incidental learning effect emerged in the recall test. To summarize the findings of Chapter 6, an incidental learning advantage was found in the explicit learning phase completed one day after the incidental learning phase as well as in a recall test completed one-week later. A summary of the incidental learning effects reported in the thesis is presented in Figure 7.1.

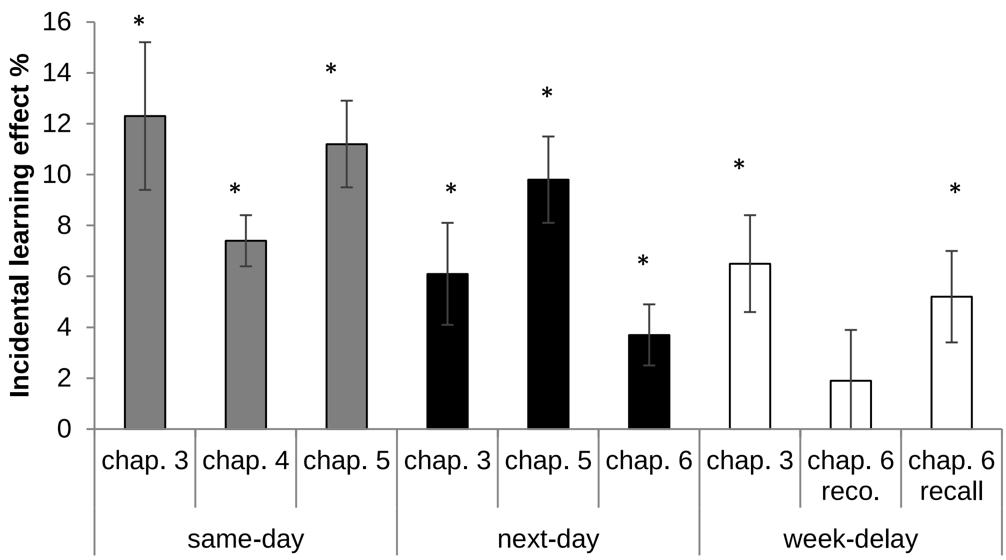

Figure 7.1: Incidental learning effect. Difference score between old and new words for the first block of trials in the explicit learning phase in Chapters 3 to 5, and between experimental and control group in Chapter 6 for the translation recognition and recall tests. The exposure to the words in the incidental learning phase was not the same in Chapter 6 (see text for further details).

Taken together, the results of Chapters 3 to 6 show convincingly that complete beginners in a FL can acquire FL vocabulary from exposure to a simple multi-modal situation. Participants were able to retrieve NL meanings when hearing FL auditory word forms, which shows higher incidental learning effects than the form recognition effect reported in Gullberg et al. (2013). However, as the same simple multi-modal situation was used in Chapters 3 to 6, it remains to be seen whether the same incidental learning effect could be found in more complex multi-modal situations. In Chapter 2, a film with subtitles was used as a multi-modal situation, and no vocabulary acquisition was found. Maybe a longer exposure is needed in a more complex multi-modal situation, and therefore a longitudinal study may be more appropriate. Furthermore, the vocabulary test used in Chapter 2 may not have been sensitive enough. Because Chapters 3 to 6 showed that a more sensitive measure of vocabulary acquisition can successfully reveal FL acquisition, future studies should therefore use a similar measure to assess the incidental acquisition of vocabulary from exposure to more complex multi-modal situations. For example, a future study could use FL sentences presented auditorily and in writing, with a simple static picture or simple animation from which the meaning of FL words could be derived. Alternatively, the incidental acquisition of FL vocabulary through watching films with subtitles could be investigated using the more sensitive measuring instrument developed in this thesis. As mentioned above, the amount of exposure to the FL words during the incidental learning phase is an important factor that may have contributed to the incidental acquisition of FL vocabulary reported in this thesis. The next section addresses this question specifically.

7.2 Research question 2: Are there frequency of exposure effects in incidental FL acquisition in the context of multi-modal situation, and how many exposures are needed to detect incidental learning?

As discussed in the previous section, clear incidental learning effects emerged in the experiments reported in Chapters 3 to 6, indicating that people can learn FL vocabulary incidentally through exposure to a simple multi-modal situation. However, two important questions remained. Firstly, how many exposures are needed to detect incidental learning, and secondly, are there frequency of exposures effects in the context of a multi-modal situation? In this section I will address each question in turn.

The minimum number of exposures necessary for incidental FL vocabulary acquisition to occur has received little attention in the literature. As was discussed in Chapter 4, one study reported incidental learning effects of FL word meanings from 2 exposures in a reading situation (Rott, 1999), whilst another reported incidental learning of FL word forms from 5 exposures in a reading-while-listening situation (Webb et al., 2013). Although many other studies have investigated frequency of exposure effects through reading (Brown et al., 2008; Hulstjin et al., 1996; Kweon & Kim, 2008; Pellicer-Sánchez & Schmitt, 2010; Vidal, 2011), listening (Brown et al., 2008; Vidal, 2011), reading-while-listening (Brown et al., 2008; Horst et al., 1998; Webb et al., 2013), and multi-modal situations (d'Ydewalle & Pavakanun, 1995, 1997; d'Ydewalle and Van de Poel, 1999; Koolstra & Beentjes, 1999; Gullberg et al., 2012) these unfortunately have not investigated the number of exposures required for incidental vocabulary acquisition to occur. Therefore, the experiment reported in Chapter 4 was designed specifically to investigate this question. The frequency of exposure during the incidental learning phase was varied between 2, 4, 6 and 8 exposures and participants completed an explicit learning phase immediately following the incidental learning phase where they learnt to recognize translation equivalents between auditory FL words and written NL translations, half of which had occurred in the incidental learning phase. Results revealed significant incidental learning effects from 2 exposures in the incidental learning phase, which is in line with the results reported in the Rott study mentioned above. Although the results of the Rott study are not directly comparable to those reported in Chapter 4 as there were many differences between the two studies (e.g., design, degree of fluency of participants, etc.), both studies reported incidental learning effects from 2 exposures. However, does the incidental learning effect increase with the frequency of exposure?

As was mentioned earlier, many incidental learning studies investigated the effects of the frequency of exposure. Although these studies varied considerably in terms of the type of design used and analyses conducted most of them reported significant frequency of exposure effects (see Table 4.1). Importantly, the only study investigating frequency of exposure effects in the context of a multi-modal situation is Gullberg et al. (2012). This study found significantly higher recognition for words that occurred frequently (8 or 16 exposures) compared to infrequently (2 or 4 exposures). Similarly to the Gullberg study, the results reported in Chapter 4 revealed a significantly larger incidental learning effect for the words occurring 8 times in the incidental learning phase compared to words occurring only twice. However, no overall frequency effect was found across frequency bands (2, 4, 6 and 8). A possible explanation for the lack of overall frequency effect was put forward in Chapter 4. I suggested that the impact of each exposure to the words may not be constant across frequency bands. In other words, it is possible that learners extracted more meaning information from the first few exposures to the words than from later exposures, and that therefore additional exposures led to smaller learning gains. This idea was confirmed by calculating a normalized frequency of exposure effect (the incidental learning effect divided by the number of exposures in each frequency band), which revealed a significant negative linear trend. However, it is possible that this effect was due to the task employed during the incidental learning phase because participants only had to search the written word to complete the letter-search task. Therefore, they might have paid less attention to the pictorial stimuli as they progressed through the experiment. The pattern of eye-tacking data obtained in the experiment reported in Chapter 6 confirmed that indeed participants spent less time fixating on the pictorial information at the end of the experiment than at the beginning (see Section 7.3 for further discussion of attention allocation in multi-modal situations). Thus although I found a negative linear trend in the normalized frequency of exposure effect in the context of a multi-modal situation, this might not be the case in other contexts.

The aim of Chapter 4 was to specifically investigate frequency of exposure effects. However Chapters 3, 5 and 6 all used slightly different frequencies of exposure. In Chapter 3, participants were exposed to each FL word 8 times during the incidental learning phase. Although the data collected during the experiment reported in Chapter 4 already included a frequency band for 8 exposures, it is interesting to note that the results for the same frequency band are similar for both experiments (see Figure 7.2 and 7.3) despite their coming from different groups of participants. Furthermore, in Chapter 3, there was a group of participants who completed the explicit learning phase one day following the incidental learning phase. The results from this group can be compared to the results of participants who were exposed to words 6 times (see Chapter 6) and 10 times (see Chapter 5) and completed both phases with a one-day delay in between. Finally, both Chapters 3 and 6 included a group of participants who returned after a one-week delay, and therefore provided information about the frequency of exposure effects in incidental learning for 8 and 6 exposures respectively. Although the effects of frequency of exposure were not compared statistically across chapters, what seems to emerge from the data presented in Figure 7.2 is that there are similar positive linear trends in the incidental learning effects for data collected the same-day, the next-day as well as one-week later. However, the negative linear trend observed in the normalized incidental learning effect seems to be restricted to the data collected on the same-day as the incidental learning phase (see Figure 7.3). It remains to be seen whether a negative relationship would emerge if data were collected for the lower frequency bands (2 and 4 exposures) in the groups of participants with a one-day and one-week delay. Furthermore, there seemed to be more gains in vocabulary knowledge between 8 and 10 frequency of exposure for participants in the one-day delay group than for participants who completed both incidental and explicit learning on the same-day. In order to further investigate the effects of frequency of exposure on incidental vocabulary acquisition, it would therefore be important to conduct a similar experiment as the one reported in Chapter 4 with a full range of frequency bands and include groups of participants both with one-day and one-week delay between the incidental and explicit learning phases.

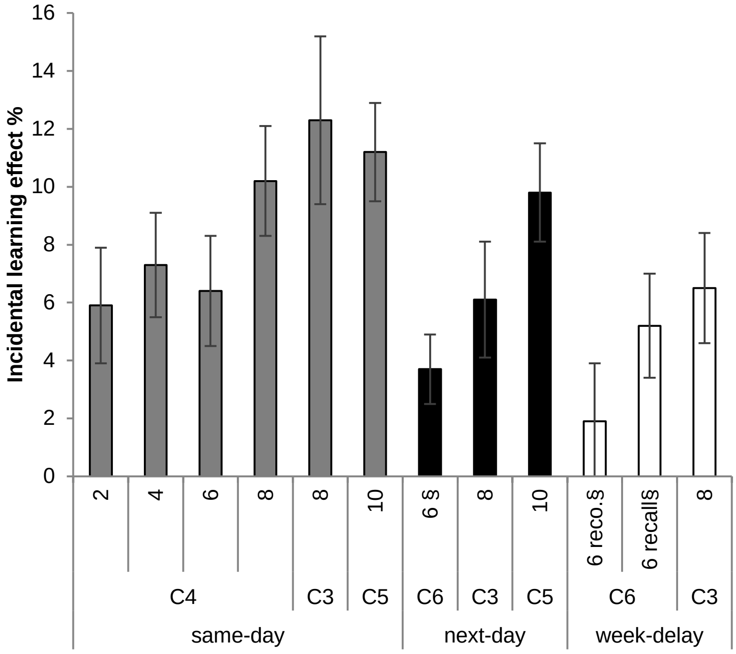

Figure 7.2: Incidental learning effect for each frequency of exposure. Difference score between old and new words for the first block of trials in the explicit learning phase in Chapters 3 to 5, and between experimental and control group in Chapter 6 for the translation recognition and recall tests (§ denotes the different exposure to the words in the incidental learning phase in Chapter 6; see text for further details).

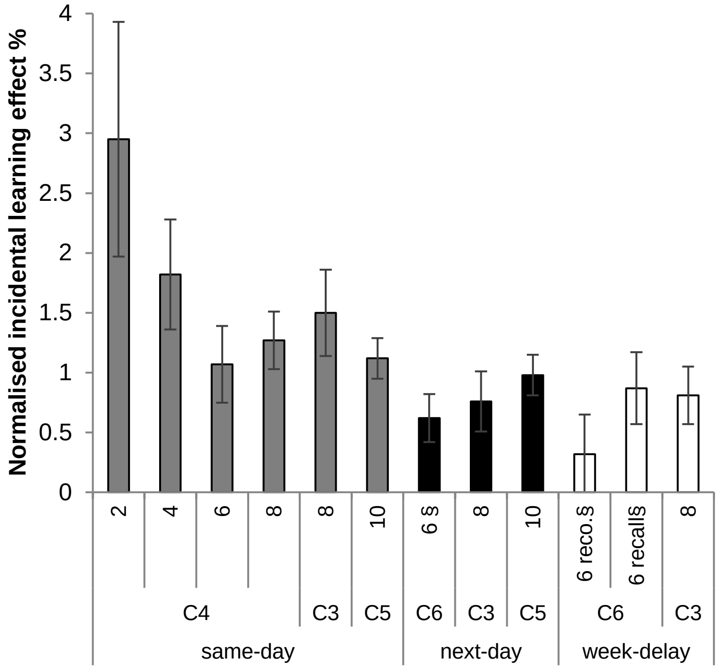

Figure 7.3: Normalized incidental learning effect for each frequency of exposure. This was calculated by dividing the incidental learning effect presented in Figure 1.2 by the number of exposure in each frequency band (§ denotes the different exposure to the words in the incidental learning phase in Chapter 6; see text for further details).

To summarize this Section, the results of Chapter 4 indicated that incidental acquisition of FL vocabulary can occur from as little as two exposures to multi-modal stimuli. Furthermore, a frequency of exposure effect was found, as more incidental learning took place for words with 8 occurrences than words with 2 occurrences during the incidental learning phase. However when taking all the frequency bands into account, no overall frequency effect was found because the impact of each additional exposure was not constant across the frequency bands. This was confirmed by the negative linear trend obtained in the normalized incidental learning effect. As explained earlier, this may have been due to participants paying more attention to the pictorial element of the multi-modal situation at the beginning of the experiment compared to at the end of the experiment. The next section will address the allocation of attention to the different elements of the multi-modal situation in more detail.

7.3 Research question 3: How do people allocate their attention to verbal and visual aspects of multi-modal situations?

The presentation of verbal and pictorial information occurs concurrently in a multi-modal situation. Therefore, people can allocate their attention to these different elements, and presumably integrate this information for comprehension and learning to happen. In the case of a film with subtitles (see Chapter 2), there is input through the visual modality, i.e., the pictorial information and the subtitles, whilst the soundtrack provides auditory information. Similarly, in a less complex multi-modal situation as was used in Chapters 3 to 6, participants were exposed to verbal information both auditorily (auditory FL words) and visually (written FL or NL words) whilst pictorial information (simple line drawings) was presented concurrently. In both types of multi-modal situation, information from different modalities had to be integrated in order for incidental FL vocabulary acquisition to occur. Therefore, an important question is: How do people allocate their attention to verbal and visual aspects of multi-modal situations? In this section I will first summarize the findings from a subtitled film situation (Chapter 2) before discussing the results from a less complex multi-modal situation with static pictures and single visual and auditory word forms (Chapters 3 to 6).

In Chapter 2, participants' eye-movements were recorded as they watched a film with subtitles. The results showed that participants attended to subtitles and images to some extent, independently of the subtitling condition. The goal of the viewer was to understand the film, and therefore they used all the available information to that end. This included the reading of the subtitles when the soundtrack was in a FL (standard and intralingual subtitling). When the soundtrack was in a NL however, there was no need to read the subtitles, as participants already understood the film. However, as was reported in Chapter 2, participants still read some of the FL subtitles. Why did participants read FL subtitles when they were unnecessary to achieve their goal of understanding the film? As was discussed in Chapter 2 (Section 2.5), the dynamic nature of the subtitles, as well as the fact that they contained words, would have made them very salient, and therefore likely to generate a saccade towards them. This reasoning is in line with visual attention deployment research showing that motion and text strongly predicts gaze allocation in both static and dynamic scene viewing, independently of task (Cerf et al., 2009; Mital et al., 2010). More evidence for this comes from Smith and Mittal (2013) who showed that across a group of participants free-viewing a dynamic scene (a video) gaze was clustering on the same salient features, in this case people and flicker. However, when participants were asked to perform a specific task no such gaze synchronization was found. Interestingly, even when the task directed participants' gaze allocation, the initial eye-movements, as well as the eye-movements following the completion of the task, returned to the salient features. The idea that the allocation of attention during a film with subtitles is driven both by the goal of the viewer, as well as salient features is therefore in line with the results of this research. In case of a film with subtitles, the goal of the viewer is to understand the film and as such, all the available information, i.e., images and subtitles, was used to that effect. Furthermore, as explained above, salient features such as motion and text attract attention regardless of the intentions of the viewer, and this made the reading of the subtitles in all subtitling conditions compelling, even when it was unnecessary. This can explain the viewing patterns in dynamic multi-modal situations but how is attention allocated during a simpler static multi-modal situation?

In the case of the simple static multi-modal situation used in Chapters 3 to 6, the goal of the participant was to complete a letter-search task. Therefore, it was expected that participants would attend to the written words, as this was necessary to complete the task. In contrast, it was not necessary to attend to either pictures or auditory words, as they were both irrelevant for the task. Importantly, FL word learning could only occur if both of these two elements were processed. The results of Chapters 3 to 6 showed that participants learnt FL word meanings incidentally through exposures to the pictures and FL word forms. Specifically, in Chapters 3 to 5, participants could only learn FL word meanings through the processing of the pictures. As significant incidental learning effects were found, this confirmed that participants attended to the pictures even though they were irrelevant for the task. Furthermore, Chapter 6 provided direct evidence of picture processing, as eye-movements were recorded whilst participants completed the letter-search task. The results showed that participants still fixated on the pictures, even though they were irrelevant for the task. Furthermore, the meaning information contained in the pictures was redundant, as meaning was already provided by the written NL words. As well as explaining the processing of subtitles in film, the saliency theory mentioned above can also be applied to the processing of the pictures during the letter-search task. In Chapter 6 specifically, the appearance of the pictures preceded the appearance of the written words, and therefore the pictures were the only element on the screen for the first 300 ms. Their appearance on the screen, on their own, would therefore have made them very salient and likely to attract attention and hence participants' gaze. Importantly, the data presented in Chapter 6 revealed that participants looked at the picture on average for 576 ms, which indicates that they either continued to look at the pictures past the written word onset, or that they returned to the image area after solving the letter-search task. The second explanation is more plausible as the task would have directed participants' attention to the word area as soon as the word appeared. Furthermore, returning to the salient features on the screen (the picture) once the task was completed is similar to the results obtained in Smith and Mittal (2013) mentioned earlier.

The results of Chapter 6 also revealed that the dwell time on the images decreased throughout the incidental learning phase. Thus participants looked more at the pictures at the beginning of the experiment than at the end of the experiment, which suggests that the effect of saliency decreased as participants' exposure to the pictures increased. In other words, participants were more interested in looking at the pictures when they were novel, however, as they appeared repeatedly in each block of trials, their 'novelty appeal' decreased. Torralba, Oliva, Castelhano, and Henderson (2006) put forward a model of eye-movement and attention guidance in real-world scene perception that can be useful in explaining this finding. Their model uses both local-features (saliency) and global features (contextual constraints) to predict eye-movements. In contrast to models based on saliency maps mentioned earlier, which are based purely on bottom-up influences on attentional and eye-movement guidance, this model also takes into account top-down influences such as the task (or goal of the viewer) and prior scene perception. The model was developed in the context of a task requiring participants to look for pedestrians in a real-world scene. Because of this task, participants would use the global features of the scene to quickly orient their attention towards the parts of the scene where pedestrians were likely to be found, for example, the pavement. Furthermore, prior scenes were used to predict the location of the pedestrians in the current scene. Applying this model to the data from Chapter 6 would suggest that the saliency of the pictures guided participants' attention at first and as they progressed through the experiment, the contextual constraint of the experiment (the repetition of the pictures) reduced the influence of saliency. Furthermore, repetition effects would also have contributed to the decrease in dwell time on the pictures across the experiment, because prior picture viewing would have reduced the time necessary to recognize the same pictures on subsequent trials.

So far, I have discussed the allocation of attention to the pictorial element, as well as to the written verbal element of the multi-modal situations. In both films with subtitles and the simpler multi-modal situation used in this thesis, there was also auditory verbal information provided. Unfortunately, it was not possible within the current experimental design to directly assess how much of the auditory stream was processed. In the film with subtitles situation, if learning of FL words had occurred in the conditions with FL soundtrack, this would have given some indication that the FL words included in the soundtrack were processed. However, no learning was detected in any of the subtitling conditions. Furthermore, although the results of Chapters 3 to 5 suggest that the auditory information was processed, this could not be ascertained directly. In the explicit learning phase, participants had to indicate whether written NL words were the correct translations of auditory FL words. Because participants performed better with words presented during the incidental learning phase this suggests that they linked *auditory* FL word forms to meaning representations. However, as both auditory and written FL word forms were presented during the incidental learning phase, it is also possible that they linked the *written* FL word forms to their meaning representations. Therefore, as both auditory and written FL word forms were presented during the incidental learning phase, it is unclear which of these representations was responsible for the incidental learning effect. In contrast, during the incidental learning phase in Chapter 6, participants were exposed to FL word forms only through the auditory modality, as the written word presented concurrently was in the NL. The data revealed an incidental learning effect, thus indicating that the FL auditory word forms were processed. This result is consistent with the results obtained in pseudoword form learning in the Saffran task (see Saffran et al., 1997). Both experiments are similar in the sense that the words presented in the auditory modality were irrelevant for the task participants were engaged in. However, in the Saffran task, the pseudowords were repeated 300 times, and participants incidentally learnt the word forms only, whereas in the experiments presented in this thesis, the frequency of exposure varied between 2 and 10 exposures only, and participants incidentally learnt both *form* and *meaning* of the FL words.

In this Section, I discussed the processing of each element of the multi-modal situation. Direct eye-tracking evidence showed that both written and pictorial elements were processed. Furthermore, auditory FL word forms were also processed because FL word learning occurred when only the auditory FL word form was presented (concurrently with meaning). In Chapter 6, FL word meaning could be accessed both from the NL translation and/or the pictorial information, whilst it was only possible for participants to acquire FL word meaning in Chapters 3 to 5 through the processing of the pictures. As discussed in Chapter 1 (Section 1.3.3) the pictorial element of a multi-modal situation is what sets it apart from other incidental learning situations such as reading or reading-while-listening. Therefore, in the next Section, I will discuss the specific role of pictorial information in FL word learning.

7.4 Research question 4: What is the importance of pictorial information in multi-modal incidental vocabulary acquisition?

The results of Chapters 3 to 5 showed that FL word meaning was learnt incidentally through exposure to pictorial information. In contrast, in Chapter 6, the design of the experiment was such that in one word condition, participants could learn the meaning of FL words through the written NL translations. Furthermore, in another condition, both translation equivalents and pictures provided participants with FL word meaning. Therefore it was possible to assess whether having access to pictorial information facilitated incidental FL word learning. As discussed in Chapter 6 (Section 6.5), the results showed that there was a special role for pictorial information in FL word learning. Specifically, participants recalled significantly more FL word meanings one week following the incidental learning phase if they had originally been presented along with pictures. Furthermore, the dwell time on the pictures was a significant predictor of the probability of obtaining correct responses on both recognition and recall tests. Furthermore, the picture superiority effect obtained in the experiment reported in Chapter 6 is in contrast to prior research on explicit FL word learning, as neither Lotto and de Groot (1998) nor Carpenter and Olson (2011) found a picture superiority effect (for further discussion see Chapter 6, Section 6.5). It is important to mention that in both these studies, learning and test phases occurred in the same session. In contrast, in the study presented in Chapter 6, a picture superiority effect was found after a one-week delay, however, immediate recall was not measured. It is therefore difficult to assess whether the different patterns of findings are the results of the learning paradigm (explicit vs. incidental) used, or that they are the result of the time of testing (immediate vs. delayed). Paivio and Csapo (1973) found superior immediate recall for pictures following an incidental learning phase, which suggests that it is the learning paradigm that explains the different pattern of results in FL word learning. Thus, there may be a special role for pictorial information in incidental learning that is not mirrored in explicit learning. A possible explanation for this was put forward by Carpenter and Olson (2011). They suggested that when learning FL words with pictures, participants are overconfident in their ability to recall the FL words and this impairs their processing of the associations between pictures and FL words during an explicit learning phase. Furthermore, in a follow-up experiment, they showed that once participants had been warned not to be overconfident, a picture superiority did emerge. The nature of the incidental learning task used in this thesis would have prevented participants from being overconfident in their ability to recall FL words, as they had not been asked to learn them, nor were they asked to look at the pictures. Furthermore, because an explicit learning paradigm was used in both Lotto and de Groot (1998) and Carpenter and Olson (2011), it was not possible to control for the strategies participants used to remember the words. In contrast, using an incidental learning paradigm ensures that any learning effects found are not contaminated by the participants' learning strategies. Thus, by giving participants a specific task in the incidental learning phase that does not involve learning per se, it is possible to control, to a certain extent, the way in which participants process the stimuli (see Craik & Lockhart, 1972). Therefore the learning paradigm used in Lotto and de Groot (1998) and Carpenter and Olson (2011) may account for their lack of picture superiority effect. However, as recall was not assessed immediately after the incidental learning phase in Chapter 6, it is not possible to rule out the test time as a possible explanation for the inconsistent findings in the literature.

To summarize, the picture superiority effect found in the experiment reported in Chapter 6 is in contrast to prior FL learning research (i.e., Lotto & de Groot, 1998; Carpenter & Olson, 2011). As discussed above, differences in the study designs might explain the contrasting results. It is therefore important that the effects of delayed vs. immediate recall, as well as incidental vs. explicit learning are explored further in future studies.

Paivio and Csapo's (1973) dual-coding theory provides an explanation for the superior recall for pictures. Their theory is based on the idea that both verbal and non-verbal information can be used during the encoding and retrieval of pictures. Furthermore, Nelson et al. (1977) suggested that pictures benefit from more diverse features compared to verbal labels, which facilitates encoding and increases the chances of retrieval. In addition, the processing of pictures has been shown to activate semantic representations first (e.g. Bajo, 1988; Smith & Magee 1980), which would encourage a deeper level of processing and hence lead to better encoding (see Craik & Lockhart, 1972). Crucially, the activation of semantic representations during picture processing may lead to the creation of direct form-meaning links, which may facilitate retrieval. The lexical-semantic connections between FL word forms and semantic representations will be discussed in more detail in the next section in relation with the current models of FL word learning.

7.5 Research question 5: What are the implications of multi-modal incidental learning for current models of FL vocabulary learning?

This thesis focused on the processing of FL words in multi-modal situations, and their potential impact on incidental FL word learning. As was reviewed in detail in Chapter 1, most research on incidental FL word learning has been conducted in the context of reading, listening, or reading-while-listening. The main difference between these situations and the multi-modal situation used in the experiments presented in this thesis, is the addition of pictorial information. An advantage of using pictorial elements during incidental learning situations is that even complete beginners in a FL can learn the meaning of FL words. This is not the case for the other incidental learning situations mentioned above because in those situations, learners must use the language context to derive the meaning of the words. Importantly, another potential benefit of using pictorial information is that it leads to superior recall (see Section 7.4). One of the potential reasons for this is that the processing of pictorial information promotes the creation of direct links between FL word forms and semantic representations. However, this would go against the predictions of current models of FL learning. Both the RHM and the BIA-d (see Chapter 1, Sections 1.4.2 and 1.4.3 for a review of these models) assume that learners initially use NL translation equivalents to access FL word meaning. Thus initially, FL word forms are linked to their NL translations via lexical links, and as such, semantic information about the FL words can be accessed via the NL word representations. As was discussed in Chapter 1 (Sections 1.4) and Chapter 5 (Section 5.1), both models were designed to account for FL word learning in an explicit learning situation (typically a classroom) where learners are presented with new FL words along with their translation equivalents. Therefore, neither model accounts for learning FL words using pictorial information. One of the aims of the current thesis was to assess whether these models are suitable to explain the learning that accrues from a multi-modal incidental learning situation that includes pictorial information.

In incidental learning, participants are not asked to learn FL words unlike an explicit learning situation. Therefore, they are not likely to use encoding strategies. Rather, learning occurs by repeated co-occurrences of FL word forms and meaning whilst participants are engaged in an unrelated task. The pairing of FL word forms with meanings will lead to associations between the two in what can be described as a form of Hebbian learning (see Grainger et al., 2010; Pulvermüller, 1999). In the incidental learning situation used in the current thesis, participants were exposed to FL word forms whilst concurrently being presented with pictures. Research has shown that picture processing first activates semantic representations (e.g. Bajo, 1988; Smith & Magee 1980), which would support the form-meaning link hypothesis. As the processing of the pictures would have activated semantic representations whilst participants processed FL word forms, this concurrent activation should have led to the creation of form-meaning connections. However, an alternative explanation (see also Chapter 3, Section 3.5) is that the processing of the pictures, whilst activating the semantic representation, would also have activated the NL lexical representations, leading to connections being created between the FL and the NL word representations. The activation of semantic representations during picture processing has been shown to cascade down to lexical representations (see Bles & Jansma, 2008; Kuipers & La Heij, 2009; Meyer & Damian 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005), which would support the creation of lexical links. However, this lexical link explanation cannot account on its own for the superior recall of FL words presented with pictures compared to FL words presented with NL translations. In other words, there must be something different happening when learning FL words with pictures.

In order to answer this question, a cross-modal priming task was inserted between the incidental and explicit learning phases in the experiment reported in Chapter 5. The results showed evidence for form-meaning links for the group of participants completing both phases on the same day. However, this was not the case for the group of participants completing both phases on different days. The results of the same-day group revealed a reversed priming effect, i.e., slower responses to pictures following FL auditory primes matching them in meaning, suggesting that the connections between FL word forms and semantic representations were still weak at this stage. This result may reflect the initial familiarization stage as put forward by the CLS account of word learning (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010; see Chapter 1, Section 1.4.1 for a review of this model). Following this initial stage, the CLS predicts that new word forms are integrated into the existing mental lexicon following overnight consolidation. Based on this model, it was expected that the results for the group of participants who completed the priming task after a day-delay (hence after an overnight consolidation period) would reflect this two-stage account of word learning, and therefore, a stronger priming effect was expected. However, no priming effect was found for this group of participants. Nevertheless, an incidental learning effect was found, showing that participants had learnt the FL words. In other words, the results of the explicit learning phase showed that participants were able to recognize translation equivalents, even though the priming task revealed no benefits of hearing FL primes on picture recognition. It is possible that once the new FL words had been integrated into the mental lexicon, the connections between the FL word forms and semantic representations were too weak to lead to priming effects. As the CLS predicts that different memory structures are involved in each stage of word learning, the connections between the FL word forms and semantic representations from one stage to the next may have undergone some structural changes or reorganization. It is also possible that the memory traces for the FL words had degraded since the initial exposure. Either way, participants may have benefitted from more time to recognize the FL word forms upon hearing the primes, and to gain access to the semantic representations linked to the FL word forms. Therefore, it would be useful to conduct in the future an experiment using a longer SOA during the initial priming task. Furthermore, EEG may be more sensitive than priming in revealing early learning effects, and as such should be used in future studies.

The group of participants completing the priming task after a day-delay, were asked to repeat the priming task following the explicit learning phase. The results of this second priming task revealed a priming effect both for pictures and word targets, thus providing evidence in support of both lexical-semantic and lexical links. As was discussed in Chapter 5 (Section 5.5), the long SOA used in this experiment may have allowed for activation of both semantic and lexical representations, and therefore it was not possible to assess which occurred first. A similar issue occurred in Guo et al. (2012), however, once they repeated their study with a shorter SOA, they were able to show the time-course of activation of semantic and lexical representations. Therefore, for the second priming task, as completed by the next-day group, it will be important to conduct a similar experiment as the one presented in Chapter 5 with a shorter SOA in order to investigate this further. Furthermore, as mentioned earlier, EEG might provide important additional information beyond the behavioral data, as was the case in Guo et al. Another interesting manipulation would involve combining the designs of the experiments from Chapters 5 and 6, such that the impact of the presentation of FL words in combination with pictures could be compared to FL words presented with NL translation equivalents during the incidental learning task.

To summarize, based on the data presented in this thesis it is not possible to conclude whether the current models of FL word learning can adequately explain learning occurring during an incidental learning situation with multi-modal stimuli. Furthermore, it remains to be seen whether having access to pictorial information during the incidental learning situation promotes the creation of direct links between FL word representations and semantic representations. The results of the same-day group reported in Chapter 5, as well as the picture superiority effect found in Chapter 6, suggest that this may be the case. However, this is speculative at this stage and requires further investigations using additional SOA manipulations and EEG recordings. Furthermore, a slightly different design including FL words paired with pictures as well as NL translations would be useful in evaluating the models of FL word learning in the context of a multi-modal incidental learning situation. Finally, both the RHM and the BIA-d attempt to explain lexical-semantic connections once a FL word representation has been acquired, but neither model accounts for the process of acquiring this new word form representation. In contrast, the CLS model of word learning assumes two distinct stages: a fast initial familiarization stage and a consolidation stage following an intervening period including sleep, in which the new word forms are integrated within the existing mental lexicon. The different pattern of results for the two groups of participants in Chapter 5 (same-day vs. day-delay) indicates that future FL word learning models need to include these two distinct stages if they are to accurately describe the process of acquiring FL vocabulary.

7.6 Limitations and future studies

Limitations specific to each experiment presented in the thesis have already been discussed in each relevant chapters as well as in Sections 7.1 to 7.5 presented above, and therefore will not be repeated here. However, it is important to highlight a few additional limitations, and how these could potentially inform future studies.

Firstly, the experiments included in this thesis all reported the results of exposure to FL words during an incidental learning situation. Furthermore, Chapter 5 investigated the lexical-semantic links resulting from this exposure. However, what the thesis did not investigate is the underlying processes involved. In order to assess whether the current models of FL word learning can account for incidental FL word learning, as opposed to explicit word learning only, it will important to compare the acquisition of FL words in both incidental and explicit learning situations *online.* In other words, are there inherent differences in the online acquisition of FL words happening during an incidental learning situation compared to an explicit learning situation?

A second important consideration concerns the ecological validity of the experiments reported in this thesis, especially the simple multi-modal situation used in the experiments reported in Chapters 3 to 6. In these chapters, participants came to the laboratory and completed a letter-search task whilst they were exposed to simple line drawings and FL word forms. This is not the type of activity that people would normally engage in, and therefore, it would important in future studies to use more appealing activities such as games or short films. Furthermore, future studies should assess the usefulness of multi-modal activities to foster language learning outside of the laboratory. For example, a future study could evaluate informal exposure to FL in a classroom situation.

Finally, this thesis investigated the impact of exposure to multi-modal stimuli on FL word learning only. However, incidental learning could have benefits beyond language learning and therefore its application to other areas would be interesting to consider.

7.7 Conclusions

The research presented in this thesis has clearly shown that informal exposure to FL words with pictures leads to the incidental acquisition of FL vocabulary. Furthermore, incidental learning occurred even after only a few exposures. The goal of the learners, as well as the saliency of each type of information, whether pictorial or verbal, was found to influence the time devoted to the processing of each type of information. In addition, the results of this research highlighted a special role for pictorial information. All these findings have practical implications for FL teachers and learners. For example, including a short activity or game at the beginning of a lesson to introduce key vocabulary may be beneficial for subsequent explicit learning. Furthermore, highlighting important words or images, hence increasing their saliency, may result in increased processing of the information, and this in turn may help learning. The results presented in this thesis suggest that processing of pictorial information leads to improvement in subsequent recall, and as such, learners may benefit from more exposure to multi-modal situations where pictorial information is provided along with FL. The variation on the savings paradigm proved useful in detecting early vocabulary learning, and as such, should be useful for vocabulary researchers. In addition, the research presented in this thesis has highlighted many possible follow-up studies including important theoretical development to inform FL word learning models. Research on the use of multi-modal situations to promote incidental vocabulary learning is still in its infancy, and much work is still needed before it can be ascertained how to best use this type of situation to facilitate FL learning. Being able to communicate in a FL is a rewarding experience, but the process of acquiring a FL can be arduous. As the results presented in this thesis showed, informal exposure to multi-modal situations can help the acquisition of FL vocabulary, and hopefully future research will shed more light on the processes involved.

References

Au, T. K., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological science*, *13*(3), 238–243.

Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417–423.

Baddeley, A. (2002). Is working memory still working? *European Psychologist, 7*(2), 85–97. doi:10.1027//1016-9040.7.2.85

Bajo, M. T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 14*(4), 579–89.

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context : An event-related potential study. *Journal of Cognitive Neuroscience*, *23*(11), 3181–3196.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy & cognition.* Cambridge: Cambridge University Press. 288 p.

Bialystok, E. (2008). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(01), 3–11.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*, 509–533.

Bles, M., & Jansma, B. M. (2008). Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC neuroscience, 9*(20).

Borrás, I., & Lafayette, R. C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal, 78,* 61-75.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289–96.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: the impact of novel word learning on the identification of visually similar words. *Cognition*, *97*(3), B45–54.

Bowers, J. S., Mattys, S. L., & Gage, S. H. (2009). Preserved Implicit Knowledge of a Forgotten Childhood Language. *Psychological Science, 20*(9), 1064-1069.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage, 25*(3), 958–68.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one, 5*(5), e10773.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.

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.

Carpenter, S. K., & Olson, K. M. (2011). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 92–101.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision, 9*(12), 1-15.

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.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning 42*(4), 497-527.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2008). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*(4), 803–20.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*, 3773–3800.

de Bot, K., Martens, V., & Stoessel, S. (2004). Finding residual lexical knowledge: The "Savings" approach to testing vocabulary. *International Journal of Bilingualism, 8*(3), 373-382.

de Bot, K., & Stoessel, S. (2000). In Search of Yesterday's Words: Reactivating a Long-Forgotten Language. *Applied Linguistics, 21*(3), 333-353.

de Groot, A. M. B., & van Hell, J. G. (2005). The Learning of Foreign Language Vocabulary. In J. F. Kroll & A. M. B. de Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 9-29). New York: Oxford University Press.

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2009). New names for known things: on the association of novel word forms with existing semantic information. *Journal of cognitive neuroscience*, *22*(6), 1251–61.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological science*, *18*(1), 35–9.

Duquette, L., & Painchaud G. (1996). A comparison of vocabulary acquisition in audio and video contexts. *Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes 53*(1), 143-172.

d'Ydewalle, G., & de Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, 12, 196-205.

d'Ydewalle, G., Muylle, P., & van Rensbergen, J. (1985). Attention shifts in partially redundant information situations. In R. Groner, G. W. McConkie & C. Menz (Eds.), *Eye Movements and Human Information Processing* (pp. 375-384). North-Holland: Elsevier Science Publishers B.V.

d'Ydewalle, G., & Pavakanun, U. (1995). Acquisition of a second/foreign language by viewing a television program. In P. Winterhoff-Spurk (Ed), *Psychology of Media in Europe: The State of the Art-Perspective for the Future,* (pp. 51-64). Opladen: Westdeutscher Verlag.

d'Ydewalle, G., & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition? In P. Winterhoff-Spurk & T.H.A. van der Voort (Eds), *New Horizons in Media Psychology*, (pp. 145-155). Opladen: Westdeutscher Verlag.

d'Ydewalle, G., Praet, C., Verfaillie, K., & van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research*, 18(5), 650-665.

d'Ydewalle, G., & van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28*(3), 227-244.

d'Ydewalle, G., & van Rensbergen, J. (1989). Developmental studies of text-picture interactions in the perception of animated cartoons with text. In H. Mandl & J. R. Levin (Eds.), *Knowledge Acquisition from Text and Pictures,* (pp. 233-248). North-Holland: Elsevier Science Publishers B.V.

d'Ydewalle, G., van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye Movements from Physiology to Cognition,* (pp. 313-321). Amsterdam, Netherlands: Elsevier Science Publishers B.V.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover Publications. (Original work published 1885). 123 p.

Ellis, N.C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N.C. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 211-282). London: Academic Press.

Gaskell, M., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132.

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*, 319–333.

Graddol, D. (2006). *English Next*. Retrieved from http://www.britishcouncil.org/learning-research-englishnext.htm

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Rethinking the bilingual interactive-activiation model from a developmental perspective (BIA-d). In M. Kail & M. Hickman (Eds.), *Language Acquisition across Linguistic and Cognitive Systems* (pp. 268–283). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Groninger, L. K., & Groninger, L. D. (1980). A comparison of recognition and savings as retrieval measures: A reexamination. *Bulletin of the Psychonomic Society*, 15, 263-266.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language ? *The International Review of Applied Linguistics in Language Teaching*, *50*, 239–276.

Hansen, L., Umeda, Y., & McKinney, M. (2002). Savings in the relearning of second language vocabulary: The effects of time and proficiency. *Language Learning, 52*(4), 653-678.

Hillenburg, S. (2004). Spongebob Squarepants The Movie. United States: Nickleodeon.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes, 8*(4), 379–411.

Holcomb, P. J., Anderson, J., & Grainger, J. (2005). An electrophysiological study of cross-modal repetition priming. *Psychophysiology*, *42*(5), 493–507.

Horst, M. (2005). Learning L2 Vocabulary through Extensive Reading: A Measurement Study. *The Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes*, *61*(3), 355–382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207–223.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use and reoccurrence of unknown words. *The Modern Language Journal*, *80*(3), 327–339.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Perception & Psychophysics, 40*(6), 431–9.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience, 2*, 194–203.

Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. *American Annals of the Deaf*, *145*(3), 275–285.

Kiyonaga, K., Grainger, J., Midgley, K., & Holcomb, P. J. (2007). Masked cross-modal repetition priming: An event-related potential Investigation. *Language and Cognitive Processes, 22*(3), 337–376.

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, *47*(1), 51–60.

Koolstra, C. M., Peeters, A. L., & Spinhof, H. (2002). The Pros and Cons of Dubbing and Subtitling. *European Journal of Communication*, *17*(3), 325–354.

Koskinen. P. S., Knable, J. E., Markham, P. L., Jensema, C. J., & Kane, K. W. (1996). Captioned television and the vocabulary acquisition of adults second language learners correlational facility residents. *Journal of Educational Technology Systems, 24*(4), 359-373.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press. 111 p. Retrieved from http://aces.ir/attachments/22d1286622494-communicative-approach-stephen-crashen.pdf

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations. *Journal of Memory and Language*, *33*, 149–174.

Kroll, J. F., 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.

Kuipers, J.-R., & La Heij, W. (2009). The limitations of cascading in the speech production system. *Language and Cognitive Processes, 24*(1), 120–135.

Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Kweon, S., & Kim, H. (2008). Beyond raw frequency : Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, *20*(2), 191–215.

Laberge, D., & Samuels, S. J. (1974). Toward a Theory of Automatic Information Processing in Reading. *Cognitive Psychology, 6*(2), 293-323.

Lambert, W., Boehler, I., & Sidoti, N. (1981). Choosing the languages of subtitles and spoken dialogues for media presentations: Implications for second language education. *Applied Psycholinguistics*, *2*, 133–148.

Lambert, W., & Holobow, N. (1984). Combinations of printed and spoken dialogue that show promise for students of a foreign language. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement. 16*(1), 1-11.

Laufer, B. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26.

Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition, 33*, 33–55.

Lightbown, P. M., & Spada, N. (1999). *How languages are learned.* Oxford: Oxford University Press. 192 p.

Lindsay, S., & Gaskell, M. G. (2010). A Complementary Systems Account of Word Learning in L1 and L2. *Language Learning*, (December), 45–63.

Lotto, L., & de Groot, A. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48*(1), 31–69.

MacLeod, C. M. (1988). Forgotten but not gone: Savings for pictures and words in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 195-212.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.

MacMillan, N.A. & Creelman, C.D. (1991). *Detection Theory: A users guide*. Mahwah: Lawrence Erlbaum Associates. 513 p.

Mayer R. E. (2005). Principles for reducing extraneous processing in multimedia learning: Coherence, signalling, redundancy, spatial contiguity, and temporal contiguity principles. In R.E. Mayer (Ed), *The Cambridge Handbook of Multimedia Learning* (pp. 183-200). Cambridge: Cambridge University Press.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35*(3), 494–503.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–4.

Mestres-Missé, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience, 20*(12), 2153–66.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2010). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation, 3*(1), 5–24.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS One*, *4*(11), e7785.

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PloS One*, *7*(5), e37527.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 555–563.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. 477 p.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*(3), 359-377.

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of experimental psychology. Human learning and memory, 2*(5), 523–8.

Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(5), 453-468.

Nelson, T. O. (1985). Ebbinghaus's contribution to the measurement of retention: Savings during relearning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 472-479.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*(1), 94–106.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206.

Paribakht, S. T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. N. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* (pp. 174-200). Cambridge: Cambridge University Press.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The Cognitive Effectiveness of Subtitle Processing. *Media Psychology, 13*(3), 243-272.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel : Do Things Fall Apart? *Reading in a Foreign Language*, *22*(1), 31–55.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading : A case study. *Reading in a Foreign Language*, *18*(1), 1–28.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*, 253–336.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search : Effects of word frequency. *Psychonomic Bulletin & Review, 3*(2), 245–248.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235.

Rebuschat, P., & Williams, J. N. (2011). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(4), 829–856.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*, 589–619.

Samuels, S. J. (1994). Toward a theory of automatic information processing in reading, revisited. In R. B. Ruddell, M. R. Ruddell & H.

Singer (Eds.), *Theoretical Models and Processes of Reading* (4th ed., pp. 816-837). Newark, DE: International Reading Association.

Sanz, C. (2013). Multilingualism and Metalinguistic Awareness. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 3933–3942). doi:10.1002/9781405198431.wbeal0802

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distribution of cognates in Europe as based on the Levenshtein distance. *Bilingualism: Language and Cognition, 15*(1), 157-166.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual.* Basingsoke: Palgrave Macmillan. 420 p.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory, 6*(2), 174–215.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109*(4), 373–392.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8):16, 1–24.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review, 113*(4), 766–86.

van Heuven, W. J. B. (2010). Eyetracker Output Utility (Version 1.20 beta 14) [Computer Program]. Retrieved from http://www.psychology.nottingham.ac.uk/staff/wvh/eou/index.html

Veroude, K., Norris, D. G., Shumskaya, E., Gullberg, M., & Indefrey, P. (2010). Functional connectivity between brain regions involved in learning words of a new language. *Brain and Language*, *113*(1), 21–7.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258.

Waring, R., & Takani, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91–120.

Williams, J. N. (2009). Implicit learning in second language acquisition. In T. K. Bhatia & W. C. Ritchie (Eds.), *The New Handbook of Second Language Acquisition* (pp. 319–353). Bingley, UK: Emerald Group Publishing.

Williams, J. N. (2010). Initial Incidental Acquisition of Word Order Regularities : Is It Just Sequence Learning ? *Language Learning, 60*(S2), 221–244.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86.

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics, 14*(04), 423–444.

Zhang, T., van Heuven, W. J. B., & Conklin, K. (2011). Fast automatic translation and morphological decomposition in Chinese-English bilinguals. *Psychological Science, 22*(10), 1237–1242.

Appendix A

List of Dutch and English words used in Chapter 2 with both match and mismatch trials.

Dutch	English	Trial
volgende	next	match
volgende	formula	mismatch
jou	your	match
jou	yawn	mismatch
iemand	someone	match
iemand	himself	mismatch
voor	for	match
voor	four	mismatch
wat	what	match
wat	at	mismatch
je	you	match
je	yeah	mismatch
restaurant	restaurant	match
restaurant	restored	mismatch
gemeen	mean	match
gemeen	main	mismatch
eens	once	match
eens	into	mismatch
krab	crab	match
krab	grab	mismatch
zee	sea	match
zee	see	mismatch
plaats	place	match
plaats	splatter	mismatch
hallo	hello	match
hallo	fellow	mismatch
iets	something	match
iets	eat	mismatch
met	with	match
met	met	mismatch
klaar	ready	match
klaar	car	mismatch
nu	now	match
nu	no	mismatch
terug	back	match
terug	truck	mismatch
en	and	match
en	an	mismatch
nieuw	new	match

nieuw	news	mismatch
hij	he	match
hij	hi	mismatch
geheime	secret	match
geheime	game	mismatch
vandaag	today	match
vandaag	wonder	mismatch
stop	stop	match
stop	spot	mismatch
onze	our	match
onze	hour	mismatch
paars	purple	match
paars	pass	mismatch
kleine	little	match
kleine	climber	mismatch
bedrijfsleider	manager	match
bedrijfsleider	believe	mismatch
wereld	world	match
wereld	word	mismatch
hoe	how	match
hoe	whom	mismatch
kijk	look	match
kijk	kill	mismatch
wacht	wait	match
wacht	fat	mismatch
beter	better	match
beter	bitter	mismatch
ja	yes	match
ja	yet	mismatch
denk	think	match
denk	thank	mismatch
vriend	friend	match
vriend	free	mismatch
twee	two	match
twee	twelve	mismatch
omdat	because	match
omdat	under	mismatch
werk	work	match
werk	jerk	mismatch
zijn	his	match
zijn	sign	mismatch
nooit	never	match
nooit	not	mismatch
over	about	match
over	cover	mismatch
wij	we	match

wij	way	mismatch
gestolen	stolen	match
gestolen	crustacean	mismatch
liefde	love	match
liefde	later	mismatch
koning	king	match
koning	coming	mismatch
plan	plan	match
plan	pan	mismatch
dit	this	match
dit	bit	mismatch
vader	father	match
vader	farther	mismatch
waar	where	match
waar	far	mismatch
baan	job	match
baan	bag	mismatch
baas	boss	match
bass	bless	mismatch
dood	dead	match
dood	don't	mismatch
kaas	cheese	match
kaas	castle	mismatch
kind	kid	match
kind	kind	mismatch
dat	that	match
dat	dad	mismatch
ogen	eyes	match
ogen	hooray	mismatch
mijn	my	match
mijn	remind	mismatch
maar	but	match
maar	may	mismatch
vingerafdrukjes	fingerprints	match
vingerafdrukjes	forbidden	mismatch
promotie	promotion	match
promotie	problem	mismatch
het	it	match
het	hit	mismatch
iedereen	everyone	match
iedereen	either	mismatch
leven	life	match
leven	leave	mismatch
hier	here	match
hier	hear	mismatch
man	man	match

man	mine	mismatch
in	in	match
in	on	mismatch
kroon	crown	match
kroon	crime	mismatch
wie	who	match
wie	be	mismatch
kaal	bald	match
kaal	call	mismatch
dochter	daughter	match
dochter	doctor	mismatch
ik	I	match
ik	sick	mismatch
verkopen	selling	match
verkopen	bucket	mismatch
dag	day	match
dag	dare	mismatch
ober	waiter	match
ober	over	mismatch
tijd	time	match
tijd	tiny	mismatch
naam	name	match
naam	none	mismatch
duivelse	evil	match
duivelse	deserve	mismatch

Appendix B

List of English and Welsh words used in Chapters 3 and 4 with list number, frequency band (for Chapter 4 only) and picture number (Snodgrass & Vandervart, 1980).

English	Welsh	List 1	List 2	Frequency band	Picture number (incidental learning)
airplane	awyren	old	new	2	2
ant	morgrug	new	old	1	5
arm	braich	old	new	1	7
arrow	saeth	old	new	3	8
axe	bwyell	old	new	2	12
barn	wen	new	old	1	17
barrel	gasgen	old	new	4	18
bed	gwely	new	old	1	22
bee	gwenyn	old	new	1	23
beetle	chwilen	new	old	4	24
bell	gloch	new	old	3	25
bird	adar	new	old	1	28
boot	cist	old	new	2	31
bow	cwlwm	new	old	4	33
box	blwch	new	old	3	35
bread	bara	old	new	2	36
broom	ysgub	new	old	1	37
carrot	moron	new	old	3	48
caterpillar	lindysyn	old	new	1	50
chair	cadair	old	new	4	53
comb	crib	old	new	4	65
deer	ceirw	new	old	4	71
dog	ci	new	old	4	73
duck	hwyaden	old	new	1	81
ear	clust	new	old	2	83
envelope	amlen	new	old	2	85
eye	llygad	new	old	3	86
finger	bys	old	new	4	88
fish	pysgodyn	old	new	2	89
flower	blodyn	new	old	4	91
fly	pryf	old	new	4	93
foot	troed	new	old	2	94
fox	cadno	new	old	2	98
frog	broga	old	new	1	100
glass	gwydr	old	new	1	104
glove	faneg	new	old	1	106
grapes	grawnwin	new	old	2	109

gun	dryll	old	new	4	112
hair	gwallt	old	new	2	113
hammer	morthwyl	new	old	3	114
hand	llaw	new	old	2	115
hanger	cambren	old	new	4	116
harp	telyn	new	old	3	117
heart	calon	old	new	2	119
horse	ceffyl	old	new	3	121
house	ty	new	old	2	122
kettle	tegell	old	new	4	127
knife	cyllell	new	old	3	130
leaf	deilen	new	old	1	133
leg	coes	old	new	3	134
lock	clo	old	new	1	143
moon	lleuad	new	old	1	146
mountain	mynydd	old	new	3	148
mouse	llygoden	old	new	3	149
mushroom	madarchen	old	new	3	150
nail	hoelen	new	old	3	151
necklace	cadwyn	old	new	3	153
nose	trwyn	new	old	1	155
pig	mochyn	old	new	4	172
rabbit	cwningen	new	old	4	182
refrigerator	oergell	new	old	3	185
rooster	ceiliog	new	old	1	191
sandwich	brechdan	new	old	2	195
seal	seinod	old	new	1	201
sheep	dafad	old	new	1	202
shirt	crys	new	old	3	203
shoe	esgid	old	new	3	204
snail	malwen	old	new	4	208
snake	neidr	new	old	2	209
spider	copyn	new	old	4	212
spoon	llwy	new	old	4	215
stove	popty	old	new	2	219
sun	haul	new	old	4	222
table	bwrdd	old	new	2	226
thumb	bawd	old	new	3	231
tree	coeden	old	new	2	241
turtle	crwban	new	old	4	244
vase	cawg	new	old	2	246
watch	oriawr	old	new	1	250
wheel	olwyn	old	new	3	254

Appendix C

List of English and Welsh words used in Chapter 5 with list number, picture number during the incidental learning phase (Snodgrass & Vandervart, 1980) and table number with picture stimuli used during priming phase.

English	Welsh	List 1	List 2	Picture number (incidental learning)	Table with picture (priming phase)
airplane	awyren	old	new	2	Table C1
ant	morgrug	new	old	5	Table C1
arrow	saeth	old	new	8	Table C1
axe	bwyell	old	new	12	Table C1
barrel	gasgen	new	old	18	Table C3
bed	gwely	new	old	22	Table C1
bee	gwenyn	old	new	23	Table C1
beetle	chwilen	new	old	24	Table C1
bell	gloch	new	old	25	Table C3
bird	adar	new	old	28	Table C1
bow	cwlwm	new	old	33	Table C2
box	blwch	new	old	35	Table C2
carrot	moron	new	old	48	Table C1
caterpillar	lindysyn	old	new	50	Table C3
chair	cadair	old	new	53	Table C1
comb	crib	old	new	65	Table C2
deer	ceirw	new	old	71	Table C3
dog	ci	new	old	73	Table C3
duck	hwyaden	old	new	81	Table C1
envelope	amlen	new	old	85	Table C2
fish	pysgodyn	old	new	89	Table C1
flower	blodyn	old	new	91	Table C1
fly	pryf	old	new	93	Table C1
fox	cadno	new	old	98	Table C3
frog	broga	old	new	100	Table C3
glass	gwydr	old	new	104	Table C2
glove	faneg	new	old	106	Table C1
grapes	grawnwin	new	old	109	Table C1
gun	dryll	old	new	112	Table C1
hammer	morthwyl	new	old	114	Table C2
hanger	cambren	old	new	116	Table C2
harp	telyn	new	old	117	Table C1
horse	ceffyl	old	new	121	Table C1
house	ty	new	old	122	Table C1
kettle	tegell	old	new	127	Table C2
knife	cyllell	new	old	130	Table C2

leaf	deilen	new	old	133	Table C2
lock	clo	old	new	143	Table C2
moon	lleuad	new	old	146	Table C1
mouse	llygoden	old	new	149	Table C3
mushroom	madarchen	old	new	150	Table C2
nail	hoelen	new	old	151	Table C1
necklace	cadwyn	old	new	153	Table C1
pig	mochyn	old	new	172	Table C3
rabbit	cwningen	new	old	182	Table C3
refrigerator	oergell	new	old	185	Table C3
rooster	ceiliog	new	old	191	Table C1
seal	seinod	old	new	201	Table C3
sheep	dafad	old	new	202	Table C3
shirt	crys	new	old	203	Table C1
shoe	esgid	old	new	204	Table C1
snail	malwen	old	new	208	Table C3
snake	neidr	new	old	209	Table C1
spider	copyn	new	old	212	Table C1
spoon	llwy	new	old	215	Table C2
table	bwrdd	old	new	226	Table C1
tree	coeden	old	new	241	Table C1
turtle	crwban	new	old	244	Table C1
watch	oriawr	old	new	250	Table C2
wheel	olwyn	old	new	254	Table C3

Table C1: Picture stimuli used during priming task, from Moreno-Martínez and Montoro (2012).

				
				
				
				

				
				
				

Table C2: Picture stimuli used during priming task, from Brodeur et al., (2010).

Table C3: Picture stimuli used during priming task, from various websites.

Appendix D

List of English and Welsh words used in Chapter 6 with duration of Welsh sound file, presentation mode in each list, and picture number during the incidental learning phase (Snodgrass & Vandervart, 1980).

English	Welsh	Sound duration	List 1	List 2	List 3	Picture number
airplane	awyren	963	AWP	A	AW	2
ant	morgrug	955	AW	AWP	A	5
arm	braich	656	A	AW	AWP	7
arrow	saeth	840	AW	AWP	A	8
axe	bwyell	959	AWP	A	AW	12
barrel	gasgen	900	AW	AWP	A	18
bed	gwely	749	A	AW	AWP	22
bee	gwenyn	661	AW	AWP	A	23
beetle	chwilen	826	A	AW	AWP	24
bell	gloch	605	AW	AWP	A	25
bird	adar	891	AWP	A	AW	28
bow	cwlwm	866	A	AW	AWP	33
box	blwch	654	AW	AWP	A	35
bread	bara	563	A	AW	AWP	36
broom	ysgub	754	A	AW	AWP	37
carrot	moron	650	A	AW	AWP	48
caterpillar	lindysyn	937	AWP	A	AW	50
chair	cadair	856	AW	AWP	A	53
comb	crib	595	AWP	A	AW	65
deer	ceirw	651	AW	AWP	A	71
dog	ci	423	AWP	A	AW	73
duck	hwyaden	1043	AWP	A	AW	81
ear	clust	575	AWP	A	AW	83
envelope	amlen	711	A	AW	AWP	85
eye	llygad	775	AWP	A	AW	86
finger	bys	556	A	AW	AWP	88
fish	pysgodyn	898	AW	AWP	A	89
flower	blodyn	640	AWP	A	AW	91
fly	pryf	555	AWP	A	AW	93
foot	troed	656	AW	AWP	A	94
fox	cadno	606	AWP	A	AW	98
frog	broga	663	AW	AWP	A	100
glass	gwydr	689	AWP	A	AW	104
glove	faneg	940	AW	AWP	A	106
grapes	grawnwin	780	AW	AWP	A	109
gun	dryll	577	AWP	A	AW	112
hair	gwallt	746	A	AW	AWP	113
hammer	morthwyl	823	AW	AWP	A	114

hand	llaw	657	AW	AWP	A	115
hanger	cambren	751	AW	AWP	A	116
harp	telyn	563	A	AW	AWP	117
heart	calon	576	AWP	A	AW	119
horse	ceffyl	536	A	AW	AWP	121
house	ty	470	AWP	A	AW	122
kettle	tegell	853	AW	AWP	A	127
knife	cyllell	734	AW	AWP	A	130
leaf	deilen	708	A	AW	AWP	133
leg	coes	648	AW	AWP	A	134
lock	clo	525	AW	AWP	A	143
moon	lleuad	866	AW	AWP	A	146
mountain	mynydd	850	A	AW	AWP	148
mouse	llygoden	1003	AWP	A	AW	149
mushroom	madarchen	1077	AW	AWP	A	150
nail	hoelen	783	A	AW	AWP	151
necklace	cadwyn	792	A	AW	AWP	153
nose	trwyn	503	AWP	A	AW	155
pig	mochyn	790	A	AW	AWP	172
rabbit	cwningen	950	A	AW	AWP	182
refrigerator	oergell	794	A	AW	AWP	185
rooster	ceiliog	791	AW	AWP	A	191
sandwich	brechdan	869	AWP	A	AW	195
seal	seinod	973	A	AW	AWP	201
sheep	dafad	649	AW	AWP	A	202
shirt	crys	684	AW	AWP	A	203
shoe	esgid	760	AWP	A	AW	204
snail	malwen	771	A	AW	AWP	208
snake	neidr	821	AWP	A	AW	209
spider	copyn	616	AWP	A	AW	212
spoon	llwy	602	AWP	A	AW	215
stove	popty	579	A	AW	AWP	219
sun	haul	708	A	AW	AWP	222
table	bwrdd	626	AWP	A	AW	226
thumb	bawd	585	AW	AWP	A	231
tree	coeden	791	AWP	A	AW	241
turtle	crwban	827	A	AW	AWP	244
vase	cawg	618	A	AW	AWP	246
watch	oriawr	1019	AWP	A	AW	250
wheel	olwyn	703	A	AW	AWP	254