



**UNIVERSITEIT VAN AMSTERDAM**  
**FACULTEIT DER GEESTESWETENSCHAPPEN**

MASTER THESIS GENERAL LINGUISTICS:

## Dutch Vocabulary of Adult L2 Learners

Marieke A. Gerrits (0474754)  
Supervisors: Paola Escudero & Paul Boersma  
August 2009

## Acknowledgement

First of all, I would like to thank all the Latin American, Spanish and Dutch subjects that participated in this study. Secondly, I would like to thank Paola Escudero for giving me the opportunity to use data of her longitudinal project and for supervising me during the first stages of writing this thesis. Paul Boersma, thanks for taking over Paola's role and advising me in the final stages of writing my thesis. Finally, of course, I would like to thank my family and friends for supporting me, special thanks to my father who helped me analyzing the data.

## Contents

Acknowledgement

1. Introduction	2
2. The acquisition of vocabulary in a second language	4
2.1 The lexicon	4
2.2 The acquisition of vocabulary in an L2	7
2.3 Vocabulary size in L2 learners of Dutch	10
3. Vocabulary size and listening comprehension	11
3.1 Cognitive processes in listening comprehension	11
3.2 The relationship between vocabulary size and listening comprehension	13
4. Testing vocabulary in a second language	16
4.1 Assessing vocabulary in a second language	16
4.2 Development of the Vocabulary Test by Brasileiro & Viellevoye Grooters	19
4.3 Requirements Vocabulary Test for adult L2 learners of Dutch	20
5. Research Questions + Hypotheses	22
6. Method	24
6.1 Participants	24
6.2 Instruments	25
6.3 Procedure	26
6.4. Data-analysis	26
7. Results	29
8. Discussion	33
9. Conclusions	36
Summary	37
References	38
Appendices	41
Appendix A	42
Appendix B	43
Appendix C	44

## Chapter 1. Introduction

Learning a second language implies learning the words of that language. Literature on the acquisition of words in a foreign language shows that there is a variety of aspects to study. With the present study I will try to contribute to research into second language acquisition. More specifically, I will study vocabulary acquisition in L2 learners of Dutch, by addressing three facets of it, namely, vocabulary size in L2 learners with different language backgrounds, vocabulary size and the connection with listening comprehension level, and the analysis of the vocabulary test used in the experiment.

To investigate this topic, data from an L2 research project carried out by Paola Escudero were analysed. For her longitudinal study into the acquisition of the Dutch vowel system of Spanish and Portuguese speaking learners of Dutch, people from Latin America and Spain were tested. They performed several tests, most of them investigating their vowel perception. To get the learners in the Dutch 'mode', the testing session started with a Dutch vocabulary test followed by a listening comprehension test. Since these tests were only used to get them in the Dutch mode, the results will not be analysed within Escudero's project, which made it possible for me to analyse these data for my thesis. On the basis of the available data, the following three aspects will be studied.

First, an analysis of the scores of the participants on the vocabulary test will be made; this makes it possible to see if there are differences between the different learner groups. They all have a Romance language as their mother language and are learning Dutch, which is a Germanic language, as their second or third language. Research into this language-learning scenario (Romance language-to-Dutch) is not abundant. Besides that, a comparison between the vocabulary score of an adult L1 speaker and the vocabulary score of an adult L2 speaker would be interesting. Therefore, a group of Dutch adults also performed the vocabulary test. It can be assumed that the vocabulary of the L2 learners, in particular in the earliest stages of second language acquisition, is much smaller than the vocabulary of native speakers. After all, the adult L1 learners have had a much longer period to learn words of their language since they started to acquire the language from birth. It is worth mentioning that studies into the vocabulary of child second language learners, who do not have the disadvantage of starting learning the L2 after the critical period (Lenneberg, 1967), also show that this group has a more limited vocabulary than monolinguals (Verhoeven & Vermeer, 1992, in Appel & Vermeer, 1994). The same holds for bilingual teenagers (Sanders, 1990, in Appel & Vermeer,

1994). A difference between the Dutch control group and the experimental groups will therefore not be surprising at all. However, it is important to examine the extent of the difference with the three learner groups. Most importantly, however, it will be interesting to establish the actual differences in vocabulary knowledge between these L2 learner groups.

Secondly, I will determine if vocabulary size correlates with listening comprehension level. There are several L2 studies that investigate the correlation between vocabulary size and other factors, such as length of residence, age, age of onset, listening comprehension, I.Q. etc. Since the participants not only performed the vocabulary test but also a listening comprehension test, it is possible for me to study the connection between these two factors. Previous studies that examine this correlation (Richards, 1983; Rost, 1990; Bonk, 2000; Mecarty, 2000) suggest a connection between vocabulary size and listening comprehension level. It would be interesting to see if an analysis of the data obtained for Escudero's project shows agreement with the results of these studies.

Thirdly, as for the practical importance of this MA research, the analysis of the scores on the vocabulary test, developed by Brasileiro and Viellevoeye Grooters in 2005, will give insight into the usefulness of this specific test as a means of assessing L2 adults' vocabulary, given that this test has been recently developed to test bilingual children and that it is the first time it has been used for L2 adults.

The analysis of the three different aspects makes the thesis consist of three parts. The first three chapters represent theoretical frameworks for each research question; this implies that in chapter 2 the acquisition of vocabulary in second language learners is discussed. This chapter addresses questions like: how do second language learners memorize words and how are words stored in the mental lexicon. In chapter 3, the relationship between listening comprehension and vocabulary size is discussed. Previous research on this topic will be described. In chapter 4, the focus is on testing vocabulary in a second language. I will describe the development of the test used in this experiment and explain what conditions it has to meet to be useful in the assessment of adults. In chapter 5 the research questions and accessory hypotheses are formulated. In chapter 6 the method of this study is discussed, including a description of the participants, the instruments, the procedure and the data analysis. Chapter 7 describes the results. The thesis ends with a discussion of the results in chapter 8 and the conclusions in chapter 9.

## Chapter 2. The acquisition of vocabulary in a second language

Appel (1996) claims that L2 vocabulary acquisition has been a neglected issue for a long time. Consequently information about the way vocabulary is learned in another language was hardly available. However, the position of L2 vocabulary acquisition research has changed since the 1990's, according to Sánchez & Manchón (2007). These authors state that from that time on many theoretical and empirical studies were carried out in this area. This indicates that there is general agreement about 'the crucial role lexis plays in second language learning [...] ' (Sánchez & Manchón, 2007, p. 7). Gass & Selinker (2001) even claim that 'the lexicon may be the most important language component for learners' (p. 372).

The first section of this chapter will start with a general description of the lexicon. It also includes a short discussion of the mental lexicon of L2 learners. Section 2.2 discusses the acquisition of vocabulary in a second language. While the acquisition of words for L1 learners has been described extensively, including different stages the learner passes through, the literature concerning the acquisition of words in a second language is scarce, according to Appel (1996). However, several researchers studied the acquisition of vocabulary in an L2. In section 2.2 I describe the prevailing theory on this topic. I will conclude with a brief review of the literature concerning vocabulary size in Dutch L2 learners.

### 2.1 The lexicon

How many words does a language count, how large is the vocabulary of an average L1 speaker and what number of words does anyone need at least to use a language? These questions are formulated and discussed by Nation (2001). First I will give an impression of the numbers by summarizing Nation. Secondly, I will describe how words fit into the structure of the internal lexicon.

How many words a language consists of, is a complex question. In order to answer this we first must agree on the definition of 'a word'. Nation distinguishes: *tokens*, *types*, *lemmas* and *word families*. Simply counting every word form in a text, and counting the same word form double as it occurs more than once is the way in which we count words as '*tokens*'. Following this counting method results in high numbers. On the other hand, we count '*types*' if we count the words in a text, but now, opposite to counting tokens, a word is counted once even if it occurs more than once in the text. This approach results in lower numbers than the token-method. Using '*lemmas*' as the counting unit reduces the number of words further. 'A

lemma consists of a headword and some of its inflected and reduced (*n't*) forms' (Nation, 2001, p. 7). Nation claims that researchers do not agree on an exact definition of a lemma. Some of them accept for instance irregular verbs as distinct words, others do not. Finally we observe '*word families*'. 'A word family consists of a headword, its inflected forms, and its closely related derived forms' (Nation, 2001, p. 8). As there is no exact definition of a word family, counting word families never leads to reproducible quantitative results. Nevertheless a few attempts were made to quantify the complete English vocabulary. This was done by counting the words in large dictionaries. A study by Goulden, Nation & Read (1990, in Nation, 2001) that followed this method yielded 114,000 word families, proper names were not included.

How large is the vocabulary of an average L1 speaker? It is beyond doubt that native speakers do not know all the words of their language. Nation claims that studies that investigated vocabulary size done in the past show mainly incorrect results while more recent studies show more reliable results. These studies estimate the vocabulary size in 'educated native speakers of English' at 'around 20,000 word families' (Nation, 2001, p. 9). Nation remarks that since only word families were counted these numbers are quite low.

Acquiring a vocabulary comparable to that of a native speaker is an idealistic goal for second language learners. Although this is realizable in the long term, in the short term learners should focus in first place on high-frequency words, as suggested by Nation (2001). He claims that within the group of high-frequency words there are certain words which are of first importance to learn. He argues that 'these words cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses in the language' (p. 13).

The next section gives a brief description of two kinds of models, namely hierarchical network models and spreading activation models. These models are described in Carroll (2004) and give insight into the possible organization of words in the lexicon. Carroll claims that hierarchical network models can partly explain the way words are stored in the mental lexicon, whereas spreading activation models, which are not 'strictly hierarchical', offer 'the most realistic picture' of the organization of the internal lexicon (Carroll, 2004, p. 113). Below I will give a summary of what Carroll states about these models.

Hierarchical network models consist of 'elements' or 'nodes' that stand below or above other elements in the network. At every node a word, supplemented with semantic information is stored. Collins and Quillian (1969, in Carroll, 2004), who developed the idea of a hierarchical network, assume that this semantic information is not stored at all the nodes, but

as a consequence of limited space, only at the highest node. Other researchers, Rosch, Mervis, Gray, Johnson, & Boyes-Braem (1976, in Carroll, 2004) suggest that it is more plausible that information is stored at more 'familiar locations' (p. 110). Whether semantic information is stored at the highest nodes, or at a more basic level, the connection between nodes makes it possible to exchange information.

Spreading activation models on the other hand 'provide a more flexible way of representing lexical information [...]' (Carroll, 2004, p. 113). The starting point is, again, a network; however the way it is built up is different. In contrast with hierarchical network models, information in a spreading activation model is stored at more or less the same level. This kind of model is often depicted like a web with interconnected nodes. Carroll remarks that some concepts in the network are more 'accessible' than others; this depends on factors such as 'frequency of usage' and 'typicality' (p. 110).

For L2 learners the question is whether they have two separated lexicons, one for their L1 and one for their L2, or one integrated lexicon. Singleton (1999) argues that we should not see the L1 and L2 lexicon either as 'completely disconnected' or as 'totally integrated'. Separated storage of the L1 and L2 lexicon, where communication between the two systems is possible, seems more plausible to him. Singleton adds to this view that the relationship between an L1 word and an L2 word depends on factors like, 'how the words have been acquired and how well they are known, and also on the degree to which formal and/or semantic similarity is perceived between the L2 and the L1 word in question' (p. 190).

A study by Wolter (2001) investigated what the L2 mental lexicon in comparison to the L1 lexicon may look like. This study explores whether the structure of both lexicons i.e. the way they are organized, shows a certain resemblance. Wolter claims that similarities were found, though after the experiment it was concluded that there also seem to be some 'fundamental differences' between both lexicons. According to Wolter (p. 66), 'semantic connections' between L1 and L2 words seem to become more dominant over 'phonological connections' when the quality of word knowledge increases. How much an individual word is integrated in the lexicon, determines according to the author also the extent to which 'syntagmatic' and 'paradigmatic connections' between words play a role. From this study it can be concluded that depth of lexical knowledge is an important factor in studies that investigate the structure of L1 and L2 lexicons.

In their study into 'Shared and separate meanings in the bilingual mental lexicon', Dong, Gui & Macwhinney (2005) propose a model that provides insight into the organization of the 'bilingual mental lexicon'. The model is called *'The shared (distributed) asymmetrical*

*model*' (p. 233); it was developed after the conduction of two experiments. From the experiments it appeared that on the one hand, conceptual representations are shared and in case they are not identical for both languages the L2 learner tries to integrate them. On the other hand, the findings of experiment 2 show a 'separatist tendency' (p. 233) in which integration of both systems was not observed. This leads to a model that according to the authors summarizes the results of the experiments: 'the diagram emphasizes both the asymmetrical nature of L1 and L2 access to meanings and the extent to which meanings are both shared and partially separated for L1 and L2' (p. 233).

## 2.2 The acquisition of vocabulary in an L2

The acquisition of words in a second language shows some differences with the acquisition of words in an L1. Three essential differences as formulated by Appel (1996, p. 384) are described below. The first difference that we have to take into account according to Appel is 'the stage of cognitive development of the learner'. L1 learners are only able to learn words if they have reached a certain stage in cognitive development. Conversely, the knowledge of words can also stimulate cognitive development. L2 learners have, in contrast with L1 learners, already reached a certain stage in cognitive development and therefore they have already learned, that objects in the world around them are concepts that can be labelled. The second difference Appel comes up with is that 'the L2 learner already has a semantic system', that for his L1, opposite to L1 learners, who have to build a complete semantic system when they start acquiring words. The L1 semantic system (already present in the L2 learner) shows similarities as well as differences with the semantic system of the L2. Similarities between both systems facilitate the acquisition of L2 words. The differences however, as claimed by Appel, make learning new words complicated, since the meaning of the majority of new words is not exactly the same as the meaning of the words in the L1. The third difference according to Appel is that children learn words in a natural way, whereas for L2 adults it often requires a great deal of effort to construct a lexicon in a foreign language. Furthermore he claims that for L2 learners it is more often a problem to learn words incidentally than for L1 children, in particular when they do not have native speakers in their environment.

According to Appel very little is known about the course of lexical development in an L2. He puts forward that since a theoretical model that explains the course of lexical development is lacking, research in this field is mainly descriptive. Schmitt (2000) argues that the large number of factors that influence L2 vocabulary learning makes it hard to develop a theory that can explain the acquisition of words. These factors can be linguistic or extra-

linguistic. Linguistic factors, and more specifically lexical factors, that affect word learning are, among other things: *'pronounceability, grammatical category of the word, semantic features of the word (e.g. abstractness, specificity and register restriction, idiomaticity and multiple meaning)'* (Laufer, 1997). Besides that, there are many extra-linguistic factors that influence the acquisition of vocabulary in a second language, I will sum up the factors discussed in Appel & Vermeer (1994): *'age, sex, language aptitude, intelligence, proficiency in L1, cognitive style, personality, social-psychological factors, language contact and instruction'*. Appel & Vermeer state that most of these factors seem to have some connection with the rate and success of L2 acquisition; a causal connection however has not been proved for all of them. It should also be noted that this list can be complemented with other factors.

To summarize the acquisition of words in one sentence, I will cite Appel (1996, p. 387): *'it can be expected that in the acquisition process the elaboration of the vocabulary proceeds in two directions: learners will learn more words, and the meaning of words will be expanded or they will learn more specific meanings'*. With respect to the acquisition of vocabulary, Schmitt (2000) gives a fairly extensive overview of the variables that play a role in vocabulary acquisition. I will give a summary of this overview. Like Appel, Schmitt (p. 117) also stresses that the acquisition of vocabulary is *'incremental in nature'*. According to him there are many aspects of a word that have to be learned before thorough command of a word can be observed. These word knowledge aspects are usually not acquired simultaneously. For example, *'learners will surely know a word's basic meaning before they have full collocational competence'* (Schmitt, 2000, p. 117). The next point Schmitt addresses is that words can be learned explicitly or incidentally, as mentioned above, children learn most words incidentally. Schmitt argues that for L2 learners a combination of both ways of word learning is necessary to gain a fairly sized lexicon. He states that the words that occur most frequently in a language should be learned explicitly; whereas infrequent words should better be learned incidentally. Another point Schmitt discusses is that the acquisition of words is not only incremental; learners also forget words they have learned. Learning and forgetting are both part of the process and occur till a word is *'fixed'* in memory. Forgetting words that are fixed in memory is also possible; this may occur when a learner finishes a language course and/or stops using the target language; this is called *'attrition'* (p. 129). Words that a language learner only knows receptively are more likely to be forgotten than words he/she knows receptively and productively (Schmitt, 2000).

Hatch and Brown (1995) describe five steps that learners have to pass through in vocabulary learning. These five steps are: *'1) encountering new words, 2) getting the word*

form, 3) getting the word meaning, 4) consolidating word form and meaning in memory, and 5) using the word'. A description of the five processes, based on Hatch and Brown, is given below. According to Hatch and Brown step 1 is essential to achieve a vocabulary size comparable to that of native speakers. They mention that encountering new words implies that the learner needs a source for words. Possible sources (as appeared from a study by Payne, 1988, in Hatch & Brown, 1995) are books, magazines, newspapers and dictionaries. Also watching TV and listening to radio can be very useful. Social interaction with the speakers of the target language also offers an opportunity to encounter new words. The second step implies getting the 'auditory' and/or 'visual' form of the word. Knowing the auditory word form is important for the retrieval of words, because learners seem to rely on sound aspects in the retrieval of words from the lexicon (Brown & McNeill, 1966 in Hatch & Brown, 1995). Hatch and Brown claim that acquiring a clear image of the visual form, in other words the spelling of a word, prevents the learner from making mistakes by confusing the target word with a word with a similar form in the native language or the target language. The third step is, according to Hatch & Brown, the first thing that comes to mind when people think of learning words namely getting the word meaning. They argue that the depth of the definition a learner may need, depends on the stage of the process of language learning and with that on the language level of the learner. Hatch & Brown mention that L2 learners use different sources to find out the definition of a word. These sources vary from a dictionary to the context in which a word is presented. The fourth step is consolidating the words in memory. There are many methods and strategies to obtain this, for a further discussion of this I refer to Hatch & Brown (1995, p. 387). Step five, using the word, is according to the authors essential for learners who want to test if they use the word in an appropriate context and gives an idea about how well the knowledge acquired during the previous steps is stored in memory. Hatch & Brown (1995, p. 391) summarize their discussion about the five essential steps in word learning as follows: 'Although there is a broad range of activities, strategies, or techniques that individuals use at each step, the necessity of the steps seems more constant. Learners need all five in order to have a full knowledge of the words they want to learn.'

A study by Schmitt (1998) contributes to the understanding of the acquisition of words. Schmitt studied the acquisition of 11 words in 3 adult L2 learners. The study had a longitudinal character, since he followed the participants during one year. Four word knowledge aspects i.e. spelling, associations, grammatical information, and meaning were investigated. The experiment revealed that some of the word knowledge aspects seem to be connected to each other; this is however not statistically proved because the author doubted

the usefulness of statistical procedures with only 3 participants. The main conclusion says that ‘the study did not show evidence of a developmental hierarchy for word knowledge types’ (Schmitt, 1998 p. 281). This study had an explorative nature; further research is required to gain more understanding of the incremental acquisition of vocabulary in a foreign language.

### 2.3 Vocabulary size in L2 learners of Dutch

As appears from the above sections the acquisition of words in a foreign language is an intensive activity. To acquire a vocabulary comparable or nearly comparable to that of a native speaker a lot of effort is required. For most foreign language learners this is too ambitious or simply not achievable. If we compare the Dutch vocabulary of an adult L1 speaker with that of an adult L2 speaker of Dutch it can be assumed that the vocabulary of the L2 learner, in particular in the earliest stages of second language acquisition, is much smaller than native speakers’ vocabulary. After all, adult L1 learners have had a much longer period to learn words of their language because they started to acquire the language from birth. It is worth mentioning that studies into vocabulary size in child second language learners (Verhoeven & Vermeer, 1992, in Appel & Vermeer, 1994), who do not have the disadvantage of starting learning the L2 after the critical period (Lenneberg, 1967, in Appel & Vermeer, 1994), which is nowadays denoted as sensitive or optimal period, also show that early L2 learners have a smaller vocabulary size than monolinguals.

## Chapter 3. Vocabulary size and listening comprehension

To understand a foreign language one should have knowledge of the different aspects of that language. A second language learner has to become familiar with the sounds of the second language and has to learn the meaning of words in order to expand his vocabulary. Also a certain understanding of the grammar of the foreign language contributes to its comprehension. To what extent these factors correlate with the understanding of language is of importance to people who learn a second language, and to those who work in the field of second language acquisition. The focus of this chapter is on the role of vocabulary in the understanding of language, more precisely, on the relationship between vocabulary size and listening comprehension. The first section of this chapter will discuss the cognitive processes involved in listening comprehension. Section 2.2 deals with the relationship between vocabulary and listening comprehension and gives an overview of the literature concerning this topic.

### 3.1 Cognitive processes in listening comprehension

In order to understand listening comprehension, I will first discuss some ‘central issues in language processing’ by summarizing Carroll (2004). Afterwards I describe how these processes can be applied to second language learners.

When trying to comprehend spoken language a few processes take place. These processes occur on different levels. Carroll distinguishes: the phonological level, the lexical level, the syntactic level and the discourse level. If language is processed ‘from the lowest to the highest level’, i.e. from the phonological level to the discourse level, it is called ‘bottom-up processing’ (Carroll, 2004, p. 52). Carroll states that in bottom-up processing, there is no interaction between the lower levels and the higher levels. For example, the segmentation of words at the phonological level is an isolated action that is not affected by one of the higher levels. The opposite of bottom-up processing is ‘top-down processing’. In top-down processing, interaction between lower and higher levels is possible, according to Carroll. For instance, it is possible to use contextual information to identify words or syllables. Another example of top-down processing is when knowledge of the world makes you have expectations of the content of a text and for example what the last words of a sentence will be. However, if you find out that these expectations were wrong you have to go back to the bottom-up approach to understand the text. According to Carroll, bottom-up and top-down processing can both explain language processing.

Another distinction discussed in Carroll (2004), is that between serial and parallel processing. Imagine that someone has an idea he wants to convey. The speaker has to go through different stages. He has to find the right words, put them in the right order, retrieve the right pronunciation, etc. It is possible that these actions, which are necessary to produce language, take place successively, then it is called serial processing (Fromkin, 1971 in Carroll, 2004). Carroll argues that bottom-up processing is often a serial process. Parallel processing, on the other hand, is when two or more actions are executed at the same time (Dell, 1986 in Carroll, 2004). Top-down processes are usually considered as parallel processes (Carroll, 2004). There is general agreement that listeners use both processes (Graham, 2005; Rönnberg, 2003, in Zekveld, Heslenfeld, Festen & Schoonhoven, 2006). However, whether a listener uses bottom-up or top-down processes in language comprehension depends, according to Vandergrift (2004, p. 4), 'on the purpose for listening'. For instance, performing a listening comprehension task in a foreign language requires a different way of processing than participating in a simple conversation in a first language. Kelly (1991, p. 136) comes up with some factors that determine whether spoken language is processed top-down or bottom-up. A 'weak, obscured or incomplete sound signal' obliges the listener to fall back on top-down processing. When the 'words are highly predictable' because of, for example, the listener's familiarity with the topic, only minimal bottom-up processing is needed. 'Low expectations', on the other hand, oblige the listener to focus more on the sound signal and to use bottom-up processes.

In a review paper, Rubin (1994) discusses L2 research that addresses the question whether listeners use top-down processing or bottom-up processing and how these two interact. The studies she discusses show contrasting results. This is not surprising since there are many factors, (linguistic, cognitive, textual) which influence the complex activity of listening comprehension. According to Vandergrift (2004) beginning L2-listeners must rely mainly on bottom-up processing when listening. They need to focus on the speech sounds they hear and try to distinguish words from concatenated speech. This inefficient way of processing often leads to poor understanding. Vandergrift suggests that the use of other listening strategies may be helpful for beginners. Kelly (1991, p. 136) claims about processing in L2 learners that '[...] only as they progress in proficiency in regard to the skills in the foreign language they bring into play semantic and other knowledge'. Summarizing research on the mental processes that play a role in language comprehension, Vandergrift (2004) points out that L2 learners should learn that there are different ways of processing, and how they can benefit from using them in different situations.

### 3.2 The relationship between vocabulary size and listening comprehension

In order to understand spoken language, vocabulary is required; on the other hand, listeners can also increase their vocabularies by listening. Mecartty (2000, p. 340) mentions that ‘the relationship between lexical knowledge and comprehension is bidirectional and interactive in nature’. According to Nation (2001), an L2 learner needs to understand at least 95% of the words in the input to warrant comprehension of spoken language. Hu and Nation (2000, in Nation, 2001) even argue that understanding of around 98% of the words is necessary for adequate comprehension. However, as mentioned above, listeners can learn new words while listening. Nation (2001) describes some conditions that are important to get the best results in this process. These are: ‘interest in the content of the story, comprehension of the story, understanding of the unknown words and retrieval of the meaning of those not yet strongly established, decontextualisation of the target words and thoughtful generative processing of the target vocabulary’ (p. 118). To gain insight in the relationship between vocabulary and listening comprehension, an overview of studies that address this topic is given below.

Mecartty (2000) investigated ‘lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish’. Since the present study only deals with the relationship between vocabulary size and listening comprehension, the results for reading comprehension and grammatical knowledge will be excluded in the discussion of this paper. A group of 154 students participated in the study of Mecartty. To operationalize lexical knowledge, the author developed a ‘word-association’ and a ‘word-antonym’ task. For the assessment of listening comprehension, the participants had to listen to two spoken text fragments. Afterwards, they had to answer 16 multiple-choice items, where the basic ideas and details of the texts were questioned. The results of the study show that both lexical and grammatical knowledge were significantly correlated to both reading and listening comprehension. However, a multiple regression analysis revealed that ‘only lexical knowledge explained the variance in listening comprehension’ (Mecartty, 2000, p. 323). This means that lexical knowledge is a significant predictor for listening comprehension, while grammatical knowledge is not.

Bonk (2000) investigated whether there is a correlation between ‘lexical knowledge’ and ‘listening comprehension’; another aim of this study was to find out *how much* lexical knowledge is necessary for good comprehension. To this end, 59 non-native speakers of English performed a listening test, consisting of four listening passages. The vocabulary used in the passages varied from easy to difficult. The participants had to note down everything they remembered after listening to a passage. The recall texts were given a score from 1-4,

based on beforehand-formulated criteria. From the experiment a correlation between lexical recognition and listening comprehension could be concluded. Generally spoken, for good comprehension an L2 learner needs to know the meaning of most of the words in a text. There were some participants in this study who showed good comprehension of the text although they had a low level of lexical knowledge. The author suggests the possibility of efficient listening strategies that make good comprehension of a text possible. Bonk himself criticises the design characteristics and the small sample size of his study. Consequently, the findings of this study should be interpreted with some caution.

Similar results come from a study carried out by Zareva, Schwanenflugel and Nikolova (2005). They studied the relationship between ‘lexical competence’ and ‘language proficiency’. In their study three different learner groups participated, i.e. native speakers of English, L2 advanced learners, and intermediate learners of English. The proficiency levels of the participants were determined on the basis of three tests, namely the Cambridge Certificate in Advanced English (CAE), the Cambridge First Certificate of English (FCE) and the TOEFL. The aim of this study was to determine which macro level features of lexical knowledge (i.e., ‘quantity, quality, and metacognitive awareness’) are L2 proficiency dependent. In other words, does lexical knowledge differ at different levels of language proficiency? In order to answer this question the authors described word knowledge of the participants after examining six variables. One of these six variables was vocabulary size, which was tested with words selected from a dictionary. After the test an estimation of the vocabulary size of the participant was derived. It was found that some of the six measures, including vocabulary size, show a relationship with language proficiency. They show growth if language proficiency improves.

Some more dated studies addressing the contribution of lexical knowledge to L2 listening comprehension are Richards (1983) and Rost (1990). They see lexical knowledge as indispensable for L2 learners to assign meaning to spoken language. Kelly (1991) argues that ‘lexical ignorance’ is the factor that keeps a foreign language learner back from becoming highly proficient in listening comprehension in a target language. This implies that paying attention to the words in spoken language results in better comprehension. Kelly states that to reach native-like listening comprehension, L2 courses should focus on the expansion of the L2 learner’s lexicon.

The existence of a relationship between listening and other language aspects is put forward in the following statement: ‘given the critical role of listening in language learning, students need to “learn to listen” so that they can better “listen to learn”’ (Vandergrift, 2004, p.

3). Vandergrift presents approaches to improve listening instruction in classrooms, one of which favours bottom-up processes, where the L2 learners have to work on the development of 'lexical segmentation' and 'word recognition skills'. In a more recent article, Vandergrift (2006) stresses the role of vocabulary development in improving L2 listening comprehension skills. The author states that to understand spoken language, it is not sufficient to just learn the meaning of individual words. Listening comprehension involves the recognition of concatenated words in spoken language. Therefore, students need to develop strategies to recognize words and match these words with the concepts in their lexicon, as suggested by Vandergrift.

Although most studies into this topic show that lexical knowledge is related to listening comprehension, some critical points formulated by Mecartty (2000) should be taken into consideration. Firstly she states that the amount of empirical research is not convincing. More evidence is needed to support the findings till now. Secondly, Mecartty (p. 326) stresses that: 'the factors that contribute to the comprehension process are complex ones', and that lexis is only a single factor in this process. Thirdly Mecartty calls into question that the operationalization of constructs like comprehension and knowledge of words is a problem. Not every researcher uses the same definitions for these language aspects. This results in tests that do not measure the same constructs. With these points in mind, the literature discussed in this section, which shows in general a relationship between listening comprehension and word knowledge, must be interpreted with caution.

Another language skill where lexical knowledge seems to play a role is reading comprehension. The relationship between reading comprehension and the role of vocabulary has been studied more intensively than the relationship between listening comprehension and vocabulary. In accordance with results of the latter relationship a strong connection between vocabulary and reading comprehension is also observed, as Mecartty (2000) remarks in her literature review. Nevertheless, as a consequence of the form in which a text is presented, findings from reading research cannot always be generalized to listening comprehension research, according to Bonk (2000). An example can clarify this: word recognition in a spoken text takes place via sound perception; this requires different skills than the recognition of written words, because this takes place via visual perception canals.

## Chapter 4. Testing vocabulary in a second language

This chapter discusses vocabulary testing in a second language. The first section introduces the topic by describing different kinds of vocabulary tests. A review of current developments in this area is also included. After this general introduction, the focus is on **vocabulary size** tests in particular. Which tests give researchers insight into breadth of vocabulary knowledge of L2 learners? Section 4.2 gives information about the development of a vocabulary size test for bilingual children. The participants of the present study (L2 adults) were also tested with this vocabulary size test. In section 4.3 it is described what conditions this test has to meet to be useful in the assessment of vocabulary size in adult L2 learners of Dutch.

### 4.1 Assessing vocabulary in a second language

‘Vocabulary is an essential building block of language and, as such, it makes sense to be able to measure learners’ knowledge of it’ (Schmitt, Schmitt & Clapham, 2001, p. 55). Testing vocabulary is on the one hand easy because the unit to be tested is obvious. On the other hand there are some factors that make it complicated. One of these factors is that lexical knowledge consists of many aspects, including ‘word meaning, spelling, pronunciation, grammatical form, relative frequency, collocations, and restrictions on the use of the word’ (Nation, 1990). In order to make the concept of ‘lexical competence’ more accessible, Henriksen (1999, p. 303) proposes to divide it into three dimensions. The ‘*partial-precise knowledge dimension*, *depth of knowledge dimension* and *receptive-productive knowledge dimension*’. I will now briefly clarify each of the three dimensions. The *partial-precise* distinction holds that word knowledge can vary from partial to precise. Levels of knowledge are usually operationalized using several different tasks. *Depth of knowledge*, also referred to as quality of knowledge, should be tested using a combination of different tasks, since it comprises all the knowledge aspects that someone can acquire of a particular word. The *receptive-productive knowledge* distinction can be seen as the distinction between language comprehension and production. Receptive knowledge of words is necessary for reading and listening while productive knowledge is required for writing and speaking. Although these distinctions give something to go on, the complexity of vocabulary knowledge makes testing complicated. There is a multitude of different vocabulary tests, which measure different aspects of lexical knowledge, and can be used for different purposes. Consequently, it is hard to decide which test to use in which situation. Now, a short description of five typical vocabulary tests, described by Nation

(2001) follows. The first typical vocabulary test is a 1000 word level true/false test (Nation, 1993). This test consists of sentences, which the test-taker has to judge with true or false. The next test Nation mentions is a vocabulary depth test (Read, 1995). In this test the person who is tested is given a word and then has to find words that go with the given word. Another kind of a vocabulary test is a definition completion test (Read, 1995). In this test, the test-taker has to select words from a given list to make sentences complete. The fourth test discussed is a so called sensitive multiple-choice test (Joe, 1994). This test is a typical multiple-choice test where the test-taker has to choose the best meaning of a given word. The last test Nation discusses is a translation test (Nurweni & Read, 1999). Here a translation of a given word in the first language of the test-taker is asked.

In a review article Read (2007) discusses recent developments in vocabulary assessment. The article is divided into three main topics, ‘testing breadth of vocabulary knowledge’, ‘testing depth of vocabulary knowledge’, and ‘measuring vocabulary knowledge in context’. I will now summarize the main points of this article, paying most attention to the assessment of breadth of vocabulary knowledge. Most vocabulary tests that examine breadth of vocabulary knowledge are based on word frequency lists. Since computer corpora are more and more developed they are very helpful in the composition of word frequency lists. However, Read (p. 109) claims that a ‘definitive word frequency list’ does not yet exist, in the case of English. A large number of test items is required to make a reliable vocabulary size test. For practical reasons a simple test format is therefore desirable. Read (p. 113) argues for the Yes/No format, since it is an ‘*informative*’ and ‘*cost-effective*’ test. Concerning tests that measure depth of vocabulary knowledge, Read argues that there are some useful tests (for instance the word association test, Read, 1993, 1998) but generally, there has been made little progress in developing tests, as well as in defining the concept. An explanation for the lack of well-developed tests is possibly the complexity of the construct of word knowledge. Read (2007) argues that since the advent of the communicative approach in language teaching and testing, testing vocabulary in isolation is not sufficient anymore. Therefore he points to the importance of testing vocabulary in context.

This section continues with a discussion of some less and more popular vocabulary size tests that have been developed to test quantity of vocabulary knowledge. First some tests for English L2 learners are described, followed by a description of two tests that are used in the assessment of vocabulary in Dutch L2 learners.

A widely used vocabulary test is the Vocabulary Levels Test (Nation, 1990). Nation states that the test is a helpful tool in deciding on which front the learner needs support with

vocabulary learning. As indicated by its name the test is composed of levels. These levels are the 2.000- and 3.000-word levels, the university word level and the 5.000- and 10.000-word levels. The university word level contains words from the ‘University Word List’ (Nation, 1990, Appendix 2) and therefore ‘represents one type of specialized vocabulary’ (Nation, 1990, p. 261). The Vocabulary Levels Test examines receptive word knowledge. The test-taker has to match a word with a definition by writing the number of the word on the line next to the corresponding meaning. Below an example of a completed test item (Nation, 1990, p. 264):

1. business
2. clock              6   part of a house
3. horse              3   animal with four legs
4. pencil             4   something used for writing
5. shoe
6. wall

Schmitt, Schmitt & Clapham (2001) claim that although the test is used by many researchers and language teachers, it has never been accurately validated. Therefore, these researchers made an attempt to test the validity of two new versions of the test. From their examination it appears that these versions show valid results. According to the authors this is only a first step in exploring the Vocabulary Levels Test, therefore they suggest further research. On the basis of the Vocabulary Levels Test, Laufer & Nation (1999) produced a test of ‘controlled productive ability’, named the Productive Vocabulary Levels Test. The test resembles the Vocabulary Levels Test in that it employs the same vocabulary frequency levels. It consists of sentences in which the test-taker has to fill in a word. The first letters of the word are given in order to prevent the test-taker to fill in another semantically correct word. The format of the test is ‘controlled’ since the design requires the use of specific words, in contrast with, for instance, writing tasks where free productive ability is tested. In their article the authors claim that the Productive Vocabulary Levels Test is valid, reliable and practical. They suggest to use it combined with the receptive levels test and a free productive ability test. In my opinion this combination is a powerful instrument to examine vocabulary knowledge. The Yes/No vocabulary test is an instrument that tests receptive vocabulary knowledge (Beeckmans, Eyckmans, Janssens, Dufranne, & Van de Velde, 2001). For this test L2 learners have to decide if they know the meaning of the words that are presented to them. According to Beeckmans et al. (2001, p. 238) the test is ‘easy to construct, administer and correct and the format permits the testing of a large number of words in a short span of time’. There are

however some shortcomings because closer examination of the test shows that it does not meet the requirements of reliability in its current version (Beeckmans et al., 2001).

To test vocabulary in Dutch L2 learners several tests exist. In the assessment of vocabulary size in L2 adults the Vocabulary Size Placement Test (VSPT), part of Dialang (Alderson & Huhta, 2005), is a useful tool. This test has a Yes/No format and consists of a list of 75 words (50 real words and 25 nonsense words); the test-taker is asked to judge for every word whether it exists in the target language or not. He can do this by giving *yes* if he thinks the word exists, or *no* if the word is a non-word in his opinion. Directly after the test, the participant receives feedback. This feedback implies an ability level from 'very low' to 'indistinguishable from a native speaker'. In assessing vocabulary knowledge in L2 or bilingual children, the 'Taaltoets Alle Kinderen' (TAK, Verhoeven & Vermeer, 2001) is commonly used. The TAK has a receptive and a productive vocabulary test. In the receptive test, ('Passieve woordenschattoets') the child has to find the picture (out of 4 pictures) that depicts the word that is given by the instructor. In the productive task ('Woordomschrijvingstaak'), the child is asked to define the words given by the instructor. After close examination of the TAK Brasileiro (2009) decided that this test was not suitable to test vocabulary size in the bilingual children participating in her study. Therefore she, in cooperation with Viellevoeye Grooters, developed a new vocabulary test. The development of this test that is based on the TAK, is described in the next section.

#### 4.2 Development of the Vocabulary Test by Brasileiro & Viellevoeye Grooters

In 2005, Brasileiro and Viellevoeye Grooters developed a test to examine vocabulary size in bilingual, Brazilian Portuguese-Dutch, children. In Brasileiro's PhD project, where the language proficiency of this particular group was examined, a suitable vocabulary test was needed. Since no test to assess vocabulary size in this particular group existed, an attempt was made to develop such a test. After the first stage in which they collected words and pictures, a pilot test with 60 children, between 5 and 7 years old and 8 adults was conducted. The results of this pilot test were analysed. This analysis consisted of a calculation of the significance between the results of the different age groups, a calculation of the percentage of correct answers for each age group, an Index of Discrimination test and a Reliability Analysis. The Index of Discrimination is an indicator for comparing best and worst performing groups. The Reliability Analysis was used to determine the reliability of the test. Then 8 native Dutch adults were tested to verify if the incorrect answers of the children were caused by 'ignorance of the children' or by 'ambiguity of the pictures' in the test (Viellevoeye Grooters, 2005, p.

13). After this preliminary examination the test was composed. During the selection of test items, items that did not meet the requirements were rejected. There were three reasons to reject an item. The first reason for rejection was too much contrast or similarity between the Dutch and the Portuguese word. As Brasileiro wanted to test the difference between Dutch and Portuguese vocabulary size, items that show too much contrast or similarity in word form were not useful. Secondly, items that did not show an upward trend for the three age groups in their study were also rejected. The third reason to remove an item was in case of an ambiguous picture. The final version of the test consists of 40 items and can be extended with six extra items for children with high scores. These six extra items are not of importance for the present study, because we did not test our participants with them. The design of the vocabulary test is more extensively described in section 6.2. Section 4.3 describes the conditions the test has to meet to be useful for L2 adults.

#### 4.3 Requirements of the vocabulary test for adult second language learners of Dutch

This section starts with a short description of the similarities and differences between the two target groups, i.e. bilingual children and adult L2 learners. Afterwards it is described what requirements the test has to meet to be useful in the assessment of L2 adults' vocabulary.

The two groups are similar in that they both have small vocabularies in comparison to adult L1 speakers' vocabulary. They are both acquiring Dutch, either as L2 or simultaneously with acquiring Portuguese. Two main differences between the children tested for Brasileiro's study and the adults tested for the current study are described here. In the first place both groups vary in age. All participants in the current study are older than eighteen, while the children were between 5 and 7 years old. The second difference concerns the L1 of both groups. The children were all bilinguals, learning Dutch and Portuguese from birth on. In contrast, the adults have Portuguese or Spanish as their L1 and learn Dutch as L2. Consequently, the setting in which they learn Dutch is different. To judge the usefulness of the vocabulary test for adults, we have to take these differences into account.

Now an enumeration of the requirements of the vocabulary test for L2 adults follows. These requirements are similar to the requirements formulated for the items of the test for bilingual children in Brasileiro's project.

- The pictures that are inserted in the test should be clear. Picture items that are ambiguous according to the control group have to be rejected.
- The Index of Discrimination (see chapter 6.4 and 7) of the test items has to be a positive value; test items with a negative value should be rejected.

- The percentage of correct answers should show an upward trend for the different levels of difficulty. This will be analysed for the different Dialang level groups.
- The test has to be reliable. If Cronbach's  $\alpha$  is  $\geq .60$  the reliability of the test should be considered as acceptable. A test item that causes an improvement of .05 or more, if deleted, should be rejected (see chapter 6 and 7).

## Chapter 5. Research Questions and Hypotheses

In brief, this study researches the vocabulary of second language learners of Dutch. Based on already existing data, obtained for a longitudinal study into the acquisition of the Dutch vowel system, three research questions are formulated below. The first aspect to be analyzed on the basis of these data are the scores on a vocabulary test, for three different groups of L2 learners and native speakers of Dutch. This leads to research question 1. The second research question focuses on the relationship between vocabulary size and listening comprehension.

Furthermore I will try to answer the question if this recently developed vocabulary test, which was developed to test bilingual children, is also useful in the assessment of L2 adults. The corresponding question is formulated in research question 3.

### *Research Questions*

1. Are the scores on the vocabulary test comparable for the four different groups (Spanish (SP), Brazilian Portuguese (BP), Latin American Spanish (LAS) and Dutch (DUT) participants)?
2. Do the scores of the participants on the vocabulary test correlate with their listening comprehension level?
3. Is the vocabulary test (Brasileiro, 2009), developed to test bilingual children's lexical abilities, also useful in the assessment of adult L2 learners?

Research question 3 is divided into three sub questions:

- How do the test items discriminate between the best and worst performing group (in this case groups A and C on the listening comprehension test, see chapter 6)?
- Does the percentage of correct answers show an upward trend for the different levels of difficulty in the vocabulary test?
- Is the vocabulary test a reliable test?

### *Hypotheses*

1. It is hypothesized that the scores on the vocabulary test will be comparable for the three research groups. The fact is that the groups are matched for listening comprehension level (more information about this procedure in chapter 6). Since listening comprehension level and vocabulary size seem to correlate (see for a

literature review chapter 3), I expect no significant differences between the three L2 groups. If the results do however show significant differences, these are probably due to other linguistic or non-linguistic factors that impact the acquisition of a second language. The comparisons between the three research groups and the Dutch control group are expected to yield significant differences. It is a well-known fact that people who learn a foreign language, especially those who just started learning the second language (in this case Dutch), do not have vocabularies that are comparable with the vocabularies of native speakers of the target language. It is therefore expected that the participants in the research groups are inferior to the subjects in the control group. The Dutch participants in the control group are supposed to make no mistakes at all on the test items, since this test is developed to test (bilingual) children and does not contain items that are too complex for L1 adults.

2. Previous studies that investigate the interaction between vocabulary size/lexical knowledge and listening comprehension in a second language (Richards, 1983; Rost, 1990; Mecartty, 2000; Bonk, 2000) all demonstrate in a certain way, a relationship between these two variables. Although these studies use different methods and investigate different L2 populations, the results display a certain resemblance. Therefore it is expected that the scores on the vocabulary test correlate positively with the listening comprehension scores of the participants.
3. It is doubtful whether the vocabulary test is useful in the assessment of adult L2 learners. On the one hand, the vocabulary of children and the vocabulary of adult L2 learners resemble each other in a way that they both have small vocabularies in comparison to L1 adults. On the other hand the test contains words that are more likely to be present in a child's vocabulary than in the vocabulary of an adult. It concerns the words in the domains of animals and school. Therefore it is hypothesized that the design of the test is not (yet) completely suitable for L2 adults.

Besides that it should be remarked that this test is not suitable for second language learners in all stages of acquisition, as it is not for children of all ages. As this test was developed to test children in the age between 3-6, it is expected to be only useful to assess L2 learners with a low or intermediate language proficiency level. Second language learners in a more proficient stage of language learning will show ceiling effects on the test. Answering the three questions formulated under research question 3 will however decide if this hypothesis is well grounded or not.

## Chapter 6. Method

### 6.1 Participants

#### *Experimental groups*

57 Adult second language (L2) learners of Dutch participated in this study. They were divided into three groups according to their native language variety, namely 15 Peninsular Spanish (PS), 29 Latin American Spanish (LAS) and 13 Brazilian Portuguese (BP). The Spanish-speaking Latin Americans came from various countries, namely Peru, Puerto Rico, Argentina, Venezuela, Chile, Mexico, Colombia, Uruguay, Honduras, Ecuador and Cuba. The Portuguese-speaking Latin Americans came from Brazil. The participants were recruited for a longitudinal L2 research study into the acquisition of the Dutch vowel system by Spanish and Portuguese learners. As part of this longitudinal study, they performed several tests, most of which investigated their vowel perception. Each test session started with a Dutch vocabulary test and a listening comprehension test. The results of the vocabulary test and the listening comprehension test are the topic of the present study. The participants are between 20 and 55 years old. Their mean age is 31,5 (s.d. 7,0). Most of them filled out a background questionnaire in their native language, which contained questions related to their length of residence in the Netherlands and their self-assessment of their general Dutch comprehension on a scale from 0-7 (0 = no understanding of Dutch, 7 = understanding of Dutch at a native-like level).

To answer research questions 2 and 3, posed in the previous chapter, the results of all 57 L2 participants were considered. To answer research question 1, a selection of the three experimental groups was made. For this selection, I decided to match the participants in the three experimental groups for listening comprehension level. This resulted in three groups of ten participants each. The listening comprehension level of the participants was determined by the results of a test that is part of Dialang (Alderson & Huhta, 2005). ‘Dialang is an on-line diagnostic language assessment system’ that tests language proficiency, containing ‘tests of reading, listening, writing, vocabulary and structures in 14 European languages’ (Alderson & Huhta, 2005, p. 301). After every test the system determines your level and provides feedback about your language skills. There are six levels, A1, A2, B1, B2, C1, C2, where A represents beginning proficiency, B intermediate and C advanced. Each of the selected groups had four participants who scored A, one participant who scored B, and five participants who scored C.

This group selection was done in order to make sure that listening comprehension was not the factor, that caused differences between the three groups.

### *Control group*

To compare the scores of the L2 learners with the scores of native speakers of Dutch, 11 Dutch adults performed the same vocabulary test. They were between 21-26 years old, with a mean age of 23,4 years (s.d. 1,6). The results of the Dutch natives also served as a check of the clarity of the pictures used in the test. A further discussion of this issue will be presented in chapter 8.

## 6.2 Instruments

### *Vocabulary Test*

A vocabulary test, which measures active vocabulary size, was used for this study. The test was recently developed and used to examine the vocabulary size of bilingual, Portuguese-Dutch, children (Brasileiro, 2009). It consists of 40 picture items (Appendix A), which can be divided into four domains, namely: animals, food, school and neutral, each domain consisting of 10 items. Brasileiro and Viellevoye Grooters (2005) point out that the reason to divide the items into different domains is that bilingual populations often use each of their two languages for certain specific domains. That is, it is possible that they have a larger vocabulary in a certain domain than in other domains. For the participants of the present study the possibility that some domains are more developed than others also exist. Most pictures have a single correct answer, but there are also some picture items where more than one answer is accepted.

Brasileiro and Viellevoye Grooters (2005) divided the words in the test into five frequency levels based on the frequency lists of the Corpus Gesproken Nederlands (Piepenbrock, 2000). Frequency 1 represents low-frequency words and frequency 5 high-frequency words. After a pilot test, Brasileiro and Viellevoye Grooters decided to turn the frequency levels into levels of difficulty. In the current version of the test, the words in the domains represent five different levels of difficulty. Each level of difficulty consists of 2 items. Item 1 and 2 are the two most difficult items and belong to the first level of difficulty, item 3 and 4 to the second level of difficulty, and so on. This classification is based on the answers of the children in the pilot test.

### *Listening comprehension test*

To test listening comprehension, participants performed a listening comprehension test which was part of an European Language Assessment i.e. Dialang (Alderson & Huhta, 2005). The Dutch version of the test consists of 30 fragments, presented auditorily; after each fragment the participant has to answer a multiple-choice question. These questions are formulated in Dutch; participants are asked to read the question and possible answers prior to hearing the fragment. After the participant completes all 30 questions, the tool calculates the corresponding score, which is linked to one of the six levels mentioned in section 6.1.

### 6.3 Procedure

The participants were tested one at a time in a quiet room. Each testing session started with the vocabulary test. They were asked to name the Dutch word for each of the 40 pictures the experimenter showed them. There was no time-limit and it was possible to correct answers. The test was audio-taped in most of the cases; in the rest of the cases the experimenter wrote down the answers given by the participant during the test. Since an answer to each of the 40 test items can only be right or wrong, each participant will have a vocabulary score between 0 and 40. If a participant gave no answer the score was 0.

After the vocabulary test participants performed the listening comprehension test on the computer. Before the experiment started they were told that they had to read the question and possible answers before listening to the fragments on headphones; after each fragment they had to answer the question before going to the next fragment. They performed this test individually without intervention of the experimenter. Important to mention is that the L2 adults performed both the vocabulary and listening comprehension test, whereas the Dutch participants only performed the vocabulary test.

### 6.4 Data-analysis

This study is based on the results of two tests (i.e. a vocabulary test and a listening comprehension test). In order to examine vocabulary size all participants were tested with a vocabulary test. After the test a total score for each participant was calculated, just as a mean score for each of the four groups. The L2 participants also performed a listening comprehension test. This resulted in a listening comprehension level, calculated by Dialang. This section will continue with a description of how the data were analysed.

To determine whether the three experimental groups and the control group show similar results on the vocabulary test, the Kruskal-Wallis test (Field, 2005) was used. This test

was conducted with SPSS 16.0. The Kruskal-Wallis test compares the mean scores of several independent groups. This is a non-parametric test, useful for analysing non-normally distributed data. Since the Kolmogorov-Smirnov test (Field, 2005), calculated with SPSS 16.0, shows that the data obtained with the vocabulary test are non-normally distributed ( $p < .05$ ), this test seems to be most appropriate. The Kruskal-Wallis test however, only tells if there exists a difference between groups, but not between which groups this possible difference is found. Therefore a post-hoc test was conducted afterwards, in this case a test described by Siegel & Castellan (1988, in Field, 2005). This test is similar to the Mann-Whitney test (Field, 2005).

In order to answer research question 2, we first determined whether the data are normally distributed. The Kolmogorov-Smirnov test shows that the data on the listening comprehension test are non-normally distributed ( $p < .05$ ). Therefore a non-parametric test was used to test for the correlation between vocabulary size and listening comprehension. Kendall's tau test (Field, 2005) seems to be the most appropriate, since our data set is small and has a large number of 'tied ranks' (p. 131).

To decide whether the vocabulary test used in the current study is useful in the assessment of L2 learners of Dutch, a few analyses were done. First the Index of Discrimination (Baker, 1989) was calculated for each of the 40 test items. With this test we determined how the test items discriminate between group A (the bottom group) and C (the top group) on the listening comprehension test. The formula to calculate the Index of Discrimination (Baker, 1989, p. 52) is:

$$\text{Index of Discrimination} = t / N_1 - b / N_2$$

t = number of correct responses in the top group

b = number of correct responses in the bottom group

$N_1$  = number of candidates in the top group ( $N_1=19$ )

$N_2$  = number of candidates in the bottom group ( $N_2=29$ )

Secondly, the percentage correct answers per level of difficulty was calculated. This calculation makes it possible to see if the subdivision in levels of difficulty in the current composition of the test is appropriate for L2 adults. Increasing percentages from level of difficulty 2 to 6 would confirm this.

Lastly, the reliability of the test was examined. With SPSS 16.0 a Reliability Analysis can be conducted. De Heus, Van der Leerden & Gazendam (1995, p. 186) give the following rules of thumb about the results of the analysis: they state that ‘een schaal ‘goed’ is als  $\alpha \geq .80$ , ‘redelijk’ als  $.60 \leq \alpha < .80$  en ‘slecht’ [...] als  $\alpha < .60$ .’ *‘A scale is ‘good’ if  $\alpha \geq .80$ , ‘reasonable’ if  $.60 \leq \alpha < .80$  and ‘bad’ [...] if  $\alpha < .60$ .’* (my translation). Furthermore, it is possible to determine whether a test item has a positive or negative contribution to the reliability of the test. De Heus et al. (1995) state that every test item that causes an improvement of .05 or more, if deleted, should be removed.

## Chapter 7. Results

To answer the three research questions formulated for this study, the scores on the two language tests were analysed. This chapter describes the results of these analyses. To answer the first research question, we compared the mean scores of the four different groups (DUT, PS, BP and LAS) on the vocabulary test, using the Kruskal-Wallis test. First, the total scores for all participants that were selected to answer research question 1 (see Appendix B) were calculated. Then these scores were ranked, not taking into consideration to which of the four groups the score belongs. After ranking the data the ranks for each group were collected and added up. Then can be determined whether the groups score significantly different on the vocabulary test by calculating the test statistic H, with the following equation (Field, 2005, p. 544):

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N + 1)$$

$R_i$  = the sum of ranks for each group

$N$  = total sample size

$n_i$  = the sample size of a particular group

$k$  = number of groups

This test shows that the mean scores are significantly different ( $H(3) = 22.84, p < .01$ ). A test described by Siegel & Castellan (1988, in Field, 2005) was used to follow up this finding.

This test ‘is essentially the same as doing Mann-Whitney tests on all possible comparisons’ (Field, 2005, p. 551). For the data of the present study it shows that at ( $p < .01$ ) the critical difference is 16.99. Table 7.1 shows the differences between the mean ranks for all groups.

Table 7.1 The differences between the mean ranks for the groups DUT, PS, BP and LAS.

<b>Comparison</b>	<b>Difference between mean ranks</b>
DUT-PS	22.45*
DUT-BP	19.35*
DUT-LAS	18.00*

LAS-PS	4.45
BP-PS	3.10
LAS-BP	1.35

\* significant value

Values that are greater than the critical difference (16.99) represent significant differences. As appears from the table the differences between the groups with language background DUT-BP, DUT-LAS and DUT-PS are significant. Thus, the control group (DUT) differs significantly with the experimental groups (BP, LAS and PS). Between the three experimental groups no significant differences are found.

For research question two, the correlation between the score on the vocabulary test and the score on the listening comprehension test was calculated for the L2 participants. Dutch participants were not included in this analysis since they did not perform the listening comprehension test. The analysis resulted in a correlation coefficient of  $\tau = .65$ ,  $p$  (one-tailed)  $< .01$ . A scatter plot of this correlation is presented in Figure 7.1. The upward trend indicates a positive relationship between vocabulary size and listening comprehension level.

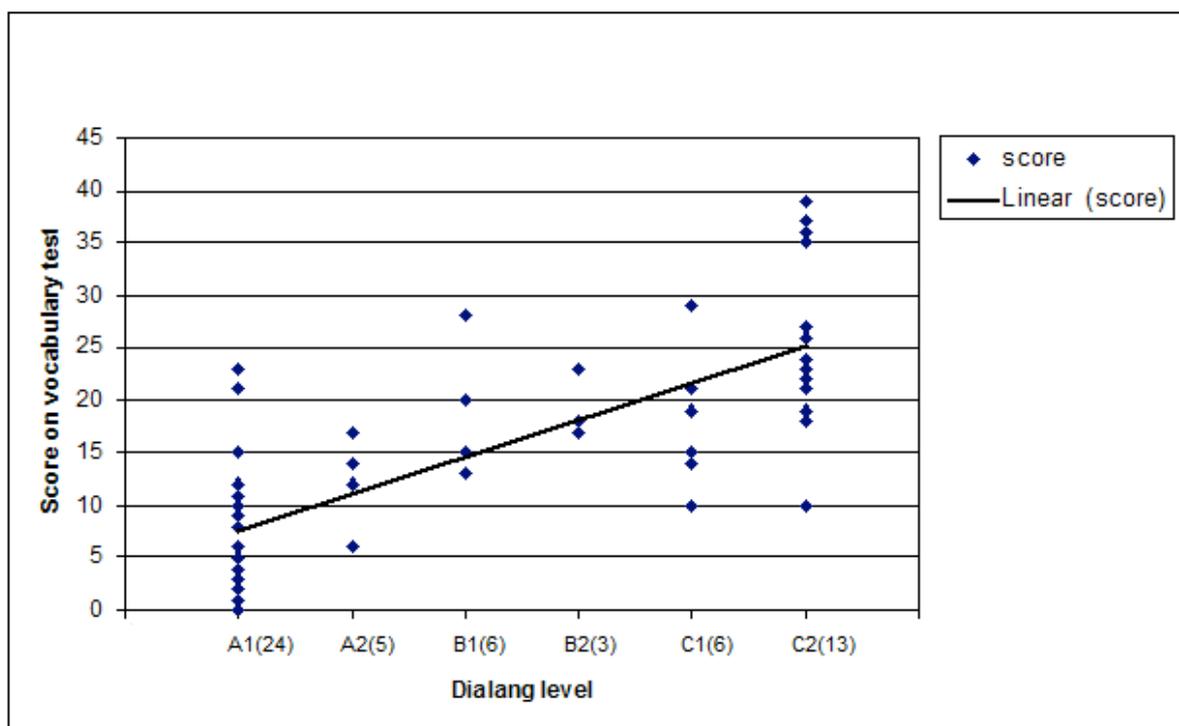


Figure 7.1. Scores on the vocabulary test and Dialang levels for all 57 L2 participants. Between parentheses: number of participants.

In order to answer research question 3, the scores of the L2 participants on the vocabulary test were analysed, also here the Dutch participants were excluded from the analysis. First the

Index of Discrimination (Baker, 1989, p. 52) was calculated. This analysis resulted in a value between  $-1$  and  $1$ . A negative value would indicate that the bottom group (group A) had more correct answers on a particular item than the top group (group C). A value of  $-1$  would indicate that all participants in the bottom group answered this item correctly, while all participants in the top group failed on this item. In Appendix C a table of all items with their corresponding Index of Discrimination is given. The table shows that all Index of Discrimination values are positive. This means that all items in the test discriminate well between group A and C. In other words, all test items show an upward trend between group A and C on the listening comprehension test. Note that some items score much higher than others, with a lowest Index of Discrimination of  $0.08$  and a highest Index of Discrimination of  $0.66$ .

Secondly, the percentage of correct answers per level of difficulty was calculated. The results are presented in Table 7.2 and depicted in Figure 7.2. The first level of difficulty is missing in the table, because items with level of difficulty 1 are added as extra items in the original test for children with very high scores. These items are not part of the test we used in the assessment of L2 adults vocabulary.

Table 7.2 Percentages of correct answers per level of difficulty and Dialang level

	Level of difficulty** →	< difficult				easy >
<b>Dialang level***</b>		<b>2 (8)</b>	<b>3 (8)</b>	<b>4 (8)</b>	<b>5 (8)</b>	<b>6 (8)</b>
↓	<b>A (29)</b>	15.09	6.90*	22.84	22.84*	33.19
	<b>B (9)</b>	36.11	25.00*	47.22	47.22*	66.67
	<b>C (19)</b>	46.71	46.05*	58.55	59.87	82.24
	Mean	32.64	25.98*	42.87	43.31	60.70

\* Percentages that interrupt the upward trend between the levels of difficulty

\*\* Between parentheses: number of items

\*\*\* Between parentheses: number of participants

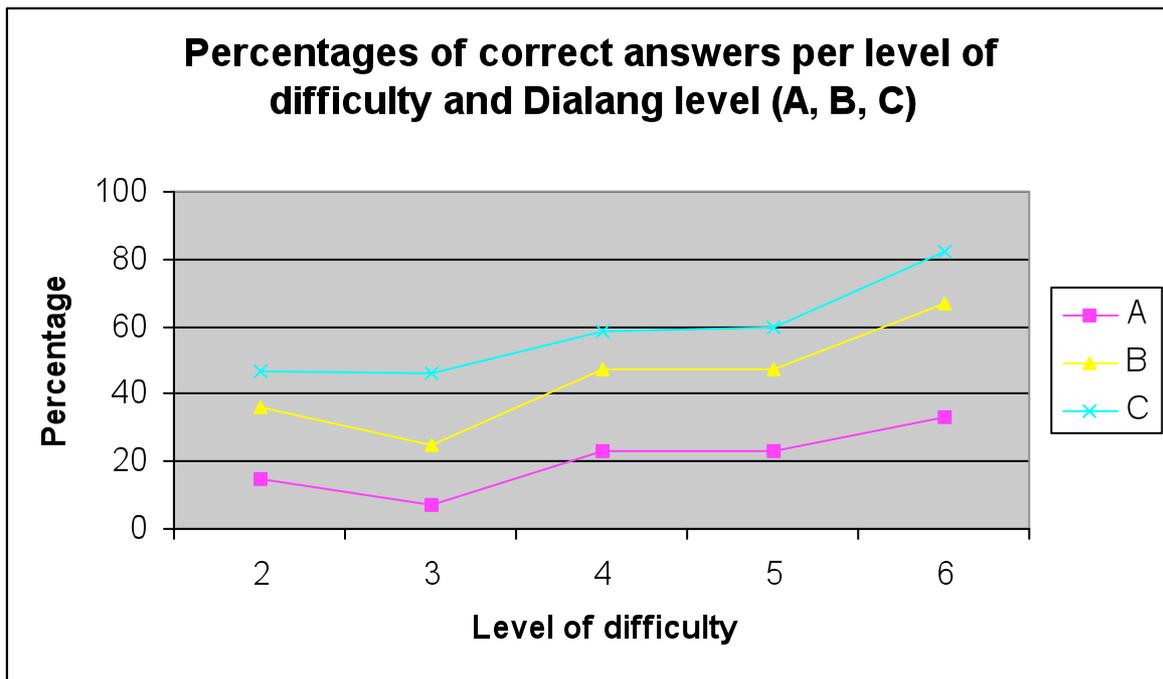


Figure 7.2 Percentages of correct answers per level of difficulty and Dialang level

As appears from the table and the graph, the percentages of correct answers do not show an upward trend between the levels of difficulty, since there are some percentages that interrupt the line. These are marked with an asterisk. It concerns the percentages at level of difficulty 3 and 5, as can be concluded from the table. An upward trend between the different Dialang levels is however perceptible. On all 5 levels of difficulty the percentages of correct answers increases from Dialang level A to C. An explanation of these results is part of the discussion in chapter 8.

Finally, a Reliability Analysis was carried out. This yielded an overall  $\alpha$  of .95. As suggested by De Heus, Van der Leerden & Gazendam (1995, p. 186) a scale with an  $\alpha \geq .80$  is a 'good' scale. Furthermore, De Heus et al. suggest that test items that cause an improvement of .05 or more, if deleted, should be removed. In case of the present study there were no test items that make the alpha increase with .05 or more.

## 8. Discussion

The study presented in this thesis was based on data obtained for a longitudinal project of Paola Escudero. The data we are talking about are the scores on a vocabulary test and on a listening comprehension test. On the basis of these data three different aspects of vocabulary acquisition in Dutch as a second language were investigated. First, the vocabulary scores of three different learner groups (Peninsular Spanish (PS), Brazilian Portuguese (BP) and Latin American Spanish (LAS)) that participated in the experiment were compared to each other and afterwards with the scores of a Dutch control group. Secondly, the connection between vocabulary size and listening comprehension level was examined. The third aspect that was analyzed was the usefulness of the vocabulary test as a tool in the assessment of L2 adults' vocabulary. Various analyses were carried out for an objective judgment of the test.

As expected, the comparison between the vocabulary scores of the three L2 groups and the Dutch control group resulted in significant differences. The mean group score of the Dutch participants was significantly higher than the mean group scores of all three L2 groups. However, none of the comparisons between the three L2 groups (which were matched for listening comprehension level) resulted in significant differences. Apparently, no other factor (linguistic or extra-linguistic) affected the scores on the vocabulary test enough to cause significant differences between the L2 groups. Possible factors that could have caused a significant difference between the experimental groups are for instance: either or not participating in a Dutch language course and the reason to take up residence in the Netherlands. The absence of significant differences between the experimental groups can possibly also be attributed to the small group sizes. To assure that the three L2 groups were similar concerning listening comprehension level we had to select people from the total group of participants; this resulted in small experimental groups. It should be noted that the Dutch participants did not perform the listening comprehension test, and consequently it was not possible to match this group for listening comprehension level. This weak point in the design of the experiment has most probably affected the outcome of the statistical test. The same experiment with larger groups, a Dutch control group matched for listening comprehension level and the possibility of controlling for more factors would be interesting. Another point of criticism is the use of this particular test to test vocabulary size in L2 adults. Whether this test is a useful tool in the assessment of adults and if the results are meaningful should appear from the discussion of research question 3 below.

Concerning research question 2, the results show a statistically significant interaction between vocabulary size and listening comprehension level. As expected the correlation is positive, which means that high scores on the vocabulary test go together with high scores on the listening comprehension test, and conversely, low scores go together with low scores. The results of the present study agree with earlier studies that show a relationship between lexical knowledge and listening comprehension (Richards, 1983; Rost, 1990; Mecartty, 2000; Bonk, 2000). Mecartty's study showed not only a correlation between lexical knowledge and listening comprehension, the results also prove that lexical knowledge is a significant predictor for listening comprehension level. The outcomes of the present study show a correlation coefficient of .65, which is a reasonable correlation. This coefficient does however not indicate causality, that is, it cannot be concluded that increasing vocabulary size results in better listening comprehension. Due to missing data of the participants, this study only analyses the relationship between vocabulary size and listening comprehension level. A study that examines the relationship between vocabulary size and listening comprehension in combination with other variables, for example length of residence or reading comprehension level, would be more meaningful. Especially if an attempt is made to explain the variance in both aspects by using (multiple) regression analyses. Another suggestion for future research is to make use of vocabulary tests that test different aspects of vocabulary knowledge and see how the outcomes correlate with listening comprehension level. The test used in our experiment tested merely productive vocabulary knowledge from words in isolation.

Chapter 4.3 describes the requirements of the vocabulary test for adult L2 learners of Dutch. The analyses of the test results will be discussed with these requirements in mind. The first requirement says that pictures that are not clear according to the control group should be rejected. This requirement was formulated as we try to avoid participants making mistakes because of ambiguous pictures. It was expected that the test did not contain unclear pictures, because these were already rejected when the test was composed in 2005. However, 4 of the 11 participants in the control group of our experiment reported that the picture of the 'apple' was unclear. They confused it with a 'paprika' or a 'tomato'. In the control group in Brasileiro and Viellevoeye Grooters' experiment there were also two participants who noticed the ambiguity of the picture of the 'apple'. At that time, it was decided not to replace this item since the children in their experiment did not make mistakes with it. Now that this picture again creates confusion it is suggested to replace this particular picture with an unambiguous one. Secondly, items with a negative Index of Discrimination should be removed. A negative value indicates that an item does not discriminate well between the best and worst performing

group, in the case of this experiment between group A and C on the listening comprehension test. The analysis of the test results by calculating the Index of Discrimination yielded positive values for all test items. This means that we see, for all test items, an upward trend between group A and C, indicating that group C always scores better than group A. The outcomes of this analysis suggest that, based on this point, the test is useful for this particular group of L2 adults. To say with certainty that these outcomes can be generalised to a bigger population of Dutch L2 learners, I recommend doing the same analysis with the results of other L2 learners of Dutch. Requirement 3 states that the percentages of correct answers should show an upward trend between the different levels of difficulty. From Table 7.2 and Figure 7.2 it appears that the results do not meet this requirement. Where the percentages of correct answers showed an upward trend for the 5-7 year old bilinguals in Brasileiro's project, the upward trend in the current experiment is interrupted at level of difficulty 3 and 5. Note that this holds for all three Dialang levels and the mean at level of difficulty 3 and Dialang level A and B at level of difficulty 5. This implicates that the L2 adults experience words of level of difficulty 3 as more difficult than the words in level of difficulty 2 and words of level of difficulty 5 as more difficult than words of level of difficulty 4. The interruption of the upward trend between level of difficulty 4 and 5 is hardly remarkable because the percentages at Dialang level A and B are equal. However, the differences between level of difficulty 2 and 3 are more obvious. Apparently the L2 adults acquire the tested words in a different order than the bilingual children. This can be due to the fact that the language input of the L2 adults differs from the input that the bilingual children receive. Fourthly, the reliability of the test was determined. From this analysis it can be concluded that this test is a reliable test in its current shape. A scale with an  $\alpha$  of .95 is according to De Heus, Van der Leerden & Gazendam (1995) a 'good' scale. None of the test items has to be deleted, since they all contribute to the reliability of the test. In view of the results of these analyses, the test seems to meet most of the requirements. However, to be useful in the assessment of adult L2 learners some changes have to be made. To begin with, a picture that is not ambiguous should replace the picture of the 'apple'. Secondly, it should be determined how the problem with the levels of difficulty can be solved.

## 9. Conclusions

The following can be concluded based on the (statistical) results obtained: 1) The Dutch participants (DUT) score significantly higher on the vocabulary test than the three L2 groups (PS, BP, LAS). No significant differences between the L2 groups were found; 2) Vocabulary size and listening comprehension level correlate significantly. A positive relationship between these two factors was found; 3) The vocabulary test does not meet all the requirements formulated in section 4.3. Therefore it can be concluded that the test in its current version is not useful in the assessment of adult L2 learners. However, the results of the analyses show that only a few changes would make the test useful to test vocabulary in adult L2 learners.

## Summary

This study examined three aspects of vocabulary acquisition in second language (L2) learners of Dutch: (1) vocabulary size in L2 learners with different language backgrounds; (2) vocabulary size and the connection with listening comprehension level; (3) an analysis of the vocabulary test used in the experiment. The participants were 57 adult L2 learners of Dutch. They were divided into three groups according to their native language variety, namely 15 Peninsular Spanish (PS), 29 Latin American Spanish (LAS) and 13 Brazilian Portuguese (BP). Furthermore there was a control group that consisted of 11 Dutch (DUT) participants. The L2 participants completed a vocabulary test and a listening comprehension test. The Dutch participants only performed the vocabulary test. Several (statistical) tests showed the following results: 1) the vocabulary scores of the three L2 groups differ significantly from the vocabulary scores of the Dutch control group; whereas the comparisons between the three L2 groups did not result in significant differences. 2) Calculating a correlation coefficient between vocabulary size and listening comprehension level showed that these two variables correlate significantly. 3) Different analyses that were conducted with the data obtained with the vocabulary test revealed that the test can be useful in the assessment of L2 adults' lexical abilities, at least after some adjustments are made.

## References

- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22: 301-320.
- Appel, R. (1996). The lexicon in second language acquisition. In P. Jordens & J. Lalleman (eds.). *Investigating second language acquisition*. (pp. 381-403) Berlin: Walter de Gruyter & Co.
- Appel, R., & Vermeer, A. (1994). *Tweede- taal verwerving en tweede-taal onderwijs*. Bussum: Coutinho.
- Baker, D. (1989). *Language Testing; A critical survey and practical guide*. London: Edward Arnold
- Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the yes/no vocabulary test: some methodological issues in theory and practice. *Language Testing*, 18: 235-274.
- Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14: 14-31.
- Brasileiro, I. (2009). The effects of bilingualism on children's perception of speech sounds. *PhD study*. Utrecht: Universiteit van Utrecht.
- Brown, R., & McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5: 325-327.
- Carroll, D. W. (2004). *Psychology of language*. USA: Wadsworth/Thomson Learning.
- Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8: 240-247.
- Dell, G.S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93: 283-321.
- Dong, Y., Gui, S., & Macwhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Language and Cognition*, 8: 221-238.
- Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.
- Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47: 27-52.
- Gass, S.M., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Goulden, R., Nation, P. and Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11: 341-63.
- Graham, S. (2005). Listening comprehension: The learners' perspective. *System*, 34: 165-182.
- Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education*. USA: Cambridge University Press.
- Henriksen, B. (1999). communicative approach in language teaching. *SSLA*, 21: 303-317.
- de Heus, P., van der Leeden, R., & Gazendam, B. (1995). *Toegepaste data-analyse; technieken voor niet-experimenteel onderzoek in de sociale wetenschappen 's-Gravenhage*: Reed Business Information
- Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13: 403-430.
- Joe, A. (1994). The effects of text-based tasks on incidental vocabulary learning. *Unpublished MA thesis*. New Zealand: Victoria University of Wellington.
- Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *International Review of Applied Linguistics in Language Teaching*, 29: 135-150.
- Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary*:

- Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16: 33-51.
- Lenneberg, E. (1967). *Biological foundations of language*. New York: John Wiley.
- Mecartty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, 11: 323-348.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House Publishers.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, 18: 161-175.
- Payne, M.E. (1988). Vocabulary learning strategies used by ESL students and their relationship to perceptual learning style preferences. *M.A. Thesis*, Linguistics Department, Brigham Young University, Provo, Utah.
- Piepenbrock, R. (2000). CGN Lexicon 7.0. Corpus Gesproken Nederlands.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10: 355-371.
- Read, J. (1995). Refining the word associates format as a measure of depth of vocabulary knowledge. *New Zealand Studies in Applied Linguistics*, 1, 1-17.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment*. Mahwah: NJ: Erlbaum, pp. 41-60.
- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, 7: 105-125.
- Richards, J.C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17: 219-240.
- Rönningberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *International Journal of Audiology*, 42, S68-S76.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8: 382-439.
- Rost, M. (1990). *Listening in language learning*. Harlow (Essex): Longman.
- Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78: 199-221.
- Sánchez, A., & Manchón, R.M. (2007). Research on second language vocabulary acquisition and learning: An introduction. *International Journal of English Studies*, 7: 7-16.
- Sanders, M. (1990). Met andere woorden. Een onderzoek naar de omvang van de receptieve woordenschat van anderstalige brugklasleerlingen. *Levende Talen*, 455: 424-427.
- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48: 281-317.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18: 55-81.
- Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioural science*. New York: McGraw Hill.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. United Kingdom: Cambridge University Press.

- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24: 3-25.
- Vandergrift, L. (2006). Second language listening: Listening ability or language proficiency? *The Modern Language Journal*, 90: 6-18.
- Verhoeven, L. & Vermeer, A. (1992). Woordenschat van leerlingen in het Basis- en MLK-onderwijs. *Pedagogische Studiën*, 69, 3: 218-234.
- Verhoeven, L. & Vermeer, A. (2001). *Taaltoets Alle Kinderen*. Arnhem: Citogroep.
- Viellevoye Grooters, W. (2005). Pilotonderzoek. Bilingualism and phonological representation. *Bachelor Thesis*. Universiteit Utrecht.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual knowledge model. *Studies in Second Language Acquisition*, 23: 41-69.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency. *Studies in Second Language Acquisition*, 27: 567-595.
- Zekveld, A.A., Heslenfeld, D.J., Festen, J.M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, 32: 1826-1836.

## Appendices

Appendix A

Appendix B

Appendix C

## Appendix A

### Vocabulary Test

<b>Item</b>	<b>Word</b>	<b>Level of difficulty</b>	<b>Domain</b>	<b>English translation</b>
1	paard	6	dieren 'animals'	horse
2	dorst	4	eten 'food'	thirst
3	tang/ (nijptang)	3	neutraal 'neutral'	nippers
4	vleermuis	4	dieren 'animals'	bat
5	kersen	3	eten 'food'	cherry
6	winter	4	school 'school'	winter
7	maïs	2	eten 'food'	corn
8	bad (kuip)	6	neutraal 'neutral'	bath
9	lezen	6	school 'school'	to read
10	aardbei	6	eten 'food'	strawberry
11	nijlpaard	4	dieren 'animals'	hippopotamus
12	kwast	5	school 'school'	paintbrush
13	speelgoed	4	school 'school'	toys
14	ui	4	eten 'food'	onion
15	fee/ (elfje)	5	neutraal 'neutral'	fairy
16	mier	5	dieren 'animals'	ant
17	toetsenbord	2	school 'school'	keyboard
18	helm	3	neutraal 'neutral'	helmet
19	uil	6	dieren 'animals'	owl
20	pompoen	3	eten 'food'	pumpkin
21	struisvogel	2	dieren 'animals'	ostrich
22	liniaal/ (meetlat)	2	school 'school'	ruler
23	bijenkorf	2	neutraal 'neutral'	beehive
24	worst	5	eten 'food'	sausage
25	knippen/ (knutselen)	6	school 'school'	to cut
26	stoom	2	neutraal 'neutral'	steam
27	popcorn	5	eten 'food'	popcorn
28	kraan	6	neutraal 'neutral'	tap
29	kreeft	2	dieren 'animals'	lobster
30	leeg	5	neutraal 'neutral'	empty
31	puntenslijper	3	school 'school'	sharpener
32	appel	6	eten 'food'	apple
33	driehoek	3	school 'school'	triangle
34	horens	4	neutraal 'neutral'	horns
35	octopus/ (inktvis)	3	dieren 'animals'	octopus
36	aansteken	4	neutraal 'neutral'	to lighten
37	ananas	2	eten 'food'	pineapple
38	walvis	3	dieren 'animals'	whale
39	lijm	5	school 'school'	glue
40	leeuw	5	dieren 'animals'	lion

## Appendix B

### Total scores on Vocabulary Test and Dialang Levels

<b>Participant Number</b>	<b>Language Background</b>	<b>Total Score Vocabulary Test</b>	<b>Dialang Level</b>
1	LAS	6	A2
2	LAS	12	A2
3	LAS	39	C2
4	LAS	12	A2
5	LAS	6	A1
6	LAS	19	C2
8	LAS	10	C1
9	LAS	26	C2
12	LAS	28	B1
16	LAS	36	C2
30	PS	15	B1
31	PS	2	A1
32	PS	37	C2
33	PS	8	A1
34	PS	14	C1
35	PS	15	C1
36	PS	22	C2
37	PS	18	C2
38	PS	0	A1
39	PS	0	A1
45	BP	13	B1
46	BP	24	C2
48	BP	19	C1
49	BP	10	A1
51	BP	5	A1
52	BP	3	A1
53	BP	29	C1
54	BP	9	A1
55	BP	35	C2
56	BP	27	C2
59	DUT	40	-
60	DUT	40	-
61	DUT	40	-
62	DUT	40	-
63	DUT	39	-
64	DUT	40	-
65	DUT	40	-
66	DUT	40	-
67	DUT	40	-
68	DUT	40	-

## Appendix C

### Index of Discrimination

<b>Item</b>	<b>Word</b>	<b>Index of Discrimination</b>
1	paard	0.56
2	dorst	0.66
3	tang / (nijptang)	0.20
4	vleermuis	0.35
5	kersen	0.63
6	winter	0.16
7	maïs	0.42
8	bad (kuip)	0.50
9	lezen	0.55
10	aardbei	0.47
11	nijlpaard	0.23
12	kwast	0.22
13	speelgoed	0.33
14	ui	0.43
15	fee / (elfje)	0.15
16	mier	0.42
17	toetsenbord	0.35
18	helm	0.37
19	uil	0.32
20	pompoen	0.58
21	struisvogel	0.10
22	liniaal / (meetlat)	0.35
23	bijenkorf	0.22
24	worst	0.46
25	knippen / (knutselen)	0.63
26	stoom	0.37
27	popcorn	0.40
28	kraan	0.56
29	kreeft	0.08
30	leeg	0.56
31	puntenslijper	0.17
32	appel	0.31
33	driehoek	0.60
34	horens	0.43
35	octopus / (inktvis)	0.22
36	aansteken	0.38
37	ananas	0.57
38	walvis	0.38
39	lijm	0.30
40	leeuw	0.44