JYVÄSKYLÄ STUDIES IN HUMANITIES
184

Helena Miettinen

Phonological Working Memory and L2 Knowledge

Finnish Children Learning English

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN HUMANITIES 184

Helena Miettinen

Phonological Working Memory and L2 Knowledge

Finnish Children Learning English

Esitetään Jyväskylän yliopiston humanistisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S 212
elokuun 31. päivänä 2012 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Humanities of the University of Jyväskylä,
in building Seminarium, auditorium S 212, on August 31, 2012 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Phonological Working Memory and L2 Knowledge

Finnish Children Learning English

JYVÄSKYLÄ STUDIES IN HUMANITIES 184

Helena Miettinen

Phonological Working Memory and L2 Knowledge

Finnish Children Learning English

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Editors
Paula Kalaja
Department of Languages, University of Jyväskylä
Pekka Olsbo
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-4825-2
ISBN 978-951-39-4825-2 (PDF)
ISSN 1459-4331 (PDF)

ISBN 978-951-39-4824-5 (nid.)
ISSN 1459-4323

Jyväskylä University Printing House, Jyväskylä 2012

ABSTRACT

Miettinen, Helena
Phonological Working Memory and L2 Knowledge: Finnish Children Learning English
Jyväskylä: University of Jyväskylä, 2012, 185 p.
(Jyväskylä Studies in Humanities
ISSN 1459-4331; 184)
ISBN 978-951-39-4824-5 (nid.)
ISBN 978-951-39-4825-2 (PDF)

Phonological working memory functions as a gateway for storing linguistic knowledge into long-term memory (Baddeley, Gathercole and Papagno 1998). It is therefore considered not only an essential part of language learning but also a potential source of individual differences in language learning. This study investigates the connection between language knowledge and phonological working memory as understood in the working memory model by Baddeley (1986). The primary aim is to study what kind of a relationship exists between phonological working memory and second/foreign language (L2) knowledge. The study also examines a number of aspects generally associated with phonological working memory, including modality and language specificity.

The participants of the study were 15 Finnish children who had studied English as their first L2 since the third grade. The children took part in a longitudinal research project, of which the present study focuses on two years, grades five and six. While some of the data used in this study are drawn from the larger project, the main body of the data was collected with measures specifically designed for this study. These measures included an English and a Finnish nonword repetition test and a number of new L2 measures designed to differ from and thus add to measures used in previous studies in this area. The data were analyzed statistically using correlation and cluster analysis.

The findings suggest that the connection between English (L2) knowledge and phonological working memory as assessed with an English nonword repetition test is fairly strong. Moreover, neither the modality (written or spoken production) nor the assessment method (structure or meaning) of the L2 measures appeared to have much effect on the relationship. Furthermore, phonological working memory did not appear to be strictly language-specific as both nonword repetition tests (based on L1 and L2) were connected to L2 language measures.

Keywords: phonological working memory, nonword repetition, L2, individual differences

Author's address

Helena Miettinen
Department of Languages / English
P.O. Box 35
40014 University of Jyväskylä, Finland
helena.miettinen@jyu.fi

Supervisors

Paula Kalaja
Professor
Department of Languages / English
University of Jyväskylä, Finland

Riikka Alanen
Professor
Department of Teacher Education
University of Jyväskylä, Finland

Reviewers

Judit Kormos, PhD
Senior Lecturer
Department of Linguistics and English Language
Lancaster University, UK

Elisabet Service
Assistant Professor
Department of Linguistics and Languages
McMaster University, Ontario, Canada

Opponent

Elisabet Service
Assistant Professor

FOREWORD

This thesis would never have seen the light of day without the guidance, support and persistence of my supervisors Paula Kalaja and Riikka Alanen. When it comes to the thesis process, I have not really had a close "peer group" to lean on since the very first year, so that has made the two of you all the more important. Thank you for always finding the time to read and discuss my work and for egging me on.

The roots of this work are in a project headed by Riikka, but it was actually Hannele Dufva who interviewed me for an internship in the project. I got the job and am fairly certain I would not be here had things turned out differently. Thank you for seeing something in me, Hannele! On the other end, in its final steps, this thesis has benefited greatly from the detailed comments and helpful suggestions of the external reviewers, Judit Kormos and Elisabet Service. I appreciate how willingly you have shared your insights.

The thesis project officially started when I came to work at the Department of Languages. Along with a grant from the Ellen and Artturi Nyyssönen Foundation, the department has kept a roof over my head over the years spent working on my thesis. I have had the honor of working under the reign of three different heads of department: Tuula Hirvonen, Bertold Fuchs and Anne Pitkänen-Huhta. Thank you for always having a job for me, and while being generous with research time, for also giving me the opportunity to work as a full-fledged member of staff involved in teaching, syllabus work, various working groups, etc. I cannot imagine what it would be like to only focus on one's research. I say this with absolutely no sarcasm, quite the contrary, I feel I have had the opportunity to learn so much more than just the ins and outs of research, and for this, I am forever grateful.

The Department of Languages has been a wonderful place to work in many ways but its most obvious perk, for me, has been the wonderful colleagues. So, a warm thanks to you all and especially the English section. Many of you have also had a hand in my research, so a special thank you to Eleanor Underwood, Mari Aro, Alicia Copp Jinkerson, Leila Kääntä and Laura McCambridge. Some of you are also dear friends and the rest I could easily imagine being such.

In addition to the colleagues mentioned above, a number of other people have contributed to the thesis in one way or another and probably do not even remember doing so. Let me remind you. Tuija Hirvelä, Ari Maijanen, Mirja Tarnanen, thank you for making this all possible. In addition, there is one person who probably remembers but wishes he could forget: Michael Freeman, thank you for your careful, yet speedy, proof-reading. I may now be understood.

There is also a very special group of people that I want to thank, who have, at one time or another, been more or less forced to deal with my research. They have always had a minute to listen but also to share their own experiences with similar hurdles. What they have *not* had is an escape, for they have been my

officemates: Marianne Toriseva, Saara Jäntti, Piia Varis, Sanna Lehtonen and Heidi Koskela. Strangely, I have also had many other office mates during the years. I like to think it is not just because I wear people out.

There are yet more colleagues who have allowed me to discuss all matters academic, and in general they have made daily life at the office (and outside) a picnic: Katja Mäntylä, Riikka Ullakonoja and the "dream team", *Unelmanaiset*, Johanna Saario, Niina Lilja and Heidi Vaarala. Thank you for your encouragement and advice at different points of the process, but especially for all the movies, dinners, lunches, drinks and coffee breaks that have so enriched my life as well as my nutrient intake.

Kiitos Päiville, Annelle, Heidille, Jennille ja Tuulalle ystävyydestä, kummilapsista ja Luiskasta. Olemme varmaankin puhuneet kaikesta maan ja taivaan välillä paitsi väitöskirjastani ja olen rakastanut joka hetkeä! Heikkiä ja Merjaa haluan kiittää etenkin yhdestä elämäni kirkkaimmasta valopilkusta, Tarmosta, jonka kanssa hengaillessa ei voi mikään, edes väitöskirja, painaa mieltä. Lämmin kiitos myös rakkaalle puolisolleni Mikolle, jonka kanssa on ihanaa jakaa niin arki kuin juhla, ahvenet ja kuhat. Tämä on nyt se juhla.

Vanhempiani kiitän aivan kaikesta. Omistan tämän työn teille.

Jyväskylässä Helenan päivänä 2012
Helena Miettinen

FIGURE

TABLE

CONTENTS

1 INTRODUCTION

Intuitively, it is obvious that memory and language are connected, and research unquestionably indicates the same. Issues to do with both memory and language separately and together have been widely studied, and the approaches have been manifold. Since both issues are complex to say the least and little is known for a fact, theories and models about the structure and functioning of both language and memory are abundant. The present study draws on research on the cognitive aspects of language learning. While it is understood that language is very much a social phenomenon, the view adopted here is that language learning is a result of cognitive processing and is subject to the rules of information processing. It is the nature of the information that sets language learning apart from all other learning.

Although the basic mechanisms of learning are supposedly the same for all, it is still the case that, in any group of learners of a second or foreign language (henceforth L2), some learn more easily than others. The individual differences that cause this variation are of interest to L2 researchers and teachers alike, as knowledge of them could potentially lead to ways to intervene and help poor learners reach levels or speed of learning closer to those of good language learners. Working memory is one of the elements recognized as a source of individual differences. It is a part of the cognitive architecture that plays a role in language learning and is considered a crucial source of individual differences at least in language aptitude and in learning strategies.

The working memory model that the present study draws on is a multi-componential one. Here, the focus is on one particular component, *phonological working memory*, which is thought to be essential to language learning due to its role as a gateway and possible bottleneck through which verbal input gains access to long-term memory. In order to operationalize and assess phonological working memory, a number of tests have been developed in the field of cognitive psychology. One of such tests is the nonword repetition test.

Nonword repetition tests have been used to examine various operations that take place in phonological working memory; however, they have most commonly been employed in studying the relationship between phonological

working memory and different aspects of language learning. This relationship has been investigated in both children and adults and subjects with and without memory or language deficiencies. The focus of studies has varied from short-term language learning to the language specificity of phonological working memory, and has included a vast range of languages as targets of learning and study. While the number of studies focusing on the relationship between phonological working memory and L2 are growing, they remain a minority in a field which originated and continues to be dominated by research on first or native language (henceforth L1) acquisition. Nevertheless, the existing research strongly points to the importance of the role of phonological working memory in L2 learning and consequent knowledge. If this is the case, gaining specific information about the mechanisms behind the connection is essential to eventually being able to provide support to learners regardless of the differences in their phonological working memory capacity.

1.1 Context of the Present Study

The present study is part of a longitudinal research project titled Situated metalinguistic awareness and foreign language learning, headed by Professor Riikka Alanen (the Centre for Applied Language Studies, University of Jyväskylä). The project aimed at defining young language learners' awareness and beliefs about the Finnish and English language and language learning. The participants were a group of Finnish school children who started school at the age of six or seven, as is the norm in Finland. All the participants spoke Finnish as their L1 and began studying English as their first L2 in the third grade. The data collection was carried out during grades one through six with the number of participants diminishing from 20 at the beginning of the project to 15 at the end of the data collection. The project has produced several publications and theses, for example, Alanen 2003 and 2006; Aro 2001, 2003, 2006; Dufva 2003; Mäntylä 2004 and 2008; Dufva, Alanen and Mäntylä 2001; Dufva, Alanen and Aro 2003.

In this larger research project, Vygotskian sociocultural and Bakhtinian dialogic frameworks were applied in studying the participants' metalinguistic development. An attempt was made in the project to combine social and cognitive aspects of language learning with a more general sociocultural and dialogic framework. Both qualitative and quantitative methods were used in analyzing the data, which is quite varied, consisting of, for example, interviews, classroom observation, various language tasks and quantitative tasks such as the nonword repetition test. The aim of the project was not to obtain generalizable results but to studying the development of the metalinguistic knowledge of the participant children as individuals. The emphasis was on the situatedness of metalinguistic awareness, the effect it has on L2 learning and the way the participants see their own L1 and the L2 they are learning.

The present study draws partly on data previously gathered for the project, but for the most part the data have been collected specifically for this study. The purpose was to provide both quantitative and qualitative findings on L2 knowledge, and to supplement the mainly qualitative data and data analyses on L2 learner beliefs provided by other studies in the project. A Finnish nonword repetition test was carried out in the longitudinal research project when the participants were in the second grade. This was done partly to assess the participants' phonological working memory and partly to find out whether it was possible to create a Finnish nonword repetition test that would yield interpretable results in the first place. This was of interest because previous studies with such nonwords had been problematic, an assumption being that repeating Finnish nonwords is possibly too easy a task for school-age children speaking Finnish as their L1. The present study is in a sense an expansion and a continuation of that first Finnish nonword repetition test designed for the longitudinal research project. The first test appeared viable, showing differences between participants, and thus it was deemed appropriate to continue its use as a functional tool in acquiring further psychometric data on the participants.

The longitudinal nature of the project made it possible to carry out tests with the same participants in grades five and six and to bring in English nonword repetition tests in addition to continuing to use the Finnish tests. However, since the participants were considerably older by the second Finnish nonword repetition test and many of the previous studies using Finnish nonword repetition tests had failed to obtain proper data, constructing a new functional Finnish nonword repetition test was one of the main concerns. Age is an important factor when it comes to children's phonological working memory, and hence it was reasonable to assume that a new and more complex version was necessary for the measurements in grades five and six. Functional tests were needed for these later years in order to investigate possible changes between school years in the participants' performance on a phonological working memory task.

1.2 Aims, Data and Methodology

In many of the studies using nonword repetition tests, especially the earliest ones, the aim has been to investigate the relationship between repetition scores and language knowledge or learning. The term *knowledge* is used to refer to performance on L2 tasks at a given point in time while *learning* refers to a change in knowledge between measurements, for example before and after an intensive L2 course or training in an experimental situation.

To date, there is abundant evidence that a strong connection exists between phonological working memory and L1 (e.g. Gathercole and Baddeley 1989, 1993b), and also between phonological working memory and L2 (e.g. Service 1989, Baddeley 1993), although such studies are fewer in number. From

the point of view of working memory research, L2 studies are integral to pinning down the role of phonological working memory in language acquisition. Equally importantly, a better understanding of the relationship between the memory construct and L2 is valuable for deciphering the processes involved in L2 learning. Knowledge of the possible differences between the connection of phonological working memory to L1 and to L2 can throw further light on the differences in the mechanisms or strategies that are available to the L2 but not L1 learner and vice versa.

Furthermore, conducting nonword repetition tests in different languages and comparing the results between different language tasks can illuminate the nature of phonological working memory, in particular whether it is language specific or general in nature. It remains an open question whether the only way to acquire knowledge of phonological working memory is to use nonwords in languages the participants do not know and whether thinking so means that phonological working memory is assumed to equal language knowledge, so that using an English nonword repetition test, for example, is considered just one way of assessing English knowledge. There appear to be some contradictions in the existing research regarding this issue: nonwords are used precisely because they do not depend on previous language knowledge, yet nonwords of an unknown language are preferred as this is thought to enable the researcher to get to the root of phonological working memory capacity without the participants' knowledge of their L1 intervening. In the present study, nonwords in the participants' L1 and L2 were used in assessing phonological working memory, and L1 and L2 language measures were also used. Comparisons between these tasks might also shed further light on the language specificity issue of phonological working memory. For example, if nonword repetition seems to predict language knowledge in both L1 and L2 regardless of the language that the nonwords are mimicking, a general phonological working memory could be assumed to be tapped by both tests. Possible differences in the connections between the nonword repetition tests and specific language measures could also be hypothesized to indicate that some of the tasks are connected to a language-specific and others to a more general phonological working memory.

Since the research participants had been learning English (L2) since the third grade, it was feasible to assess their English nonword repetition and English knowledge in grades five and six. Previous studies have focused heavily on L2 vocabulary, although some very comprehensive assessments of L2 knowledge exist as well. The present study attempted to further diversify the field in this regard. The fifth-grade English assessment was conducted for Miettinen (2003) and comprised novel tasks, those done as part of the data collection for the longitudinal research project, and also tasks originally administered as part of regular school work. The scores on these tasks were used in the present study to answer some of the research questions. In the sixth grade a more experimental approach was adopted. In order to gain a more global assessment of the participants' English knowledge, the participants took

part in three types of tasks all of which required them to produce English based on what they had heard. The tasks were thus similar to the nonword repetition test, but also required the use of different modalities, speaking and writing.

Another aim of the present study was to investigate the language specificity of phonological working memory, the extent to which the language of the nonword repetition test appears to have an effect on the result, and the association found between phonological working memory and language knowledge. For this purpose data already available in the research project on the participants' knowledge about L1 Finnish were used.

To summarize, the present study seeks answers to two general research questions: Is there a connection between children's knowledge of L2 and their success in repeating L2 nonwords? Is phonological working memory language specific?

The approach taken here was mainly quantitative, since the tasks and tests used yielded quantifiable data in the form of number scores. First, the connections between the English and Finnish nonword repetition tests and English and Finnish language knowledge were studied using statistical analyses, by calculating correlations between the scores. To get a better understanding of the nuances of the connections and differences between the individuals, cluster analysis was conducted and the participants' individual profiles were also investigated.

It is important to note that in the present study, the angle was that of L2 knowledge. The language tasks used were carefully constructed to add to the existing research not only through what is assessed but also how. The tasks take into account different levels of language as well as different modalities and scoring perspectives. According to the literature, working memory, and especially phonological working memory, appears to have a close connection with language development yet, given that the amount of research on L2 in this respect is far from overwhelming, it was an interesting topic to research further. Also in the language aptitude literature, working memory is seen to play an important role, it even being claimed that working memory may be the whole of language aptitude. Simply put, this study focused on one aspect of working memory, i.e. phonological working memory as envisioned by Baddeley and Hitch (1974), which was measured by the nonword repetition test and its connection to L2 knowledge studied in a group of children to see what relationships with different L2 measures emerged as compared to previous studies on the subject.

1.3 Outline of the Study

This dissertation is structured so that in Chapter 2 a brief introduction to the cognitive perspective and the study of individual differences in language learning is provided in an attempt to shed light on the role and importance of

memory in language learning and use. The multitude of working memory models and their differences are touched on in Chapter 3, leading to a more detailed description of the particular model used in the present study and the nonword repetition test applied to operationalize the key concept here, phonological working memory. Chapter 4 introduces previous research on the connections between phonological working memory and both L1 and L2, after which, in Chapter 5, the design of the present study, including the research questions, the participants, and data collection and processing, is made explicit. The results of the statistical analyses are reported and discussed initially in Chapter 6 and further discussed and compared with those of previous studies in Chapter 7 together with concluding remarks and suggestions for further research.

2 COGNITIVE PERSPECTIVE AND INDIVIDUAL DIFFERENCES IN SLA

The roots of the present study are in research on individual differences, especially the cognitive processes involving working memory. In the present chapter, the role of working memory is first discussed in the context of information processing and the cognitive view on second language acquisition (SLA) more generally. After this, the field of individual differences in SLA is described and discussed again, focusing on what role working memory is seen to play in individual differences among language learners.

2.1 Working Memory in Information Processing and Language Learning

There are a number of cognitive SLA theories through which L2, as a form of cognition, ties in with the important role of working memory. In fact, memory and attention in L2 learning are what SLA research on cognition is most concerned with (see e.g. Robinson 2001). Memory – and especially working memory – is in a key role in the information processing paradigm, according to which the human mind is constantly involved in the processing of mental representations (see e.g. Altarriba and Basnight-Brown 2009). There are fundamental assumptions in information processing that apply well to SLA. These are described in, for example, McLaughlin (1987). First, mental processing is considered dual: it is automatic (unconscious) on the one hand and controlled (conscious) on the other. In addition, cognitive resources, for example attention and memory, are limited. If a process is automatic, several of them can occur simultaneously, since they do not take up many cognitive resources. Controlled processes, by contrast, start with the individual and his or her motivation, as these are believed to depend on attentional control.

The application of controlled processes takes place when an automatic approach has not yet been learnt, or if there is something amiss in automatic

processing. Controlled processing is a strain on the limited capacity of the individual's cognitive resources, and it is, in fact, serial, meaning that conscious processing is applied to one task at a time, and thus a bottle-neck can be created in a situation where many issues require controlled processing at the same time (see e.g. Segalowitz and Hulstijn 2005; McLaughlin 1987).

When we learn, information is moved to long-term memory (McLaughlin 1987: 135). In other words, long-term memory embodies representation: our knowledge, what we know, is in our long-term memory, which is unlimited. Working memory, on the other hand, is limited and represents access, processing. It is capable of holding information and most importantly integrating existing knowledge (information in long-term memory) with new information (see e.g. Gathercole and Baddeley 1993a, Robinson 1995). In sum, memory is the construct that the very idea of information processing is based on: all information goes through and involves some form of memory, and long-term memory and working memory are the two types of memory crucial in all cognitive procedures, including the processing and learning of L2.

A more detailed description of the connection between working memory and language learning can be found, for example, in Gernsbacher (1994: 1075, 1112). It appears that a lot of language learning is in fact a way of freeing up processing resources in working memory. First of all, language production and comprehension both rely on working memory since they involve processing something linear – sequences of symbols – which entails the simultaneous processing and storing of "computations" in working memory. Furthermore, it is the current understanding that representations of thought are not sequentially organized, and consequently working memory is also needed in transforming such representations into linear language. This turning of thoughts into language and comprehending and producing language makes great demands on working memory. Because working memory has a limited capacity to perform this processing and storing, comprehension performance (speed and/or accuracy), for example, declines with increased intrinsic complexity, extrinsic memory load, or temporal constraints. The end result is that the possible individual differences in how this crucial working memory functions most likely also lead to individual differences in language learning.

Working memory is an important gateway for information to reach long-term memory. According to Ellis and Sinclair (1996: 246–247), long-term sequence information for language, i.e. knowing stretches of language, is possible through accurate short-term representation and rehearsal. If language learning is largely just acquiring "memorized sequences of language" then the role of working memory is absolutely crucial. N. Ellis (2001: 44, 49) describes the process as follows. As a novel word, be it in L1 or L2, is repeated over and over in phonological working memory, more and more regularities and chunks can be abstracted. This increases the chance of bringing the correct forms into working memory to be accurately processed during language use (pronouncing or labeling representations). Repeating L2 forms aids long-term retention, and long-term knowledge of vocabulary items has an effect on phonological short-

term memory. Gaining knowledge of the phonotactics of the L2 happens automatically upon hearing and uttering L2 words, and this knowledge then supports the short-term repetition of unfamiliar words in the L2. Moreover, a similar process takes place in learning the structures of the L2, as knowledge of structural regularities is abstracted from utterances heard if they appear to serve the same function and exhibit structural similarities.

On the level of the individual, learners differ in how well they are able to repeat phonological sequences; these differences in turn partly explain individual differences in language learning aptitude (Ellis 2001: 48). However, the relationship between working memory and long-term memory is thought to be reciprocal. The information we have about language in our long-term memory, for example sequences of sounds, can add to our working memory capacity through chunking (Ellis and Sinclair 1996: 247). Our long-term knowledge filters, patterns and chunks what we perceive, thereby, to some extent, easing the processing load on working memory. In sum, then, the interaction between the long-term and short-term systems is bi-directional. It is not just that working memory is a gateway to long-term memory, but information from long-term memory can be held in working memory when it is relevant for the information being processed.

On a more general note, there is often too much input in any communicative situation, and thus the working memory system plays a role in screening out what is not relevant for comprehension at any given time (Skehan 1998: 44–45). Perception as input into working memory is automatically filtered and patterned by our existing long-term memory schema (Ellis 2001: 35). This filtering idea is a more interactive view of short-term memory than the mechanism proposed by Baddeley (discussed below) in that short-term memory or working memory reflects activated and attended subsets of long-term memory. Indeed, views differ on how working memory and input interact and how some input is processed by long-term memory without the intervention of working memory (Skehan 1998: 45). There does not seem to be any doubt about the overall relevance of memory as a component of language aptitude, however. As Skehan (1998: 218) puts it, "memory is important all along, goes hand in hand with success in language learning until an advanced level is reached, and then becomes even more important".

In the theoretical discussions here several references have been made to memory, long-term memory and working memory without much attempt to explain the terms. As there is no single unanimous view on memory or its subsystems and as grounding for the working memory model applied in the present study, a short introduction to the most prominent theories on memory and ways of dividing it into subsystems is provided in Chapter 3.

Finally, working memory has been studied quite extensively in the SLA framework during the past 15 years or so and its role is especially clear in the area of individual differences (see e.g. Skehan 2002). If memory is important in information processing, then it can be assumed that better working memory capacity leads to more effective learning of an L2. It is to the study of individual

differences in language learning and the details of the role of working memory in them that we turn next.

2.2 Individual Differences in L2 Learning: The Role of Working Memory

In SLA research, interest has first and foremost been in the similarities between language learners and what is common in language learning processes (Skehan 1989:1). This is, of course, natural when the focus has been on understanding a phenomenon in broader terms. It also makes sense from a pedagogical perspective. At least for the purposes of formal L2 teaching, the general principles and, for example, the order of acquisition of the language are essential cornerstones. However, language is always learnt by individuals, and it is easy to imagine that there are differences in the individuals that in many respects may have a bearing on their language learning. The fact that in similar circumstances some learners succeed in acquiring an L2 while others do not raises the question of what the differences among individuals are that lead to this difference.

Before going into more detail, it should be noted that the terminology around individual differences has varied somewhat over the years. Rubin (1975) talked about good language learners, as did Griffiths (2008), while the terms *learner characteristics* and *individual differences* have also been used to refer to the same research interest (e.g. Skehan 1989). The present study opts for the term *individual differences* due to its more general nature and extensive use in SLA studies.

As an area of SLA research, the study of individual differences in L2 learning employs theories and methods from cognitive psychology in particular (Dörnyei and Skehan 2003). What is meant by individual differences is not a simple issue. At least, the individual differences that are generally of interest should be somewhat stable characteristics of individuals (Dörnyei 2005: 4). Also, they should be present in everybody, only differing in their level or extent between individuals.

Individual differences have in turn been used to study the more general nature of language learning in the sense that there have been attempts to categorize individuals based on their individual difference profiles. Once it is known on what parameters individuals differ, they can be categorized and grouped according to their profiles with similar learners. This is often accompanied with what seems to be a strong (pedagogical) interest in the field, i.e. to figure out how different types of learners (different in terms of some individual differences) respond to different instructional settings (see e.g. Robinson 2002a; Wesche 1981).

While individual differences in general have been the focus of study in other disciplines for some time, the research on individual differences in L2

learning dates back to about the mid-1970s when researchers such as Rubin (1975) started looking at good language learners and the features that set them apart from less successful ones. The aim was to see what the good language learner does differently, for example, in terms of learning strategies, and use that information in teaching less proficient ones. A narrower approach to individual differences in L2 learning had, of course, already been taken earlier by Carroll and Sapon as well as by Pimsleur, who focused on language aptitude and developed the assessment batteries MLAT and PLAB, discussed in more detail below in section 2.2.3.1.

In 1989 Skehan wrote a seminal book on individual differences in L2 learning. He focused on the major areas of difference between language learners: aptitude, motivation and language learning strategies. Since then, there have been many different listings, but to date there appears to be no consensus regarding what to include under the concept of individual differences or even what any single component actually involves. Dörnyei (2005: 7) presents many and often very long lists of components considered to factor in individual differences in psychology generally, but, as he focuses on the learner, he ends up settling for personality, ability/aptitude and motivation, with learning styles and language learning strategies added to the list as the individual difference research tradition considers them important factors. He also includes anxiety, self-esteem, creativity, willingness to communicate and learner beliefs as individual difference variables worthy of note. In their article on individual differences in L2 learning, Dörnyei and Skehan (2003) limit their list to language learning styles, cognitive and learning strategies, aptitude and motivation, the last two of which they view as the most fruitful areas for future research on individual differences.

The main objective of the present study was to see if and how differences in learners' phonological working memory are reflected in their success in a number of L2 tasks. It is therefore of interest to pinpoint, in particular, the areas of individual differences where memory is seen to play a part. This topic is addressed below.

2.2.1 Cognitive and Learning Styles

People, language learners included, differ in how they prefer to use their cognitive abilities, not just in how good their cognitive abilities are. To have a certain cognitive style is to be prone to process information in a certain way (Dörnyei and Skehan 2003: 602). Dörnyei (2005: 124) expresses it thus: "Cognitive styles are usually defined as an individual's preferred and habitual modes of perceiving, remembering, organizing, processing, and representing information."

Dörnyei (2005: 123–124) talks at length about the concept of learning style and how complex and elusive it is. It is not entirely clear how innate learning styles are or whether they actually exist independently of learning strategies. According to Dörnyei (2005: 124), the problem lies mostly in the fact that it is not very well understood what learning is, what ultimately is involved;

however, the idea of styles becomes more manageable when cognitive style is separated from learning style. Cognitive style can be seen as the nucleus of learning style, as its somehow biologically predetermined and stable component but it is not to be confused with cognitive abilities, as it a question of, for example, a preference for visual or auditory information in a learning situation instead of intelligence (Dörnyei 2005: 124–125). In fact, one reason for considering cognitive style as a potential source of individual differences was the realization that not all variation could be explained through differences in ability; instead, some of it was thought to be attributable to style. A person's abilities in something can be good or poor or something in between but with style it is a question of dichotomies, preferences for A or B without either being necessarily superior in relation to the other.

Dörnyei (2005: 124–125) cites Brown (2000) when he discusses the relationship between learning and cognitive styles. When cognitive style is considered in the context of education together with "affective, psychological and behavioral factors", the end result is what is usually referred to as learning style. Sensory preference (visual, auditory, kinesthetic and tactile) is one approach to the study of learning styles (Dörnyei 2005: 139), although memory does not seem explicitly to figure in styles. However, it is obvious that memory issues may be behind some preferred ways to process information.

As Dörnyei (2005: 126) puts it, there are some fairly serious problems with the concept of cognitive style. It appears that each scholar has his own list of styles and measures to assess them, but very little is commonly agreed. This is also the reason why they are not discussed in much detail here. However, one of the first and probably the most widely recognized notions under cognitive style is the division into field-dependent and field-independent learners (Dörnyei 2005: 136; Dörnyei and Skehan 2003: 603-604). This is classically about a contrast between analytic and holistic processing, and the concept has not passed without (fairly strong) criticism (see e.g. Dörnyei 2005: 138). Skehan (1998) has suggested that more interesting than distinguishing between analytic and holistic is a division into analytic and memory orientations which need not detract from each other. It is possible to be strong in one, both or neither. It should also be noted that this latter division can be seen on both the level of ability and of style, and thus all combinations are possible; for example, a learner with abilities leaning toward the analytic might nevertheless prefer a learning style that is more geared towards employing memory.

Indeed, when discussing learning styles, memory comes most clearly into play in Skehan's dual-coding approach to language learning and performance, even though there has been some discussion about whether this is a question of ability rather than style (Dörnyei 2005: 152). In this approach, there are considered to be analysis-oriented and memory-oriented learners (also referred to as grammarians versus chunkers) and, if it is considered a preference, this division is to do with learning style. Some prefer to focus on systematic rules, others on communication. However, the root of the difference may also be in the demands of the task at hand or the underlying aptitudinal abilities, rather

than style, i.e. learners may go by what is possible or easy for them instead of actually preferring one way over the other and making choices accordingly (Dörnyei 2005: 152; Skehan 1998: 251–252). Among the many learning style conceptualizations introduced by Dörnyei, Skehan's is also a more language-oriented one. The others are suitable for basically any subject matter, and in them the special nature of language learning is not really considered (Dörnyei 2005: 154).

2.2.2 Learning Strategies and Self-regulation

The idea of the good language learner – what they do that the others do not – was what gave rise to research into language learning strategies. There are a number of definitions of strategies and these and their problems are discussed at length, for example, in Dörnyei (2005). One of the more recent definitions is provided by Griffiths (2008: 85–87). The basic idea is that in addition to the effects of motivation and aptitude learners are different in the kinds of learning techniques they apply. Learning techniques or strategies have been a research interest since the 1970s (Dörnyei and Skehan 2003: 607–608).

Rubin (1981, as cited in Chamot and O'Malley 1994: 373–374) classified strategies into those which affect learning directly and those with indirect effects. As an example of the kinds of issues that may fall under learning strategies, Dörnyei and Skehan (2003: 609) present a list of strategy categories that they have drawn from previous listings. In short, their taxonomy includes cognitive, metacognitive, social and affective strategies, and is very similar to that presented in Chamot and O'Malley (1994: 374–375). Although memory is not explicitly mentioned by these scholars, it is obviously required in many strategies. For example, imagery and auditory representation, which go under cognitive strategies (e.g. in Chamot and O'Malley 1994), seem quite directly related to the use of the visuo-spatial sketchpad and the phonological loop of the Baddeley and Hitch (1974) working memory model (section 3.1.1).

A condensed, combined taxonomy such as the one by Dörnyei and Skehan mentioned above is a good starting point, as the taxonomies available are varied and plentiful. The two that Dörnyei (2005: 168–169) mentions as well-known are O'Malley and Chamot (1990) and Oxford (1990). The first-mentioned divide strategies into three main classes: cognitive, metacognitive and social/affective strategies. Cognitive strategies are said to involve mental manipulation of the material to be learned, for example through repetition of words or phrases during a language task (O'Malley and Chamot 1990: 138), which is the most obvious route for working memory to have an effect on learning strategies in this categorization.

Oxford (1990), with six classes of strategies, is more detailed. Three of these – memory, cognitive and compensation strategies – she calls direct strategies because they concern the target language directly. The indirect strategies – metacognitive, affective and social strategies – do not directly involve the language being learnt but are used to support and manage language learning (Oxford 1990: 135). Memory strategies are of special interest in the

context of the present study and these Oxford (1990: 38–43) divides into four sets: creating mental linkages (grouping, associating/elaborating and placing new words into a context), applying images and sounds (using imagery, semantic mapping, using keywords and representing sounds in memory), reviewing well (structured reviewing) and employing action (using physical response or sensation and using mechanical techniques). The use of these strategies may partly depend on learning style preferences, for example, being visually or aurally inclined. Dörnyei (2005: 168) offers criticism of two features of Oxford's taxonomy. First, while language learning and use may be difficult to separate in practice, compensation strategies are to do with language use and should not be included in strategies of learning. Second, memory should not be treated as its own class as it is obviously a subclass of cognitive categories.

A typical method for getting to the bottom of which learners apply which strategies would be a self-report questionnaire (Dörnyei 2005: 178). Typically, such a questionnaire includes statements about strategies, for each of which learners select the most suitable option on a scale from something along the lines of 'not suitable for me / I never use it' to 'very suitable / use always'. The idea is not to think about a particular learning situation but to generalize. The number of items varies greatly between questionnaires. Dörnyei (2005: 179–183) mentions a couple of questionnaires, for example, the Motivated Strategies for Learning Questionnaire, which covers cognitive and metacognitive strategies and resource management strategies with a total of nine subscales and 50 items, and the Language Strategy Use Inventory Index, a practical check-list type questionnaire which focuses mainly on language skills (listening, speaking, reading, writing, vocabulary learning and translation skills). The one that, according to Dörnyei (2005: 181), has been the most widely used is Oxford's Strategy Inventory for Language Learning, which is obviously based on Oxford's strategy system (1990) discussed above. The version for learners of English as an L2 has 50 items which aim at assessing the frequency of particular "strategic behaviors" (Dörnyei 2005: 181).

In comparison to learning styles, memory has a clearer role among language learning strategies – on its own in some taxonomies or as a part of cognitive strategies (Dörnyei and Skehan 2003: 608–609). With respect to metacognitive/cognitive strategies, it is evident that memory-dependent learners use memorization strategies (Skehan 1989: 96). However, these dependencies do not have to be permanent. It has been found that the less successful learner is not somehow inactive but instead does not have the metacognitive knowledge to be able to use an appropriate strategy in a given task (Chamot and O'Malley 1994: 381). Also, less effective learners can use more effective learning strategies if they receive instruction on them (Chamot and O'Malley 1994: 386).

As with learning styles, the concept of strategies is not very clear; educational psychology, in particular, has adopted the more flexible concept of *self-regulation* instead (Dörnyei and Skehan 2003: 610). For this, the literature provides concise definitions. For example, Dörnyei and Skehan (2003: 611) state

that it stands for "the learner's conscious and proactive contribution to the enhancement of her or his own learning process" and "the degree to which individuals are active participants in their own learning". Dörnyei (2005: 191) describes the notion of self-regulation of academic learning as a multi-dimensional construct which includes cognitive, metacognitive, motivational, behavioral, and environmental processes that the learners can apply in order to improve their learning. In addition, it is a process-oriented construct: according to Zimmerman (2001: 1, as cited in Dörnyei 2005: 191) it is not a mental ability or a skill but a process of transforming mental abilities into academic skills.

All in all, the problem with the concept of learning strategies, although intuitively reasonable, has been that there seems to be no sound theoretical basis for the current understandings of what these are (Dörnyei and Skehan 2003: 622). Furthermore, Dörnyei (2005: 195) states that recent research has been geared to more dynamic and process-oriented variables than to learning/cognitive strategies. However, as Griffiths (2008: 85) puts it, opting for the concept of self-regulation does not help much as a self-regulating learner still selects strategies and that is why addressing the issue of strategies on some level cannot be completely disregarded in the end.

After considering the range of learning styles, strategies, aptitude and motivation, Dörnyei and Skehan (2003: 622) are very straightforward about the somewhat questionable nature of both learning styles and strategies, and consider aptitude and motivation the most promising research areas among individual differences. As memory does not figure very clearly in motivation, the final area of individual differences discussed here is aptitude.

2.2.3 Aptitude

Aptitude is a mostly cognitive construct and, along with motivation, it is the most thoroughly researched source of individual differences in L2 learning. Perhaps the simplest way to introduce the concept of aptitude is through the tests designed to assess it. Two classic aptitude measures, the Modern Languages Aptitude Test (MLAT) and the Pimsleur Language Aptitude Battery (PLAB) are described briefly here along with a more recent one, the Cognitive Ability for Novelty in Acquisition and Learning as applied to foreign language test, CANAL-FT (Grigorenko, Sternberg and Ehrman 2000).

2.2.3.1 Aptitude Tests

MLAT

The MLAT came into being in the late 1950s as a result of a process whereby Carroll and Sapon first conducted numerous potential tests to see what factors best predicted efficient language learning. The tests eventually included in the MLAT had to sample these factors, the components of aptitude, and not overlap with each other (Skehan 1989: 28). Dörnyei (2005: 35–36), at least, is very vocal about the MLAT not having an actual theoretical basis but instead including

what seemed to discriminate between better and worse students. The same he claims to be true of the PLAB.

Carroll and Sapon ended up with five subtests (e.g. Skehan 1989: 28–29): The first subtest is that of *number learning*, which involves partly learning the Kurdish number system. This subtest supposedly involves associative memory (rote learning ability) and perhaps also inductive language learning ability (Carroll 1981: 109). The second subtest, *phonetic script*, assesses phonemic/phonetic coding ability through participants attempting to learn a partial system of phonetic notation for English. The subtest apparently also taps general intellectual ability and memory (Carroll 1990: 18, as cited in Sawyer and Ranta 2001: 324–325). The third subtest, *spelling clues*, involves finding synonyms from a written list of words based on pronunciation clues, i.e. identifying pseudohomophones, and is a highly speeded test. The test is assumed to draw on native language vocabulary and phonemic/phonetic coding ability. The fourth subtest, *words in sentences*, is a test of grammatical sensitivity and requires deciding which of the underlined words in a sentence fills the same grammatical function as the one underlined word in another sentence. Finally, the fifth subtest, *paired associates*, a multiple choice vocabulary task in which Kurdish-English word pairs are learned, taps associative memory (rote learning).

The MLAT was originally designed for native speakers of English aged 14 or older but different versions exist for different ages and L1s (for examples, see e.g. Carroll 1981: 92). According to Carroll (1981: 91), the MLAT enables valid prediction of success in foreign language courses (non-intensive) and goes well beyond verbal intelligence in its predictive power. In Carroll's words (1981: 94): "Foreign language aptitude as measured by the MLAT, then, seems to consist of some special cognitive talent or group of talents that is largely independent of intelligence, and operates independently of the motivations and attitudes of the learner."

PLAB

Pimsleur designed his test battery in the 1960s for learners aged 13–19 in the belief that poor auditory ability was behind many problems in foreign language learning (e.g. Skehan 1989: 29). Compared to the MLAT, the emphasis in the PLAB is on inductive language learning and auditory ability (Skehan 1989: 29). Language learning aptitude was thought by Pimsleur to consist in the main of verbal intelligence, motivation and auditory ability (Dörnyei 2005: 40 and Wesche 1981: 120).

The battery comprises six components (e.g. Dörnyei 2005: 38–39). The first is *grade point average*, including the most recent grades in English, history, math and science. The second is *interest in foreign language learning* (modern foreign languages), which the participants self-rate on a five-point scale. The third part is the first actual test, a multiple choice measure of *vocabulary*. Participants are asked to select synonyms for a total of 24 adjectives. For each one, four options

are given. The fourth part is *language analysis*, a test where a nonsense language equivalent is given for English words and, using those correspondences as clues, the participants have to figure out which of the nonsense language options presented goes with a given English phrase. The fifth part assesses *sound discrimination* by requiring the participants to learn three words in a foreign language and after hearing a stretch of the language spoken, to indicate which of the previously learned words were included in the spoken part. The sixth, and final, part deals with *sound-symbol association*. The participants hear nonsense words and pick them out from among four written alternatives.

Dörnyei (2005: 36) comments that while the first two parts, grade point average and interest in foreign language learning clearly stand out, they are included, in accordance with what appears to be the tradition in psychometric test development, because they seem to discriminate between learners, not because the grade in history as such, for example, is actually thought to reflect language learning aptitude. As Skehan (1989: 29) notes, the PLAB does not include a test of grammatical sensitivity or really cover memory and consequently, as it appears that aptitude to Pimsleur means language analytic ability and auditory ability, the view of aptitude is narrower than that reflected in Carroll's MLAT. On the other hand, Pimsleur's view of language learning ability could be viewed as broader since he includes motivation, which is more commonly seen as an independent factor, and not part of aptitude (Dörnyei 2005: 40). Skehan (1989: 29) attributes these differences to Carroll's and Pimsleur's backgrounds in psychology and linguistics, respectively.

As Wesche (1981: 121) puts it, both Carroll and Pimsleur had the idea that the knowledge about the strengths and weaknesses of students, gained by an aptitude test, could be used to modify teaching or curricula. The PLAB, at least, has been used diagnostically to screen students for remedial instruction (Dörnyei and Skehan 2003: 594). Furthermore, Wesche (1981) describes a Canadian language training program in which aptitude tests were used to match learners with the teaching methods that best suited them. It is reasonable to believe that learners who are, for example, either analytic or memory-oriented do better and are happier in conditions (teaching methods) suited to their particular aptitude.

CANAL-FT

A representative of one of the newer research directions regarding aptitude is the CANAL-FT (Grigorenko et al. 2000). It is based on Sternberg's (e.g. 1988) theory of human intelligence in which intelligence is seen as a complex of three components: analytical, creative and practical, or metacomponents, knowledge-acquisition components and performance components.

According to Grigorenko et al. (2000), the CANAL-F theory, and the test that goes with it, emphasizes being able to cope with novelty and ambiguity as cornerstones of foreign language learning. The test itself is a simulated, naturalistic language learning situation where an artificial language, Ursulu, is

gradually introduced to the participants, who then take part in several small learning tasks which entail five knowledge acquisition processes: selective encoding, accidental encoding, selective comparison, selective transfer, and selective combination (for details on these, see e.g. Grigorenko et al. 2000: 392). The processes are then operationalized at the lexical, morphological, semantic and syntactic levels of language and both input and output are visual and oral. Furthermore, in the CANAL-FT, language learning is understood to involve encoding in working memory and storing in and retrieving from long-term memory, all of which are assessed by immediate and delayed recall tasks.

The test comprises nine sections (Grigorenko et al. 2000: 394–396): five for immediate recall and identical ones for delayed recall, except for the fifth immediate recall section which does not have a corresponding delayed section. The first section is *learning meanings of neologisms from context*, where 24 short paragraphs are presented to participants, half of them orally, half visually, and the participants have to guess which of five English alternatives corresponds to each unknown neologism in the paragraph. The second section, *understanding the meaning of passages*, consists of six items that are identical to those in the previous section, but this time the understanding of the whole passage is under scrutiny. The third section deals with *continuous paired-associate learning*, and involves the learning of 60 word-pairs, again half visual, half oral, and in the assessment the participants are required to produce the correct paired associate in English in half of the cases and in Ursulu in the other half. In *sentential inference*, the fourth section, 20 sets of three to five Ursulu sentences are presented to participants (half visually, half orally) together with their English translations. After this, the participants have to pick the best translation out of five options for a completely new sentence (in both directions of translation). The final, fifth, section entails *learning language rules*. Throughout the test, participants are given some vocabulary, grammar and examples of the workings of Ursulu and expected to learn some of the most salient rules of the language. This learning is tested with 12 items measuring understanding of Ursulu.

Dörnyei (2005: 50) emphasizes that the CANAL-FT is theory driven, unlike the MLAT and PLAB, and Grigorenko et al. (2000: 392) take the same view: their test is not empirically derived but based on a cognitive theory of knowledge acquisition. It is also dynamic, as participants are learning as they are being tested.

2.2.3.2 Aptitude: Definitions and Relation to Working Memory

On a more general note, aptitude appears as an important concept in L2 acquisition/learning. For example, Ehrman and Oxford (1995, cited in Dörnyei 2005) found that out of the individual difference variables used in their large-scale survey, the aptitude measures shared the highest correlation with L2 proficiency. Going into memory, the most likely channel for memory to figure in individual differences is aptitude. However, before going into detail about

what aptitude is currently considered to mean, a few words on the history of the concept are in order.

Aptitude, which was strongly associated with the audio-lingual method, fell into disfavor with the emergence of communicative language learning (Ranta 2008: 142). It was thought that because aptitude was so strongly linked to the old methods, its measurement, with the existing test batteries, in the context of a new and different method of teaching and learning would probably bring few, if any, benefits. Also, the idea that people were categorized according to their scores on these aptitude measures, and thus predetermined to succeed or fail, did not sit well with researchers (e.g. Dörnyei and Skehan 2003: 593). Furthermore, aptitude was defined quite simplistically as speed of language learning (Ranta 2008: 142). For a more detailed review of the criticism on aptitude see Skehan (1989: 38–44).

There has also been some debate on the kinds of learning contexts in which aptitude may play a role. Krashen (1981) assumed that aptitude is only relevant in the classroom through its role in conscious learning, whereas Dörnyei and Skehan (2003: 595) conclude that aptitude differences mean at least as much in more naturalistic acquisition environments as in formal learning largely because there are no suggested learning sequences and more is left up to the learners themselves. However, aptitude does appear to presuppose a focus on form (Dörnyei and Skehan 2003: 600).

Furthermore, Dörnyei and Skehan (2003: 503) even attribute some of the blame for the decline of interest in aptitude to the MLAT. Carroll's model has four components whereas the MLAT has five subtests that mix the underlying components. This discrepancy did not help to clarify a rather shaky concept or bode well for the reputation of aptitude. The reasons for the fairly recent revival of aptitude, according to Dörnyei (2005: 42–43), are due to there having been advances in cognitive psychology which have led to a better understanding of "the various mental skills and aptitudes that made up the composite language learning ability". In fact, Dörnyei and Skehan (2003: 622) declare that relating aptitude constructs to acquisitional processes is crucial for aptitude studies.

Just as with the concept of individual differences, the problem is one of varying terminology, different definitions and no real agreement on what aptitude actually is. Dörnyei (2005: 33–34) states that language aptitude as such does not exist; instead, it is a composite measure of several cognitive factors. One often cited set of factors considered to constitute aptitude is that proposed by Carroll (1965, as cited in Skehan 1989: 26). This standard view on aptitude has four components: phonemic coding ability, grammatical sensitivity, inductive language learning ability, and rote learning activity for foreign language materials. It is the last of these components that explicitly deals with memory in learning, more specifically associative memory. This component is to do with making connections between stimuli and responses (words in the native and target languages), and learners are different in how efficient they are in making these connections and correspondingly different in language achievement (especially vocabulary growth) (Skehan 1989: 27). On a more

general level, Robinson (2005: 46) defines L2 learning aptitude as the individual strengths "in the cognitive abilities information processing draws on during L2 learning and performance in various contexts and at different stages".

According to Skehan (2002: 75–76), memory is the component of aptitude that has been looked at the most. In Carroll's four-component view of aptitude and in the corresponding MLAT test battery memory is linked to the auditory component of phonemic coding ability, although only associative memory is explicitly mentioned (Dörnyei and Skehan 2003: 592). This, according to Dörnyei and Skehan (2003: 593), reflects the time when the test was created and when associative memory, memory as bonds, was the prevailing understanding of memory, however limited that may seem today.

Nevertheless, working memory has been studied in the context of aptitude as well, and Dörnyei and Skehan (2003: 596–597) list it as a potential aptitude construct in their table of SLA stages and corresponding aptitude constructs. Skehan (1982, as cited in Skehan 1989: 30–31) did not find a connection between size of working memory and success in language learning, but very little is said about the methods used. However, Skehan (2002: 75–76) mentions Harrington and Sawyer (1992) and Robinson (2002b) as having found positive correlations between working memory and language learning. They used the Reading Span test by Daneman and Carpenter to assess working memory. Yoshimura (2001, cited in Ranta 2008: 143) also found a correlation between aptitude test scores and working memory span.

Sáfár and Kormos (2008), among other issues, studied the relations between complex verbal working memory capacity, phonological short-term memory and language aptitude in Hungarian beginning learners of English. The participants were teenagers taking part in an intensive English language training program. Their working memory capacity (backward digit span), phonological short-term memory (nonword span) and language aptitude (HUNLAT) were assessed both at the beginning and end of the academic year. It was found that the digit span scores correlated moderately with aptitude, especially its language analysis subtest. From this, it was concluded that working memory capacity is an important factor in language aptitude and language learning. In contrast, nonword span did not correlate with aptitude, leading Sáfár and Kormos to deduce that phonological short-term capacity is not related to foreign language aptitude.

In further support of the prominent role of working memory in language learning aptitude, Miyake and Friedman (1998, cited in Dörnyei 2005: 55) state that working memory for language may be an essential part of language aptitude, while Sawyer and Ranta (2001) consider working memory as possibly crucial in refining the idea of language aptitude and considering aptitude in relation to the SLA process. Furthermore, McLaughlin (1995, cited in Ranta 2008: 143) has suggested that the predictive power of aptitude tests may be due to working memory capacity.

How language and language learning are viewed obviously has an effect on the perceived role of memory in the latter. According to Skehan (1998, 2002,

cited in Ranta 2008: 143), the components of aptitude may play a part at certain points in language acquisition, and that memory would be important in the final lexicalizing stage in achieving fluency. In more detail, Skehan (2002: 90–91) lists SLA processing stages (noticing, patterning, controlling and lexicalizing) and what aptitude components are involved in or contribute to them. Working memory is mentioned in the context of the first stage, noticing, and memory in the final stage, lexicalizing. Sawyer and Ranta (2001: 342) track down the role of working memory in learning through its effect on attention: working memory capacity sets limits on what attention can be focused on, and this in turn has an effect on noticing, which is thought to be crucial for learning. Following this logic, a greater working memory capacity would then lead to more learning.

From a more pedagogical perspective, aptitude test scores have been used to identify learner types, i.e. whether there are learners with a particular set of strengths and weaknesses as far as measures of aptitude are concerned (Skehan 1989: 36). This may be used, for example, to assign learners into groups receiving different kinds of instruction. Skehan (1989: 37) seems to arrive at the conclusion that learners may be divided at least into analytic/linguistic and memory-based learners, both of whom are included in the aptitude research.

In a similar vein, Robinson (2002a: 114) proposes a framework where aptitude complexes (sets of cognitive abilities) "are differentially related to language learning under different psycholinguistic processing conditions". This means that learners have different strengths and are likely to benefit from different learning conditions, among others.

Robinson (2002b) studied individual differences during implicit, incidental and explicit learning, and also gained insights into the role of working memory. According to Robinson (2002b: 211–212), in an implicit learning task, memory-based learning is encouraged by the learning condition but there is no awareness of any rules regarding what is being learnt or any intention to find out what the rules are. Incidental learning takes place when stimuli are processed for meaning, but rule-governed structures are learnt unintentionally on the side. Usually the learner is aware of picking up structures. Finally, in an explicit learning task, the learner is told that the material follows a rule-governed structure and is asked intentionally to figure out the rules while processing the material. The rules may even be revealed beforehand, in which case the task is to apply them.

Working memory was found to predict incidental learning and the individual difference measures of aptitude and working memory were those that incidental learning was the most sensitive to (Robinson 2002b: 256). There was also some support for the claim that "L2 learning under any condition of exposure will be sensitive to IDs in cognitive abilities and resources, where these are relevant to the processing demands of the particular learning task or condition" (Robinson 2002b: 262).

Based on the above, it seems safe to say that in terms of individual differences and especially aptitude, working memory is a central element. In fact, Dörnyei (2005: 35) suggests that the concept of aptitude or the use of such

an umbrella term needs to be revised, among other reasons because of "recent research into specific cognitive skills and capacities related to learning", such as working memory. Aptitude also appears to be closely connected to the information processing view of language learning discussed above in section 1 as Skehan (1998: 203) states that the different components of aptitude can be related to stages of information processing. Dörnyei and Skehan (2003: 596) also affirm that "aptitude, at a fairly general level, is consistent with a cognitive view of SLA".

Now that the role of working memory in cognitive theories of language learning in general and in the study of individual differences has been discussed in some detail, it is time to focus on the concept of working memory. In the next chapter, different perceptions of working memory are considered before moving on to a more detailed description of the model of working memory used as the starting point of the present study.

3 (WORKING) MEMORY FOR LANGUAGE

There are numerous theories and models of how human memory may be structured, and while some models have gained more support than others, there is no unanimity on this issue. One of the most widely researched memory models which contains a separate unit for working memory is the Baddeley and Hitch (1974) model, later revised by Baddeley (e.g. 1986, 2000). The present study is largely based on the model in question and a detailed description of it is given below in section 3.1; nevertheless, however dominant in the field, it is still just one of several models of working memory. In other words, while the present study is linked to a specific model and the methods chosen accordingly, not all research on the connection between short-term memory and L2 stems from the same model. Depending on the model studied, methods can be found in the literature that are not universally compatible with all the memory models. It is important to be aware of some of this variety and realize that when the relationship between short-term memory and language, be it L1 or L2, is discussed, it is not always the same theory of memory that is being tested. The purpose of this section is briefly to draw attention to the variety of models and foreground the most obvious differences between them in order to show where the model used in this study stands among them. Also the origins of the wide terminology used in the field are addressed, as are doubts about the whole concept of working memory.

Whether memory is even divisible into different systems has been debated, but the view that memory consists of two subsystems has gained increasing support. Already by the late 1960s, numerous models existed that were built around the dual concept of memory, i.e. that there are two types of memory, short- and long-term memory. By the early 1970s, there seemed to be agreement that a separate short-term store or memory had to be assumed to exist (Baddeley 1986: 3–17). The term *short-term memory* is often used to refer to the counterpart for more long-term memory without necessarily taking a stand on the actual functions taking place in the memory system. However, a distinction between short-term and working memory is usually made, short-term memory referring to an earlier conception of passive storage, and working memory to a

more active structure that actually processes information in addition to storing it. The term *working memory* is a much later invention and has been around since the mid-1970s (Lehto 1996: 1). Although this difference may seem clear cut difference and a somewhat radical change in theoretical thinking, the exact difference between short-term memory and working memory remains unclear (Miyake and Shah 1999: 2).

In comparing views on working memory, several issues should be borne in mind. First of all, the term working memory can be used with varying meanings, rendering comparisons difficult. Gathercole and Baddeley (1993a: 2–3) mention at least two different approaches: theories that use working memory to mean a general processing system with limited capacity (e.g. Kintsch and van Dijk 1978) and theories in which the term working memory refers to the necessary unit of production system models of cognition without stating for certain that it is of limited capacity (Anderson 1983). However, common to all the working memory models is that they involve both temporary storage of information and transforming or processing it (see e.g. Andrade 2001: 5).

As discussed in section 2.1, the relationship between working memory and long-term memory can be considered to be reciprocal (see e.g. Ellis and Sinclair 1996). According to such views, short-term representation and rehearsal have to be accurate for long-term representations of language to be formed. This comes down to a well functioning working memory which offers the temporary storage and rehearsal space necessary in the process. Existing language information in the long-term memory then adds to working memory capacity by easing the processing load through, for example, chunking, allowing better use to be made of the limited capacity of working memory (Ellis and Sinclair 1996). Working memory is thus an important part of language learning and use, and a good working memory capacity can be considered to contribute to the speedy forming of accurate long-term representations of language.

This division of labor, however, is not as clear in all models. Kintsch, Healy, Hegarty, Penington and Salthouse (1999: 427–428) present a categorization of working memory models based precisely on how the models view the relationship of working memory to long-term memory and knowledge. In three models – the multiple-component model, the EPIC model and the CAP2 architecture – clear distinctions are made between working memory and long-term memory. In contrast, in a number of models there is thought to be a more continuous relationship between the two types of memory, with working memory considered as an activated part of long-term memory. This is the case in the embedded-processes model, the "controlled attention" framework, the ACT-R model and the biologically based computational model. Finally, some models do not express such a clear view regarding the relationship between working memory and long-term memory (e.g. the ICS model) or treat the issue as a more complex one (the Soar architecture and the long-term working memory).

Another interesting question addressed by Miyake and Shah (1999) is the structure of working memory itself. A very basic division in working memory models is made between models that view working memory as a unitary construct and those that consider it non-unitary. Furthermore, if working memory is regarded as non-unitary, the question of the types of subsystems has to be addressed: on what grounds is the division into subsystems made, and is there anything specifically devoted to language? Kintsch et al. (1999: 418) note that while researchers disagree on how many domain-specific representational systems there are in working memory, most seem more or less to agree that such systems exist.

When it comes to language skills, the focus in fractionated and integrated models is different (Adams and Willis 2001: 91): the fractionated models are more interested in language production whereas the integrated models, with their complex span tasks, aim to understand language comprehension. This division into fractionated and integrated is the basis of French's (2003) grouping of working memory models into three categories. First, there are models that view working memory as a unitary construct (e.g. Engle, Cantor and Carullo 1992; Kyllonen and Christal 1990). Second, working memory is similarly viewed as unitary but also as limited when it comes to the types of tasks it is involved in. Just and Carpenter (1992), for example, limit it to language, whereas Conway and Engle (1996), for example, consider it relevant for language and possibly other tasks as well. Finally, there are the multi-componential models such as that of Baddeley and Hitch (1974), which we focus on in the following section.

3.1 The Baddeley and Hitch Model of Working Memory

Among the dozens of theoretical options, the working memory model that was originally created by Baddeley and Hitch, and later developed further by Baddeley, is the core of the present study, since the principal interest is precisely in the phonological loop component of the model and its relation to various language tasks. The phonological loop concept in the Baddeley and Hitch model (1974; Baddeley 1986) has been considered to have been developed to explain performance in language processing and verbal memory tasks (Service 1989: 8). Furthermore, the model has been deemed to be the dominant one and it has been applied because of the surmounting evidence of the link between the phonological loop and language development (French 2003: 16–19). Indeed, out of the numerous working memory models that exist, that of Baddeley and Hitch is perhaps the most detailed one in the sense that it makes claims about certain memory mechanisms that are well suited to language learning and use. Such detailed theorizing about such language-crucial issues are not, apparently, found in the other models. This could make it much more difficult to try and develop practical applications for studying language learning or development based on any other model, meaning that researchers

with such interests have adopted the Baddeley and Hitch model in great numbers.

The strengths of the Baddeley and Hitch working memory model that have emerged in the writings of researchers using this particular model in their work are summed up by Andrade (2001: 281–282). In short, they are the breadth, specificity and simplicity of the model, and also the fact that it already has an important position in cognitive psychology. The breadth of the model means that it covers both auditory and visual representations, and both manipulation and temporary storage of those representations. It is also specific since it separates verbal and visuo-spatial subsystems and also storage and processing. Especially in the view of immediate serial recall performance, the simplicity of the model is an obvious advantage for researchers, who can use it as a basis for their own quantitative models of decay and rehearsal.

Some alternative models are considered by Adams and Willis (2001) as options for a working memory model that could possibly be used in their line of research. One of those considered is Engle's model (an adaptation of Cowan's model) where working memory is a combination of long-term memory representations activated above a certain threshold and working memory capacity, which is considered domain-free limited capacity controlled attention (Engle et al. 1999, cited in Adams and Willis 2001: 90). However, the fractionated model has been found more appealing because in it separate processes are indicated as affecting comprehension and production skills (the central executive and phonological working memory respectively) (Adams and Willis 2001: 94).

Several models that aim to explain the phenomenon of serial recall of verbal material are discussed by Page and Henson (2001: 194–195). These models favor the Baddeley and Hitch model because it is based on such a sound theoretical framework that building models of language processes on it is easy. Lovatt and Avons (2001: 202) also lean towards this particular model and see the parsimony of the phonological loop model as its strongest attraction. It is simple, yet it makes strong predictions. The predictions that the model makes about immediate serial recall, temporary decay and rehearsal are considered "clear and testable", in addition to which the connections are well covered in a broad array of applied research (Lovatt and Avons 2001: 213). Finally, Service (1989: 8) valued the way in which Baddeley and Hitch (1974) bring together memory theory and information processing. This makes the model ideal for those interested in language processing in language development. In comparison to other models, this model has also created a vast amount of empirical research. The next section describes the model and some ways of operationalizing it in more detail.

As evident from the above, several theories and models have been proposed on how human memory, especially working memory, operates and how it is structured. The present study is built on one of the first and best established working memory models, that of Baddeley and Hitch (1974), which is introduced in the present section with special focus on one aspect of it,

phonological working memory. Since the nonword repetition test was the key instrument in the present study, the basic structure of the nonword test along with the idea behind using nonwords are explained in brief. The description of the model and its operationalization are then followed in Chapter 4 by a review of previous studies using mainly nonword repetition tests in studying phonological working memory in connection with language development. Studies conducted in the L1 context are reviewed first, followed by a more detailed account of studies on the role of phonological working memory in L2 development.

3.1.1 Structure of the Model

Information processing is the key factor behind the shift from simple short-term memory thinking to working memory. The idea is that when people are involved in a cognitive task, learning or using a language, for example, the information needed in that task is held and processed in working memory (Baddeley 1986: 33–34). Baddeley and Hitch first published their theory of such processing, the working memory model, in 1974, but it has since been revised on several occasions (for further details, see Baddeley 1986 and Baddeley 2000). Originally the working memory model comprised three components: the central executive, the visuo-spatial sketchpad, and the phonological loop, as illustrated in Figure 1 (Baddeley and Hitch 1974; Baddeley 1986; Baddeley 1992; Gathercole and Baddeley 1993a: 4). A fourth component, the episodic buffer, was added to the model at a later stage (see Baddeley 2000).

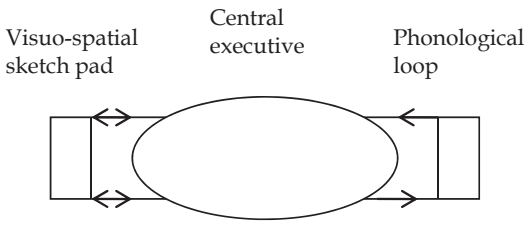

FIGURE 1 The working memory model as illustrated in Gathercole and Baddeley (1993a: 4).

The central executive is in a sense the most crucial element of working memory with the visuo-spatial sketchpad and phonological working memory as mere "slave-systems" of it (Gathercole and Baddeley 1993a: 4–5). The central executive has numerous different functions to perform, but limited capacity to do all of these. For example, it regulates information flow within the working memory system, retrieves information from other parts of memory, such as long-term memory, and it processes as well as stores information. The efficiency of the central executive depends on the amount of simultaneous demands on it:

The more there is for the central executive to cope with at one time, the less successful it is in its functioning.

The tasks of the two slave systems of the central executive are more specialized, as they deal with two different types of information. The visuo-spatial sketchpad is in charge of processing and maintaining visual and spatial information, whereas the phonological loop, also known as phonological working memory, deals with verbal information (Lehto 1996: 15; Gathercole and Baddeley 1993a). The later addition to the working memory model, the episodic buffer, is a limited capacity system that is dependent on executive processing and incapable of controlling attention (Baddeley 2000: 421–422; Baddeley 2003: 203; Andrade 2001: 302). It mainly stores information that is combined into chunks or episodes from several sources and from different modalities. That is to say it holds integrated information from the two slave systems and long-term memory as one representation, thus forming a type of interface between working memory and long-term memory. Furthermore, it is assumed that the episodic buffer forms a basis for conscious awareness.

Out of the three original parts of the working memory model the phonological loop or phonological working memory has attracted perhaps the most research interest. It is also of the most importance for the present study, since the task most commonly used in operationalizing phonological memory, the nonword repetition test, is under scrutiny here. Therefore, both phonological working memory and the nonword repetition test are described in some detail below.

3.1.2 Phonological Working Memory

The phonological memory or the phonological loop component of working memory consists of two parts: a phonological store and an articulatory rehearsal process. The structure of the phonological loop is illustrated in Figure 2. The figure also shows that speech input gains access to the phonological store directly whereas nonspeech input has to enter through the articulatory rehearsal process. Another function of the rehearsal process has to do with maintaining information. A stored phonological code fades gradually in the phonological store if not kept fresh through the rehearsal process (Baddeley 1986: 84; Gathercole and Baddeley 1993a: 8). Baddeley (1992: 558) lists several phenomena associated with and supporting the hypothetical phonological loop component of working memory, including the acoustic similarity effect, the irrelevant speech effect, the word-length effect, and articulatory suppression.

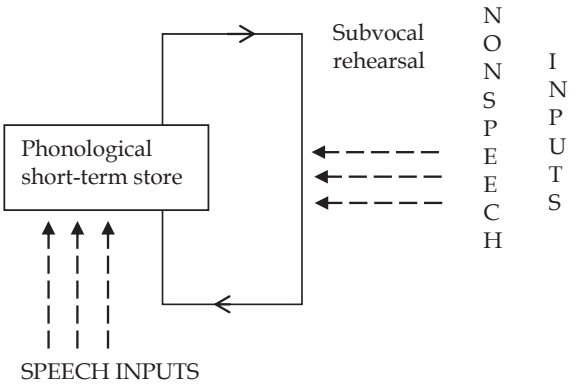

FIGURE 2 The phonological loop model as illustrated in Gathercole and Baddeley (1993a: 8)

With respect to working memory, age is a factor. According to Gathercole and Baddeley (1993a: 25–26), working memory is something that already exists in newborns and develops further with age. Research indicates that from early childhood to adulthood there is a great increase in the amount of verbal information that can be held in memory for short periods of time. The likely reasons behind this development are twofold. First, the operating efficiency of each of the components of working memory appears to increase over the years and second, different strategies are learnt with age and applied more efficiently in order to maximize the functioning of working memory.

Looking at phonological working memory development specifically, as described by Gathercole, Willis, Emslie, and Baddeley (1991: 365), phonological working memory develops very rapidly in the early school years, reaching the adult level at about the age of 12. According to Gathercole and Baddeley (1993a: 26–31), this development is a reflection of changes in the rate of subvocal rehearsal. While the phonological loop appears to be fully functional in children as young as four years old, its rehearsal component is not very efficient before children are somewhat older (e.g. Gathercole and Hitch 1993). There seems to be a connection between the subvocal rehearsal rate and speaking rate: as the speaking rate increases, so does the subvocal rehearsal rate, meaning that more information can be held in the phonological loop, resulting in a better phonological working memory.

Whether phonological working memory functions properly or not is significant since it is reflected in the ability to process verbal material. In fact, Gathercole and Baddeley (1993a: 70–73) consider it a key factor in the long-term learning of languages. In learning new vocabulary, for example, the phonological information is only transferred "into some more permanent knowledge structure in the lexical-semantic memory system" after being temporarily held in phonological working memory (Gathercole and Baddeley 1993a: 71). Should the temporary trace in phonological working memory not be

distinct and durable enough, forming more permanent traces in long-term memory is unlikely, or at least very difficult. Problems with phonological working memory can lead to the phonological material being encoded defectively or being lost quickly (Gathercole and Baddeley (1993a: 71). By assessing phonological working memory performance it is possible to predict or explain possible language learning problems. One such assessment tool, the nonword repetition test, is introduced below in section 3.1.3; however, before this, the issue of the trainability of phonological working memory is discussed briefly.

There are conflicting views about whether phonological working memory is a fixed trait or trainable. Whether individuals can be trained to improve their phonological working memory and whether a possible improvement would have an effect on L2 learning has not been much studied (Hummel and French 2010: 379). M. Dufva and Vauras (2002) found that an intervention program where Finnish learners of English focused on typical English phonemes and combinations of phonemes by listening to them and pronouncing them had a significant effect on their phonological working memory. They also cite an experiment by Service (1993) in which extra teaching of L2 English phonology for children with poor phonological working memory was found to support the functioning of phonological working memory.

On the basis of the existing research it is not clear whether targeted teaching or other training is actually about improving phonological working memory or whether it represents ways to get around or compensate for poor phonological working memory. Also, as suggested by Hummel and French (2010), in some cases the training actually results in more and better long-term knowledge of the L2, which is then reflected in better success on certain measures of phonological working memory while the basic or general phonological working memory may have remained unaltered. Indeed, most of the literature appears to support the idea that learners should be guided to make the best out of the phonological working memory they have, regardless of its capacity, instead of trying to mold the memory component itself.

Service stated in 1989 (155–156) that phonological working memory may not change but suggested that teaching compensation strategies might be possible as, for example, teaching associations between phonological, articulatory and orthographic forms might help learners to deal with new phonological forms. French (2003: 151) suggests that if accurate and permanent representations of vocabulary items in the L2 mental lexicon are thought to be due to rehearsing them in phonological working memory as, for example, Ellis and Sinclair (1996: 246–247) claim, then it might be wise for teachers to include such rehearsal in classroom activities. However, metacognitive strategies should also be taught to learners so that they can practice this kind of rehearsal on their own.

Gathercole and Alloway (2008) suggest adapting classroom techniques and materials so that learners can make maximal use of their (phonological) working memory capacity whatever it may be like. They suggest specific

classroom strategies that can be used to compensate for poor working memory in general. However, many of them also ease the load on phonological working memory. The strategies suggested include, for example, cutting down on material and using more meaningful and familiar material in order to lessen the load on working memory, repeating things and developing learners' use of memory supporting strategies such as note-taking.

Hummel and French (2010) speculate that communicative classrooms may be problematic due to their heavy emphasis on oral input, which in turn places considerable demands on information processing in phonological working memory. This, of course, applies to any learning context that relies primarily on oral input. Making more use of written and other visual formats might be beneficial as those with limited memory capacity might be able to compensate for this through support gained from visual cues.

3.1.3 Nonword Repetition Test

There are several ways to operationalize phonological working memory. One of the most widely used is the nonword repetition test. Numerous variants of the test can be found in the literature. One standardized version, described here, is the Children's test of nonword repetition (CNRep) by Gathercole, Willis, Baddeley, and Emslie (1994), a widely used and translated test of nonword repetition. The description of a basic nonword repetition task below is primarily based on that of CNRep by Gathercole et al. (1994).

In the present context, nonwords are phonological items that conform to the phonotactics of a given language, but are meaningless and as such not real words of that language (see e.g. Gathercole et al. 1994). In a way then, they mimic whichever language is needed for the purposes of the test in a particular study, usually the L1 or a language that the participants are learning. The idea is that while the nonwords could be real words in the language, the made-up forms are equally unknown to all of the participants.

One feature deemed crucial as regards the usefulness of nonwords is their wordlikeness (e.g. Gathercole, Willis, Emslie and Baddeley 1991). Participants will use their long-term knowledge of the target language as an aid in nonword repetition as much as possible. This is easier with nonwords that resemble existing, known words in their sound structure. Therefore, the more unfamiliar the structure of the nonword is, the less wordlike it is and as a result the repetition score is more a reflection of phonological working memory capacity than existing lexical knowledge. There is no absolute measure of wordlikeness but, for example, subjective ratings by native speakers have been used. Determining the wordlikeness of nonwords is not simple as it may mean different things to people with different vocabularies, as, for example, when adults assess the wordlikeness of nonwords to be used in assessing children (e.g. Smith 2006).

Other issues suggested to be of importance in selecting nonwords for a working memory measure are prosodic structure (syllabic and metric structure) (van der Lely and Gallon 2006: 592) and phonotactic probability and density of

phonological neighborhoods, which may enable existing phonological knowledge to influence nonword repetition (Vitevitch 2006: 595)

In a typical setting (e.g. Gathercole et al. 1994), the nonword repetition test participant hears nonwords that are either taped or spoken live and is requested to attempt to repeat the nonwords immediately after hearing the stimuli. The extent of the test, the number and length of the nonwords, varies according to the research agenda, but in the original CNRep, the participants repeat a total of 40 nonwords, ten of each length from one to four syllables (or two to five syllables in later versions, e.g. Masoura and Gathercole 1999). Usually the participant is allowed to get used to the idea of the task by repeating a few practice items before the actual test. The repetition attempts are usually taped and later scored by nonword or syllable. The score is then taken to represent the participants' phonological working memory capacity and often used to study the relationship between phonological working memory and some aspect of the participants' language knowledge or development. Using taped stimuli and taping the repetition attempts add to the reliability of the test as all the participants hear the exact same versions (pronunciation, stress etc.) of the nonwords and their output can be scored by several raters and listened to as many times as necessary.

Gathercole and Baddeley (1993a: 48–49) mention two advantages of using a nonword repetition test instead of some other assessment method to measure phonological working memory. First, it may be more sensitive than some of the other measures, for example digit span, since none of the participants have long-term lexical representations matching the nonwords used. This should prevent long-term memory knowledge influencing the assessment of working memory, although not even nonwords completely eliminate the effects of prior language knowledge. According to Gathercole (2006: 519–521), the lexical characteristics of the stimuli of any test, with or without nonwords, affect immediate memory performance but redintegration does not apply to the retrieval of nonwords as obviously as to the stimuli in some serial recall measures. However, nonwords, too, can be more or less wordlike, contain segments that are phonotactically frequent or syllables that are lexical units, and language knowledge can influence nonword repetition through such more familiar items. Gathercole (2006: 522) suggests that this may happen either when the item is retrieved (redintegration with the help of partially activated overlapping lexical representations) or before the storage stage during the perceptual analysis and construction of phonological representations. Most likely, sublexical familiarity facilitates phonological processing, and the phonological representation of such a nonword is of better quality than that of a less familiar nonword.

The second advantage of nonword repetition tests, according to Gathercole and Baddeley (1993a), is that the nonword repetition test is considered simple and somewhat more natural compared to some other measures, since both children and adult language learners anyway hear and repeat novel vocabulary items in the course of their learning process. The

nonword repetition test can also be viewed as simple because of its lower cognitive demands compared to digit span tasks, which benefit from the use of "higher-level strategic processes" such as cumulative rehearsal, something to which the nonword repetition test is not sensitive (Gathercole and Baddeley 1993a: 49). In Gathercole et al. (1994), comparisons were made between the CNRep and auditory digit span, a widely used index of short-term memory. In their case, digit span equaled the longest list length at which the participants could correctly repeat two out of three lists of digits. However, digit span tests vary just as nonword repetition tests do. For example, there could be just two lists of each length (e.g. Papagno and Vallar 1995).

As will be discussed in more detail in the following chapter, nonword repetition ability and vocabulary learning have been found to be closely linked. Gathercole (2006) proposes a framework for this close link which suggests that both rely on phonological storage and are multiply determined. Phonological storage is required by both nonword repetition and vocabulary learning; however, how well a nonword is stored is affected by several factors. These include the quality and persistence of phonological representation, learning conditions, and factors that affect how the phonological representation is constructed in the first place. However, phonological storage alone is not enough to determine how a person succeeds in repeating nonwords. Several cognitive, perceptual, and motor processes constrain nonword repetition and as all these processes vary between individuals and change over time, the factors behind nonword repetition success can be said to be complex at the least. The framework proposed by Gathercole (2006), the phonological storage capacity account, is by no means the only attempt to explain the link between nonword repetition and word learning. Others, such as the phonological processing account, are discussed by Gathercole (2006) and in the commentaries to her article.

In the following sections a number of studies using some form of the nonword repetition test are reviewed. They provide an array of the ways in which nonword repetition has been used to assess phonological working memory and the broad range of objectives present in such research. Below, the major division has been made between studies in L1 and L2 contexts, but more attention is paid to the latter, since they cater to the starting point of the present study and give rise to the research questions that this study attempts to answer.

4 PHONOLOGICAL WORKING MEMORY AND LANGUAGE

As should be clear from Chapter 3, there are several models of working memory; indeed, even the model by Baddeley and Hitch (1974) used here has other components besides phonological working memory. However, as the aim of the present study is to examine phonological working memory in particular, the focus of the present chapter is mostly on studies investigating language and *phonological working memory* and excludes studies looking at language and *working* from the discussion below.

4.1 Phonological Working Memory and the L1

Most studies focusing on the relationship between phonological working memory and language concern L1 learning, and usually also language development in young children. In this section a brief overview of such studies is provided, divided into sections relating to findings on certain key elements of language development.

4.1.1 L1 Vocabulary

The connection between phonological working memory and vocabulary is by far the most widely researched issue in the field. To some extent, researchers disagree on the nature and the basis of the relationship, but most claim a connection of some kind.

Among the first to reveal a significant positive correlation between phonological working memory and L1 vocabulary development were Gathercole and Baddeley (1989). They tested 104 children at ages four and five using multiple nonword tasks including nonword repetition. Vocabulary was assessed with the Short Form of the British Picture Vocabulary Scale which entailed the participants pointing to pictures named by the experimenter.

Similar results were obtained later by, among others, Avons, Wragg, Cupples and Lovegrove (1998) and Gathercole, Service, Hitch, Adams and Martin (1999).

Whether phonological memory abilities can be used to predict vocabulary learning has also interested many scholars. Gathercole, Willis, Emslie and Baddeley (1992) found a relationship between phonological memory skills and vocabulary knowledge, which appeared to be reciprocal. Michas and Henry (1994) were also able to show that phonological memory measures predict new word learning ability. In fact, studies such as that of Michas and Henry (1994) form an interesting category of their own, because they did not study natural language development, for example by gathering data about the vocabulary that the participant children had learned in the course of their lives or within a certain longer time period. Instead they introduced a vocabulary learning task to study the relationship between phonological working memory and the participants' ability to learn vocabulary items. Another example of this approach is Gathercole, Hitch, Service and Martin (1997), who gave their participants altogether four vocabulary learning tasks, three of which tested the learning of nonwords. In their study, phonological working memory was found to be significantly involved in learning the sounds of new words.

Mostly satisfied that there indeed is a meaningful connection between phonological working memory and vocabulary knowledge (and learning), many studies have been conducted in an attempt to pinpoint the mechanisms behind the connection. It is most likely that the two constructs are linked, although several other components have been found that may also influence the relationship. Gathercole, Willis, Emslie, and Baddeley (1991) proposed that there are significant correlations between nonword repetition results and vocabulary knowledge because both draw on phonological working memory capacity. An alternative explanation is that good vocabulary knowledge aids in nonword repetition. A third possibility is that participants' performance in tasks assessing phonological memory and vocabulary has a common source. Snowling, Chiat and Hulme (1991) suspected that a number of phonological processes could lie behind the connection, such as phonological segmentation, assembly of articulatory instructions, perceptual problems, or difficulties with the storage of phonological information. They were also critical of the stimuli used in nonword tasks, emphasising the need to pay attention to the prosodic structure of nonwords and the help from affixes. Gathercole, Willis and Baddeley (1991), however, argued against the influence of such non-memory processes on the basis of the findings from their previous studies.

Others, too, have paid attention to the role of nonwords in the relationship between phonological memory and vocabulary. Gathercole (1995) found the wordlikeness of nonwords to have a positive influence on the relationship, whereas Dollaghan, Biber and Campbell (1995) argued that both lexical and non-lexical morphemes integrated into nonwords have an effect on the results of a nonword test because of the inevitable use of existing lexical information in repeating nonwords. Metsala (1999) went as far as to claim that the associations found between nonword repetition and vocabulary are not due to phonological

working memory but to the underlying structure of lexical items, with phonological awareness mediating the relationship.

In a study assessing the role of speech output skills, Gathercole et al. (1999) arrived at the conclusion that they do not explain the link between nonword repetition and vocabulary skills. A similar result can be read in Adams and Gathercole (2000), who used verbal memory tasks that required both spoken and non-spoken output and were able to determine that the relationship between memory and language development is not merely due to similar output constraints prevailing in both memory and language tasks. Bowey (1996) argued against phonological working memory having a direct causal role in vocabulary acquisition, and attributed the connection to a latent phonological processing factor affecting both phonological memory and phonological sensitivity alike. Alloway, Gathercole, Adams, Willis, Eaglen and Lamont (2005) argued against such an explanation on the basis of their results of phonological awareness measures failing to show a similar connection with language assessment as phonological working memory measures.

In summary, most research to date indicates that a relationship exists between phonological working memory and L1 vocabulary development. However, there is some doubt about the nonword repetition test as a measure of phonological memory and about the basis of the relationship altogether. Common output constraints or a more general phonological processing component have been suggested as alternative explanatory factors for the link. At this point, there appears to be a relationship between phonological working memory and vocabulary development, but the exact nature of the connection remains controversial.

4.1.2 L1 Reading

Another point of interest in connection with phonological working memory and L1 is the development of reading. Compared to vocabulary, however, reading has received much less attention. Based on the studies reviewed in brief below, there appear to be quite a few ways in which (impaired) phonological working memory may affect L1 reading. First of all, it has been argued that poor readers do not have poor vocabulary but problems with forming accurate phonological representations, which in turn lead to language encoding difficulties, and thence to memory difficulties and poorer reading performance (Brady, Poggie and Rapala 1989).

Several studies have looked into the possibility of using phonological memory ability to predict later reading skills. In one of them, the participants' phonological working memory, as assessed using nonword repetition at a stage when they were yet unable to read, turned out to predict reading skill in childhood: better phonological memory in prereading children seemed to lead to better reading achievement at age eight (Gathercole and Baddeley 1993b). Gathercole (1995) ended up with similar results when she compared nonword repetition results at age four to reading performance at age five: phonological

memory skills had an important role in the early stages of reading development, but after one year of reading its influence had diminished greatly.

Further support for a connection between phonological memory and L1 reading is provided by Muter (1998), who compared nonword repetition results at ages five and six to reading at age nine and found them both to discriminate between good and poor readers and to predict reading skill. To add a dissenting voice, van Bon and van der Pijl (1997) also found that good and poor readers differed in their nonword repetition results, but rather than explain the result by differences in phonological working memory abilities, they attributed them to the participants' phonological segmentation facility.

Working memory more broadly has been under investigation in many reading comprehension studies. Seigneuric, Ehrlich, Oakhill and Yuill (2000) found working memory to predict reading comprehension in nine-year-olds even when compared to vocabulary and decoding skills. Working memory was understood as consisting of both storage and processing of information and was assessed with two verbal tasks, two numerical tasks and one spatial task. However, only the verbal and numerical tasks predicted reading comprehension.

Swanson and Howell (2001) studied age-related differences in reading (word recognition and reading comprehension) in nine- and fourteen-year-olds. The role of two working memory systems in these differences was investigated. The central executive – the role of executive processing in coordinating processing and storage – was assessed with measures of verbal and visual working memory (sentence span, auditory digit sequencing, visual matrix and mapping and directions). The phonological loop (verbal short-term memory) was assessed with a digit span task and a word span task. Working memory and short-term memory were found to contribute independent variance to reading: domain-general working memory contributed variance to changes in word recognition and reading comprehension with age. This contribution went beyond that of short-term memory and articulatory speed. Also, verbal short-term memory processes contributed variance to word recognition and reading comprehension, and separated age groups.

Holsgrove and Garton's (2006) results were somewhat contradictory to those of Swanson and Howell (2001) on the role of the central executive. They studied the reading comprehension of 13-year-olds and the role played in this by phonological and syntactic processes. In addition, the contribution of working memory to the processes behind reading comprehension was of interest. Tests from the Working Memory Test Battery for Children were used to assess the phonological loop and the central executive separately. The storage component of the phonological loop was assessed with the NonWord List Recall Test while the Word List Matching Test was used for measuring subvocal rehearsal. The central executive was assessed with the Listening Recall Test. Both phonological and syntactic processing were found to predict reading comprehension. Phonological processing predicted reading ability better with the less skilled readers whereas syntactic processing predicted the ability better

in readers with higher skills. The phonological loop, but not the central executive, played a role in the processes involved in reading comprehension.

Kintsch and Rawson (2005: 224) conclude that while there are several models of reading comprehension, something they all have in common is the assumption that information processing takes place in a working memory that is of finite capacity. Because reading comprehension places great demands on processing facilities, individual differences in working memory capacity are reflected in comprehension. Nation (2005: 258) states that at least in poor comprehenders there is a close link between language comprehension and verbal memory, but that issues of causality are in need of further research.

4.1.3 Phonological Working Memory, L1 Language Production and School Attainment

The following four studies serve to introduce some less researched areas in the field of studies on phonological working memory in the L1 context. First, Adams and Gathercole (2000) compared groups of four-year-old children who performed relatively poorly or well on a nonword repetition task on measures of verbal and visual short-term memory and language development, more particularly spoken language. The higher nonword repetition group turned out to use a greater number of different words in their speech, and they also produced longer utterances and more varied syntactic constructions. The two groups of children also differed on measures of visual memory span, and one of the visual memory tasks in turn was connected to some of the language measures; hence although the results clearly show a significant relationship between spoken language development and phonological memory, it is not clear whether the processing mechanisms of the visuo-spatial sketchpad are also linked to language development.

Alloway et al. (2005) studied the links between numerous working memory measures and children's skills in key scholastic domains at the time of school entry, age four or five. In addition to measures of complex memory and the episodic buffer, phonological short-term memory was assessed with the CNRep as well as the digit recall test and the word recall test of the Working Memory Test Battery for Children. Such a link has been found in later school years, and the attempt was to find out whether one exists at the very beginning of the school career. Phonological working memory was indeed found to be uniquely linked to reading, speaking and listening abilities as assessed by teachers using, the Stockton on Tees Baseline Scheme. Based, in particular, on the links with speaking and listening, Alloway et al. (2005) drew the conclusion that the phonological loop has a specific role in vocabulary acquisition by affecting the long-term learning of the phonological forms of novel vocabulary.

Finally, Gathercole, Tiffany, Briscoe, Thorn and the ALSPAC team (2005) compared the school attainment of two groups of children, one with an early phonological memory deficit detected at age five and one in which the deficit persisted until retesting at age eight. The results showed that poor phonological working memory did no significant harm to language acquisition during the

early school years, as the persistent phonological memory deficit group did not show learning difficulties in any of the key domains of school attainment assessed (vocabulary, language, literacy and number skills). However, this was not viewed as evidence against the link between phonological working memory and vocabulary knowledge, because a strong connection exists in earlier childhood. The assumption was that since language abilities in middle childhood already approach those of adults, other factors besides phonological working memory become more important. In addition, phonological working memory loses in importance to the central executive in the acquisition of complex skills and knowledge in the early school years.

The same authors reported data from the same memory groups in a later article (Gathercole et al. 2008). They proposed that even though the L1 vocabulary knowledge of the low-memory children was at an age-appropriate level, they might still have problems in new word learning. The highly redundant exposure to L1 vocabulary and, at a later stage, reliance on lexical mediation in learning words could have led to the children's normal levels; however, an experimental learning task might reveal whether or not there was impaired new word learning. In fact, out of a number of learning measures (at age eight), only unfamiliar verbal material was problematic for the low-memory participants. They were impaired in learning the names of unfamiliar people (verbal recall of the Doors and People test) and in learning paired associates where words were paired with nonwords. Such learning of new phonological material is considered to be mediated by the ability to store phonological representations temporarily; however, this mechanism of word learning is a primitive one and likely to be most important at the early stages of language learning. Therefore, even though an experimental learning task showed a significant impairment, as found in Gathercole et al. (2005) above, other factors may compensate for low phonological working memory in L1 vocabulary learning across a longer time period, leading to age-appropriate levels of vocabulary knowledge.

In sum, although this review of studies on phonological working memory and L1 above has been kept brief, it is clear that the connection has been rather widely studied. The role of phonological working memory in vocabulary development has received most of the attention, but a number of other aspects of L1 knowledge have also been found to show a connection to the memory system. It is also clear that, in particular, moving from vocabulary to other areas of language knowledge necessitates the consideration of multiple predictors of L1 development, along or through which phonological working memory capacity imposes its influence. In other words, while the connection between phonological working memory and L1 development is seldom contested, the exact source or route of the connection is not agreed upon and investigations are ongoing.

The studies reviewed above in this section, although covering a wide range of issues to do with language knowledge, are a very homogenous group

in the sense that studying L1 development quite understandably entails dealing with fairly young participants. Less explicably, the language in question has been English in most cases. With the shift in focus from L1 to L2 in the following section, the participants and languages to be learned become much more varied.

4.2 Phonological Working Memory and the L2

Relatively few studies have specifically investigated the relationship between phonological working memory and L2 learning or knowledge. Below, a number of these studies are summarized and reviewed in detail. Most of these studies have focused on English as L2. This aside, the research interests appear to be very varied, ranging from the influence of phonological working memory on vocabulary learning to the language specificity of phonological working memory. Below, an attempt has been made to provide a clear and concise picture of what is currently known about the relationship between phonological working memory and L2 learning and knowledge. This is done by presenting some of the key findings of a number of previous studies individually. This presentation of the individual studies is then followed by a review of two key aspects of the studies summarized: the L2 and the phonological working memory measures used. This is done to highlight the multitude of variables that need to be considered in this field of inquiry. It also makes it easier to assess their possible contributions to our understanding of the field and the need for further study.

No specific details about the statistical methods or scoring procedures used in the studies reviewed are provided here. Similarly, not all of the tasks and results of the studies described, but only those considered of interest with respect to the present research questions on phonological working memory and L2. The studies discussed in this section are also combined in a concise format in Tables 1 and 2 below. The terminology in the sections below is somewhat varied as a deliberate attempt has been made to preserve the original terms used in these sources.

TABLE 1 Studies on phonological working memory and vocabulary.

	N	Age of Participants	L1	L2	L2 Tasks	Working Memory Tasks	Selected Results
Cheung 1996	84	12 (average)	Cantonese Chinese	English	various, main focus on vocabulary learning task	nonword span (L2)	nonword span predicted vocabulary learning speed in poorer learners
Papagno, Valentine and Baddeley 1991	24	adults	Italian or English	not relevant	learning word pairs (visual or auditory)	secondary task (tapping/"bla")	articulatory suppression more harmful to L2 than L1 vocabulary learning; phonological coding affects L2 vocabulary learning
Papagno and Vallar 1995	10 and 10 controls	adults	Italian	not relevant / polyglots	learning word pairs (Italian-Italian, Italian-Russian)	auditory digit span, nonword repetition etc., language n/a	polyglots: better phonological working memory and faster L2 learning; phonological working memory and L2 learning connected
Baddeley 1993	1 and 6 controls	adults	English (presumably)	-	learning word pairs (English-English, English-Finnish)	adult version of CNRep (+ various), language n/a	problems in phonological working memory lead to defective long-term phonological learning
Atkins and Baddeley 1998	30	adults	English (presumably)	-	learning words in pairs / short sentences, English-Finnish	several, visual and verbal	learning speed connected to verbal memory span only
Speciale, Ellis and Bywater 2004, experiment 1	38	adults	English	-	learning word pairs (English-German)	nonword repetition (nonwordlike)	nonword repetition related to receptive but not productive L2 vocabulary measures
Speciale, Ellis and Bywater 2004, experiment 2	44	adults	mostly English but also Welsh, Greek and German	Spanish	receptive vocabulary at the beginning; written exam at the end; varied tasks	nonword repetition (nonwordlike, at the end also Spanish)	long-term knowledge affects nonword repetition (nonwordlike connected to L2 knowledge less than Spanish nonword repetition)
Thorn and Gathercole 1999, experiment 1	45	4 to 9	English-French bilinguals, English monolinguals and L1English with L2French	(French)	receptive vocabulary	CNRep (English and French)	better vocabulary knowledge connected to better CNRep
Thorn and Gathercole 1999, experiment 2	25	4 to 8	native and nonnative bilinguals, English-French and French-English	English/ French	receptive and expressive vocabulary	CNRep (English and French	language-specific knowledge affects phonological short-term memory
Masoura and Gathercole 1999	45	8 to 11	Greek	English	2 translation tasks (vocabulary)	CNRep (English and Greek)	general nonword repetition capacity linked to L2 but not L1
Farnia and Geva 2011	91 + 50	6 at the beginning	LLs: Punjabi, Tamil or Portuguese; L1 English	English	receptive vocabulary	English and Hebrew nonword repetition	both repetition of familiar and unfamiliar nonwords predicted vocabulary learning but tapped different cognitive systems
Kaushanskaya and Yoo 2011	22 + 20	20	English	-	learning word pairs, familiar/ unfamiliar sound structure	vocal vs. subvocal rehearsal	vocal rehearsal benefits learning familiar sound structures; subvocal rehearsal benefits learning unfamiliar sound structures; rehearsal methods tap different cognitive processes.

TABLE 2 Studies on phonological working memory and diverse components of language knowledge.

	N	Age of Participants	L1	L2	L2 Tasks	Working Memory Tasks	Key Results
Service 1989	44	9 at the beginning	Finnish	English	English grade, listening and reading comprehension, written production	nonword repetition in L1 and L2	language learning stage affects nonword repetition; nonword repetition related to L2 skills in general, not more to some subskills
Service and Kohonen 1995	42	9 during memory assessment, 12 during L2 assessment	Finnish	English	English grade, vocabulary translation, communicative test, traditional test	data from Service 1989	connection between phonological working memory and language learning depends on vocabulary knowledge
Dufva and Voeten 1999	160	7 (average) at the beginning	Finnish	English	vocabulary, reading comprehension, writing (gaps), listening comprehension	nonword repetition (L2)	word recognition (one of literacy measures) and phonological working memory predicted L2 knowledge; the two stem from same phonological skill
Dufva and Vauras 2002	24	10 at first	Finnish	English	phonological awareness and memory, word recognition, spelling	no separate task	intervention (focusing on English sound structure) had a positive effect on phonological memory, word recognition and spelling
French 2003	54	11	French	English	overall proficiency (2 standardized tests), vocabulary, grammar	CNRep (L2 and Arabic)	phonological working memory and L2 ability connected; phonological working memory predicted L2 production in low but not high proficiency learners
Swanson, Sáez, Gerber and Leafstedt 2004	101	6 to 7	English, Spanish or bilinguals	-	vocabulary, word identification, word attack	various, phonological and working memory measures separately	better language knowledge means better access to resources from working memory; language-specific phonological working memory behind reading disabilities
O'Brien Segalowitz, Collentine and Freed 2006	43	adults	English	Spanish	OPI: productive vocabulary, narrative ability, inflectional morphemes, clauses	serial nonword recognition (L1)	phonological working memory related to vocabulary use but not development; differences between low and high ability participants
O'Brien, Segalowitz, Freed and Collentine 2007	43	adults	English	Spanish	OPI: general overall oral ability and fluidity	serial nonword recognition (L1)	phonological working memory valuable for adult L2 oral proficiency development regardless of learning environment
Kormos and Sáfár 2008	121	15 to16	Hungarian	English	reading comprehension, listening comprehension, composition, vocabulary, grammar and oral skills	nonword repetition (L1), backward digit span	phonological short-term memory and general working memory separate constructs; affect L2 learning in different ways

4.2.1 Focus on L2 Vocabulary

Although other aspects of language knowledge have not been ignored, vocabulary is by far the most investigated area in phonological working memory studies in the L2 context. While it was not initially the main focus of study, it quickly assumed this position.

One of the first studies to investigate the effect phonological working memory has on L2 learning was Service (1989). The participants were Finnish children who were learning English (L2) in school. At each testing time, four occasions at one-year intervals, the participants were required to complete a nonword repetition test both in Finnish and in English. It turned out that there was a ceiling effect on the Finnish tests, and these results were thus excluded from the study; however, the English nonword repetition test scores were compared to the participants' English grades two and a half years after the study began. A significant positive correlation was found, leading Service (1989) to the conclusion that it is possible to predict children's learning of English in the first few years by looking at their nonword repetition scores. General academic ability did not seem to have influenced the correlation.

Nonword repetition also correlated with several subskills of English equally significantly. The subskills were assessed using measures for listening and reading comprehension as well as production taken from tests conducted during the study. Because of similar correlations, Service (1989) concluded that the ability to correctly repeat nonwords is related to English skills in general and not to any particular subskills more than others.

However, using partly the same data and the same participants as in Service (1989) above, Service and Kohonen (1995) paid close attention to the relationship between nonword repetition and L2 vocabulary learning. The nonword repetition tests and scores in Service (1989) were used, but were now compared to new measures of school achievement. There were altogether four English (L2) measures: English grade, a vocabulary task where 40 items were translated in writing from English to Finnish and vice versa, a communicative test, and a traditional test. In the communicative test the participants' communicative skills were measured with four tasks: reading comprehension, reproduction of structures (writing down what was dictated), listening comprehension, and written production (for details, see Service and Kohonen 1995: 161). The traditional test measured similar skills in more traditional ways. The measures included listening and reading comprehension, knowledge of phrases, recognition of adjective forms, knowledge of words, correct word forms, and past tense forms (for details, see Service and Kohonen 1995: 161–162).

Service and Kohonen (1995) found that a relationship existed between nonword repetition scores and the different measures of English, and that the connection was not mediated by general academic achievement. However, as suspected by the researchers, when the effect of English vocabulary knowledge was controlled for, a relationship no longer existed between nonword repetitions and traditional or communicative tests of English. From this Service

and Kohonen (1995) concluded that the relationship between phonological working memory and language learning appeared to depend on vocabulary knowledge.

On the basis of these two early studies it appears that there is a fairly clear connection between phonological working memory and L2 and that it may be justified to focus on vocabulary. On the other hand, it is difficult to imagine vocabulary knowledge not having an effect on any aspect of L2 that one might wish to assess.

4.2.1.1 L2 Vocabulary Learning Speed

Within the group of studies that focus on vocabulary, a number of studies have investigated how phonological working memory may be related to the rate of vocabulary learning. Cheung (1996) studied this in an experimental situation with children who spoke Cantonese Chinese as their L1 and had been learning English since the age of three or four.

The English measures used were various, but the main focus was on the vocabulary learning task, which introduced the participants to three new English vocabulary items. The participants were required to learn the correct pronunciation and translation into Cantonese Chinese for each of the items. The score was the number of trials it took them to learn the items. Three consecutive errorless trials were considered as learning.

Phonological working memory was assessed with a nonword span task, a version of nonword repetition suitable for older participants. The task consisted of English two-syllable nonwords which the participants heard and repeated in lengthening sequences of one to seven nonwords. There were two trials at each length and the task continued until errors were made in both trials of the same length.

Cheung (1996) found that nonword span was the only measure that uniquely predicted vocabulary learning speed. However, these results were only significant in the case of the participants with English vocabulary knowledge below the group median, indicating some interaction between the effect of phonological working memory on vocabulary learning and prior vocabulary knowledge. Phonological working memory seemed thus not to predict the vocabulary learning of the better participants, the ones with English vocabulary knowledge above the group median.

Atkins and Baddeley (1998), on the other hand, studied the connection between verbal memory span and L2 vocabulary learning in adults. Verbal short-term memory span was measured with several tasks: auditory and visual versions of digit span, phonologically similar letter span, and phonologically dissimilar letter span. For visual short-term memory there was only one task, a test of visuo-spatial recall. It required the participants to reproduce patterns they were shown.

The vocabulary learning experiment included four lists of 14 English words paired with their Finnish translations. The items were either learned as pairs of words or as pairs of short two-word sentences. There were two learning

sessions one week apart and the lists were divided so that half of the pairs were learnt in the first session, half in the latter one. Learning the pairs and testing them were done individually on a computer so that the participants could take their time and focus on the items which they found difficult. Vocabulary learning speed was measured with the average number of errors the participants made during the learning process. There were two different scores: a strict one allowing no errors or omissions, and a lenient one, which allowed mistakes in spelling.

A verbal and a visual factor were found in a principal components analysis of the short-term memory tasks. Speed of learning the vocabulary items only correlated with the verbal factor, irrespective of the details, meaning all of the learning speed scores (pairs or sentences, auditory or visual presentation, strict or lenient score) correlated with verbal span only. Visuo-spatial span did not correlate with any aspect of vocabulary learning speed. Errors made one week after learning the words did not correlate with either visual or verbal short-term memory span. Atkins and Baddeley (1998) interpreted the results to mean that speed of learning new vocabulary could be predicted by the verbal span, but not by the visuo-spatial span, and that problems in short-term memory did not lead to more rapid forgetting of the vocabulary items learned.

In sum, using different kinds of tasks and studying different kinds of participants, a connection has been found between verbal memory and vocabulary learning rate. These studies are by nature somewhat artificial as they have had to isolate their target, vocabulary learning rate, in an experimental setting; however, it is easy to imagine how such a factor may be influential, at least at the early stages of language learning, even in the typical context of classroom learning.

4.2.1.2 Rehearsal in Phonological Working Memory as a Prerequisite to L2 Vocabulary Learning

The importance of subvocal rehearsal to the learning of L2 vocabulary has been the focus of some studies. Papagno, Valentine and Baddeley (1991) studied the role of the short-term phonological store in learning L2 vocabulary with Italian and English adults. The participants took part in several learning experiments in which they either heard or saw lists of eight word-pairs and were required to learn them. Half of the pairs consisted of two L1 words; in the other half L1 words were paired with L2 items. The practice and learning trials were followed by a maximum of five test trials. The total number of correctly remembered pairs signified the score for each trial.

The role of the phonological loop in all this was assessed with a secondary task the participants did while learning the word pairs. They either tapped their finger or repeated the syllable “bla”, which was assumed to prevent the participants from rehearsing the word to be learnt subvocally thus disrupting the operation of the phonological loop, or to interfere with the phonological recoding of words when they were presented visually. Articulatory suppression

turned out to be more harmful to learning L2 than L1 vocabulary, which was considered to show that short-term phonological coding is needed in acquiring L2 vocabulary. A more recent study by Kaushanskaya and Yoo (2011) reached a very similar result. Phonologically familiar words were better learnt after vocal rehearsal but subvocal rehearsal aided the retrieval of phonologically unfamiliar words.

Also Baddeley (1993) reached a similar conclusion in his case study of a man (S.R.) with a suspected phonological short-term memory deficit. The memory tests included two tasks of visual and verbal long-term memory, a task of visual short-term memory, and eight tasks of verbal short-term memory, one of which was a phonological working memory measure, an adult version of the CNRep (Gathercole and Baddeley 1989). This test required repeating nonwords in trials of three nonwords of the same length. The first nonwords only had one syllable and the length increased until an error in repetition was made.

There were also several language measures. L1 reading, spelling and vocabulary were assessed, in addition to which there was an L2 vocabulary learning task, which included several learning and test trials to learn eight pairs of English (L1) words and eight English-Finnish pairs.

It turned out that the participant S.R. had a very poor verbal memory span compared to the peer controls and that the deficit was only detectable with verbal material. The verbal recognition and name recall measures revealed that the effect of the defective phonological short-term memory was reflected in poor long-term phonological memory and learning. L1 development had not been affected, but S.R. had not been able to learn other languages and was similarly unable to do so in the learning experiment. While semantic coding enabled him to learn the L1 items, he did poorly on the English-Finnish pairs. S.R. seemed unable to employ rote rehearsal, which learning of phonologically new items would have required. Based on these findings, it was concluded that problems in phonological short-term memory were connected to defective long-term phonological learning. A well-functioning phonological loop seems therefore important largely because of its effect on long-term phonological learning, and hence poor phonological working memory can impede L2 learning.

The studies reviewed in this section have found a particular verbal memory mechanism, subvocal rehearsal, to be connected to a certain kind of vocabulary learning and that there, consequently, are differences between learning new L1 and L2 vocabulary. What is detrimental to learning novel lexical items in L2 does not necessarily affect L1 learning. This kind of information is important because it is further validation for the existence of different learning paths for L1 and L2 learners and the need to study L2 development separately in the phonological working memory framework. However, as is well known, L2 learning is not a unitary research object, as it takes place in varying contexts and learners have very different language backgrounds, all of which can have an effect on the kind of memory needed and utilized in L2 learning. This is the topic of the next section.

4.2.2 Influence of Language Background

The language background of participants has been found influential in how they succeed in tasks of phonological working memory. Papagno and Vallar (1995) were interested to see whether polyglots and non-polyglots were different when it came to phonological short-term memory and new vocabulary learning. The polyglot participants were fluent in at least three languages, and all had Italian as their L1. Also the non-polyglots were native speakers of Italian, but had studied only one L2 in school.

Both groups were tested using several measures, for example L1 vocabulary knowledge, visuo-spatial span, visuo-spatial learning, paired-associate learning of real and nonwords, and phonological working memory tasks, which included auditory digit span and a nonword repetition test. The nonword repetition test was a more complicated version suitable for the adult participant; however, no details were provided as to how the test was constructed. There was no reference to the nonwords conforming to the phonotactics of any language, but they were described as "meaningless" and "pronounceable" (Papagno and Vallar 1995: 100).

An experimental task was conducted to assess the participants' ability to learn new words. The goal of the task was to learn eight Italian (L1) word-pairs and eight pairs where Italian words were coupled with pronounceable Russian words deemed nonwords. The participants first heard the pairs once, after which they heard the first item of each pair and had to supply the appropriate word or nonword to go with the stimuli. The task was continued until all eight items (Italian words in the L1-L1 list and nonwords in the L1-nonword list) were remembered correctly in two consecutive trials.

Papagno and Vallar (1995) found that the polyglots were significantly better at nonword repetition and auditory digit span tasks than the non-polyglots. The groups did not differ in learning pairs of Italian (L1) words, but the polyglots were faster at learning nonwords than the non-polyglots. These results together indicated that a close association exists between phonological working memory and L2 learning.

Thorn and Gathercole (1999) compared different learner groups in specific languages and found that language-specific (lexical and sublexical) knowledge has an effect on phonological working memory capacity. In their first experiment they compared three groups of children: native English-French bilinguals, native English-speakers learning French as an L2, and finally English monolinguals. The first group, bilinguals, had been learning French since before the age of three, whereas the L2 learners started French in school at age three or older. The basis for selecting the L2 learners was equal command of French vocabulary compared to the bilingual group, which lead to them being older than the bilingual participants.

Receptive vocabulary in English and French was assessed using the short form of the British Picture Vocabulary Scale and its French translation, where the participants pointed to a picture corresponding to the word spoken by the

experimenter. Phonological short-term memory was assessed using two measures: digit span and nonword repetition, both in English and French. The nonword repetition test merged the CNRep (e.g. Gathercole et al. 1994) and a French version of it, meaning the participants repeated two lists of 40 nonwords, but half of the nonwords in each list sounded like English, and the other half like French.

The results were as had been expected by the researchers. The L2 learners and monolinguals were better at repeating English than French nonwords, whereas the native bilinguals did equally well on nonwords in both languages. The bilinguals and L2 learners did not differ in their performance in repeating French nonwords. All this combined with the participants' success in the vocabulary tasks was considered as indication of the language specificity of phonological short-term memory.

In the second experiment the aim was similar, only this time the participants were native and non-native bilinguals in English and French. Half of the non-native bilinguals had English as L1, half had French, and all had been learning the other language since age two, on average.

The participants' vocabulary knowledge was assessed more thoroughly than in the first experiment. The receptive test was similar to the one in the first experiment: the participants indicated a picture that corresponded to the word spoken by the experimenter. In an expressive test the participants named objects in pictures shown to them. Both of the vocabulary tasks were carried out in English and French. It turned out that the non-native bilinguals repeated nonwords in their L1 more successfully than the ones in the language they started to learn later, whereas the native bilinguals did equally well on nonwords and vocabulary in both languages. These results were considered to support the claim that phonological short-term memory is affected by language-specific knowledge.

A different angle on the possible language-specificity of phonological working memory comes from Masoura and Gathercole (1999). They studied the connection between phonological short-term memory and L2 vocabulary and between L1 and L2 vocabulary knowledge with Greek children learning English as L2 in school.

The language tasks used by Masoura and Gathercole (1999) included a productive and receptive Greek (L1) vocabulary task and two translation tasks to assess English (L2) vocabulary knowledge. Phonological short-term memory was assessed in both L1 and L2 by the CNRep (e.g. Gathercole et al. 1994) and its Greek version.

The results showed that L2 nonwords were repeated less successfully, but also that there was a significant link between phonological short-term memory and vocabulary: both L1 and L2 vocabulary scores correlated with both Greek and English nonword repetition test scores. L1 and L2 vocabulary scores remained significantly connected even after a composite repetition score (English and Greek nonword repetition tests combined) was statistically partialled out, indicating that the connection between vocabulary knowledge in

L1 and L2 was independent of the effect of phonological short-term memory. Furthermore, nonword repetition was linked to L2 vocabulary scores after partialling out L1 vocabulary; however, when L2 vocabulary scores were partialled out, L1 vocabulary was no longer associated with nonword repetition. This was taken as a sign of language specificity, not, however, in the sense that language knowledge in one language is related to nonword repetition in that language, but that a more general nonword repetition capacity is connected to L2 but not L1.

The complex relationship between L1 and L2 was also at the heart of a study by Dufva and Voeten (1999). Using a longitudinal format, they examined whether L1 literacy and phonological working memory predicted L2 learning. The participants were Finnish children who spoke Finnish as their L1 and started to learn English (L2) in the third grade. The participants were first tested in the spring of the first grade and the study continued until the end of the third grade.

Phonological working memory was tested at the end of the second grade with a test modeled after Service (1989). In the test, the participants heard two lists of ten taped nonwords made from real English words. Half of the nonwords had two syllables, half had four. The repetition attempts were scored and rated by syllable, so that a point was awarded for each correctly reproduced syllable.

Three tasks were used to measure knowledge of English at the end of the third grade. In a vocabulary task the participants produced short lists of English words and their Finnish translations on given topics. A communicative task involved reading comprehension and written production: the participants were required to fill gaps in an English text using Finnish clues. In a listening comprehension task an English story was read twice to the participants followed by Finnish questions which the participants also answered in Finnish, in writing.

Dufva and Voeten (1999) found that phonological working memory and some L1 literacy measures predicted L2 knowledge after one year of learning. Based on their results they suspected that both word recognition (one of the literacy measures used) and phonological working memory stem from the same phonological skill, and went on to suggest that phonological working memory may have appeared to have too significant of a role in previous studies since the effect of word recognition had not been considered.

Instead of keeping to languages that the participants are familiar with, nonword tasks with stimuli based on a language that is equally unknown to all participants or on no real language at all have been used in some studies. The latter was done by Speciale, Ellis and Bywater (2004) who studied novice L2 learners by assessing them at the beginning and end of a ten-week Spanish course. Most of the participants were L1-speakers of English.

Phonological sequence learning ability (PSI, recognizing previously heard nonwords) was measured both at the beginning and end of the course as was nonword repetition of nonwordlike (CV sequences in computer

synthesized speech) nonwords. At the end of the course a nonword repetition task with Spanish-sounding stimuli was also administered.

At the beginning of the course, knowledge of Spanish was assessed with a receptive vocabulary task. This involved distinguishing Spanish words from nonwords read on a computer screen. L2 assessment at the end of the course was based on seven written questions in a written exam. The questions were video comprehension (writing down information from Spanish interviews), listening comprehension (four questions about Spanish dialogues), reading comprehension (four content questions on a short Spanish passage), productive and receptive vocabulary (asking and understanding directions on a map), written production (a 50-word passage), receptive vocabulary knowledge (making opposite pairs of a list of Spanish words), and reading comprehension and written production (writing suitable questions to precede Spanish statements presented).

Nonword repetition, PSI, and Spanish receptive vocabulary at the beginning of the course all predicted Spanish nonword repetition at the end of the course. The contributions of these three measures were furthermore independent of each other. The relationship between the nonwordlike nonword repetition task (both beginning and end of course) and Spanish exam performance was not significant overall, but there was a connection with the question involving reading comprehension. The Spanish nonword repetition task, on the other hand, was significantly associated with almost all of the questions on the exam. Repetition of Spanish nonwords was more accurate than that of nonwordlike nonwords, especially with better learners of Spanish.

Speciale et al. (2004) conclude that long-term memory affects repetition of wordlike materials because the more exposure there has been to the language, the better the learner recognizes repeated phonological sequences and is able to abstract their regularities. On the other hand, the capacity of the short-term phonological store places constraints on acquiring new words. The combined effect of the two – the store and the ability to learn phonological regularities – is related to productive and receptive L2 lexical competence more than to store capacity alone.

Farnia and Geva (2011) used a Hebrew-based nonword repetition task in addition to an English-like one in order to eliminate language knowledge from their assessment of phonological short-term memory. The participants were Canadian school children, either L1- or L2-speakers of English. Their receptive vocabulary and phonological working memory were tested once a year during their first six years of school.

The English nonword repetition test was adapted from the Students' Test of Nonword Repetition where the participants repeated 25 nonwords of two to five syllables one by one. The Hebrew-like version was similar in structure. Receptive vocabulary was assessed with the revised Peabody Picture Vocabulary Test.

The L1-speakers scored higher than the L2-speakers on the English-like nonword repetition test in the first grade, whereas in grades two through six

both groups performed equally well. In the Hebrew-like nonword repetition test there were no differences between the groups at any time of testing. Phonological working memory was found to predict vocabulary both concurrently and developmentally as both nonword repetition tests were related to vocabulary. Although the relationships between the two different tests and vocabulary were similar, the contributions made to predicting vocabulary were distinct and the tests were thus assumed to tap separate cognitive systems functioning behind vocabulary development. In light of these results, Farnia and Geva (2011) came to the conclusion that a nonword repetition test based on an unfamiliar language provides a purer and thus a more reliable measure of phonological working memory. They concluded that a nonword repetition test in a typologically distant language predicts vocabulary development in both learners of English and native English-speakers irrespective of the level of English knowledge.

The language-specificity issue has not been studied only in the context of vocabulary. Swanson, Sáez, Gerber and Leafstedt 2004 studied the role of processes in two major memory components (phonological (short-term memory) and executive (working memory)) in children's acquisition of L2 and reading. The participants were American children who either had English or Spanish as their L1 or were Spanish-English bilinguals. The English learning history of the participants was not described in much detail, but most likely even the Spanish monolinguals had started learning English in school, at the latest, since all the reading instruction in school was in English.

Both English and Spanish versions of each measure were administered to all of the participants. The Peabody Picture Vocabulary Test and its Spanish version were used to assess vocabulary. In these tasks the participants saw four pictures and were required to select the one that went with the word spoken by the examiner. Reading was assessed with two tasks: in a word identification task the participants read out loud words that increased in difficulty, and in a word attack task, they read pseudowords.

Working memory was assessed with several tasks. A rhyming task involved listening to nine sets of words, 2–14 words per set, that rhymed, and then recalling the set of words. In a semantic association task, the participants heard a list of words that they had to divide into categories. The difficulty of the task ranged from having to make two categories of two words to making five categories with four words in each. Finally, a visual matrix task was carried out to assess the participants' ability to remember visual sequences within a matrix. During all of the working memory tasks, the participants were asked a question that functioned as a distractor to item recall. In the rhyming task, for example, the participants were asked after each set of words whether a particular word was included in the set.

Phonological measures included the word attack task described above (used to assess reading), letter- and object-naming speed, and digits forward and digits backward where the participants were required to recall sequences of digits that increased in number to as many as eight digits per sequence.

Swanson et al. (2004) found that English word identification and vocabulary were predicted by a language-general working memory factor, and English pseudoword reading was predicted by Spanish pseudoword reading and working memory. Better language proficiency enabled the children to access resources from working memory. Swanson et al. (2004) also attributed L2 difficulties to accessing a language-independent system. As for reading disabilities, the children with such disabilities did poorly on the Spanish short-term memory measures, and hence it appeared that deficits in language-specific phonological memory underlay reading disabilities.

In sum, the language background of participants appears to affect the results on what kind of a connection is found between L2 and phonological working memory. In other words, learners who are different in terms of the number of L1s and the ages at which they have started learning languages appear to have differences in the capacity or use of their phonological working memory. Simply the fact that a person is a polyglot may help in performing well in a nonword repetition task or in learning nonwords faster than nonpolyglots do. This does not even have to depend on exactly which languages the participant knows; however, language knowledge seems to affect nonword repetition success leading to the conclusion that the language of the nonword repetition test does matter. Furthermore, there is an indication that some sort of a more general phonological working memory is connected to L2 more than to L1, once again indicating that the memory components at work behind L1 and L2 learning or knowledge are not the same. There have been attempts at eliminating the effect of language background by using nonwords that are not modeled after any real language or at least not a language familiar to the participants. The results of these studies vary. Either this does away with almost all connections between phonological working memory and L2 (Speciale et al. 2004) or predicts vocabulary learning in different populations (Farnia and Geva 2011).

4.2.3 Influence of Language Proficiency

Similarly to Swanson et al. (2004), Kormos and Sáfár (2008) were not only interested in different memory structures and their connection with L2 acquisition, but also in whether phonological memory had a different role for lower and higher level learners. The participants were teen age Hungarian native speakers participating in an intensive language training program in English (L2). Based on whether they had previously studied English, the participants were divided into beginners and those on a pre-intermediate level.

Hungarian versions of the nonword span test and the backward digit span test were conducted to assess the participants' short-term memory capacity and working memory capacity, respectively. The nonword span test consisted of 36 Hungarian nonwords the length of which varied from one to nine syllables. There were four nonwords of each length and the participants' memory score equaled the number of syllables in the longest nonword when at least two nonwords of the same length were correctly repeated. Nonwords based on the

L1 of the participants were used in order to eliminate the effect of L2 knowledge on the memory score.

L2 English was assessed with the Cambridge First Certificate Exam at the end of the term. The written part of the exam consists of reading comprehension, listening comprehension, composition and a use of English test (vocabulary and grammar), and the oral part includes an interview and tasks of picture description and problem-solving.

Kormos and Sáfár (2008) found no significant connection between the nonword span score and success on the L2 exam for beginners, whereas the nonword score of the pre-intermediate participants correlated moderately with their scores on writing, use of English and total points, and the fluency and range of vocabulary scores of the oral exam. The backward digit span test, however, was found to correlate with all the components of the L2 exam except the writing one. Furthermore, the digit span and nonword scores were not correlated.

Kormos and Sáfár (2008) took these findings as evidence of working memory (assessed by the backward digit span task) and phonological loop (measured by the nonword task) likely being separate constructs and affecting language learning in different ways. They also proposed that working memory affects the acquisition of syntax and vocabulary through attention regulation. What is also interesting is the explicit-implicit division in learning processes that Kormos and Sáfár (2008) bring up as a possible reason behind the differences in the memory-L2 connections between the beginners and pre-intermediate learners. They propose that the lower level learners were exposed to explicit learning processes whereas the learning processes of the higher level learners were more implicit, as there was no longer such a great need for explicit instruction in grammar and vocabulary. This together with previous findings that the ability to repeat nonwords has an effect on implicit vocabulary acquisition renders the results reasonable even though they appear to contradict those of many previous studies where connections to explicit L2 learning have also been found.

The level of language proficiency also turned out to be a factor in a study by O'Brien, Segalowitz, Collentine and Freed (2006). They studied the L2 lexical, narrative and grammatical ability of adult students who were all L1-speakers of English and studied Spanish as an L2. Their oral proficiency and phonological memory were assessed both at the beginning and end of a semester which part of the participants spent abroad in Spain while the rest attended regular formal Spanish classes as part of their university studies in the United States.

As a measure of phonological memory, O'Brien et al. (2006) used a serial nonword recognition task (SNWR) which required no spoken production. The stimuli were English-sounding 1-syllable (CVC) nonwords that were presented to the participants in lists of five, six and seven nonwords, eight lists of each length. Each list was presented to the participant auditorily twice so that on half of the occasions the nonwords in the second presentation of the list were in a different order. The participants were required to indicate whether they

thought the latter presentation was exactly the same as the previous one and the score was the number of correct responses.

To assess the participants' L2 Spanish speech production, an oral proficiency interview of 20 to 30 minutes was conducted at the beginning and end of the semester. The actual data used were two two-minute extracts of each interview. The measures used in the analysis were productive vocabulary, narrative ability, inflectional morphology and clauses. Also, total words were calculated to control for the effect of producing more language at the end of the semester. Productive vocabulary meant the number of unique words in the speech samples, whereas narrative ability was based on five elements considered to indicate narrative discourse. The occurrences of those elements – past tense verbs, third person morphology, past participles, present participles and public words – were summed to form the narrative ability measure. Grammatical abilities were assessed by two means: calculating the number of instances where the participants used any of 14 inflectional morphemes (free and bound) correctly, and counting the subordinated and coordinated clauses in their speech samples. In addition to these more specific measures, O'Brien et al. (2006) assessed the participants' overall knowledge of Spanish at the beginning of the study with the SAT II Spanish test. The participant's success on this test determined whether they were placed in the high or low ability group for further analysis.

O'Brien et al. (2006) found phonological memory to be related to vocabulary use but not to its development. They also found differences between the low and high proficiency groups. For the low proficiency group, phonological memory was related to the development of narrative abilities. For the high proficiency participants only, phonological memory appeared to be related to the gain in both the correct use of function words and in the use of subordinated clauses.

Yet more evidence on the influence of the level of language learners is provided by French (2003). The participants in the study were Canadian francophone children who had received formal ESL instruction in school in grades four and five. In the sixth grade they participated in an intensive English program (English immersion) where they spent five months of their school year learning English 25 hours a week. The data were collected at the very beginning of the program (first week of month 1) and at the end (last week of month 5).

Phonological working memory was assessed with two nonword repetition tests both in the beginning and end of the intensive program. First, the Children's Test of Nonword Repetition (see e.g. Gathercole et al. 1994) was adapted to a North American pronunciation and the participants were required to repeat 40 English-sounding nonwords. Second, there was a similar nonword task with Arabic nonwords, which was scored by native Arabic speakers. This task was included to control for changes in phonological working memory as L2 knowledge of English increased. While such an increase might have affected the ability to repeat English nonwords, it should not have made a difference to the participants' performance on Arabic nonwords.

Three types of measures were used to assess the participants' knowledge of English as L2. Overall L2 proficiency was assessed both before and after the intensive program by adapted versions of two standardized tests, the Baldwin-Cartier English Test (BCT) and the Ministry of Quebec English Test (MEQ). There was also a vocabulary translation task requiring the students to translate 60 items from French to English and another 60 items from English to French. The test was essentially a paper-and-pencil task with the participants writing down their translations of a list of words, but the participants also heard taped spoken versions of each word to be translated. Finally, grammatical knowledge was assessed with a cloze passage with ten gaps.

The participants were divided into low and high proficiency groups based on their performance in the pre-program overall L2 proficiency test results. The Arabic and English nonword repetition test correlated highly at both testing times, and a significant difference was observed between the low and high proficiency groups on the accuracy of the Arabic nonword repetition test at both times, with the high proficiency group performing better on the nonword test. In both groups, performance on the English nonword repetition test improved over time, possibly because of increased familiarity with the L2. The phonological working memory scores also correlated significantly with overall L2 proficiency and vocabulary in both groups on both testing occasions.

The results indicated a close relationship between phonological memory and L2 proficiency, and also L2 subskills that are based on vocabulary. It also appeared that there may be a causal relationship between phonological memory and later L2 proficiency that is not mediated by previous L2 knowledge. The relationship between phonological working memory and grammar ability, however, appeared to be mediated by vocabulary.

Phonological working memory predicted L2 proficiency both in low and high proficiency learners. However, it only predicted overall L2 learning in the low proficiency group, lending further support to the assumption that phonological working memory plays a smaller role in learning as L2 proficiency increases. Because it appears that previous knowledge of language structure affects new vocabulary learning, French (2003) suggests that the relationship between phonological memory and long-term language knowledge is reciprocal.

Again findings on when and how working memory affects L2 vary somewhat. Kormos and Sáfár (2008) found no connections between phonological working memory and L2 in beginning learners, whereas already in pre-intermediate learners connections were found to several L2 measures. It should be noted, however, that Kormos and Sáfár (2008) used L1-nonwords and as discussed above, the language of the task appears to make a difference. O'Brien et al. (2006) found connections in both low and high ability groups, but the aspects of L2 knowledge involved differed. Finally, French (2003) found phonological working memory to predict L2 learning in low but not high proficiency learners. In sum, the findings are somewhat contradictory but definitely indicate that the level of language ability has a bearing on what kinds

of connections there appear to be between phonological working memory and L2, and that an overall connection in a group of participants – whether one exists or not – does not necessarily tell the whole story of the influence of phonological working memory on L2.

4.2.4 External Influences

The possible effects of language background were already discussed above in section 4.2.2, but the effect of instruction or the learning environment on people with similar language histories has also been of interest to researchers.

Using the same data as in O'Brien et al. (2006) above, O'Brien, Segalowitz, Freed and Collentine (2007) set out to investigate the role of phonological memory in the acquisition of L2 oral fluency in adult students with English as their L1 and Spanish as an L2.

What was used to assess speech production were two two-minute extracts of an oral proficiency interview both at the beginning and end of a semester. Oral fluency was conceived as general overall oral ability (measured by the total number of words and the number of words in the longest turn) and fluidity (measures by speech rate, absence of hesitations, absence of filled pauses and the longest fluent run in words containing no silent or filled pauses).

When looking at all the participants, phonological memory as assessed by SNWR (for details, see O'Brien et al. 2006 above) appeared to have a minor role in the growth of oral fluency, but when the learning context (university course or study abroad) was statistically partialled out, SNWR accounted for a significant amount of variance of all but one oral development variable, absence of hesitation. The development in oral fluency as measured by the total number of words and absence of filled pauses was significantly linked to SNWR performance only in the participants who studied abroad. Based on the results, O'Brien et al. (2007) deemed phonological memory valuable for adult L2 oral proficiency development, regardless of the learning environment.

Dufva and Vauras' (2002) interests lay in two subskills of phonological processing: phonological memory and phonological awareness. They were specifically interested in whether an intervention program they had developed enabled the intervention group to outperform the control group in skills tapping proficiency in the English sound structure. The participants were Finnish children who had started to study English as L2 in the previous school year.

Phonological memory was assessed with a task consisting of two lists of two- and four-syllable pseudowords that sounded like English. In the word recognition task the participants read a text of 12–14 sentences aloud as quickly and correctly as they could. Finally, there was a dictated spelling test. The tasks were administered four times. There was a pre-test in the fall of the fourth grade, and an intermediate test 4 months later, a post-test another four months later, and a delayed test in the fifth grade, nine months after the end of the intervention program.

The participants were divided into intervention and control groups matched for nonverbal intelligence. For the intervention program the participants in the intervention group were divided into three groups of four. Each group met once a week for 45 minutes a total of 26 times. The control group only attended regular English classes. The purpose of the intervention program was to advance the participants phonological processing skills in English by focusing on their comprehension of English sound structure. Dufva and Vauras' (2002) assumption was that better phonological processing skills would be reflected in improved phonological awareness, phonological memory, word recognition, and spelling, the four measures of English they used.

When the development of the aspects of English proficiency (phonological memory, word recognition and spelling) between the pre-test and delayed test was compared between the intervention and control groups, the differences were statistically significant. Intervention seemed to have a positive effect on phonological memory as well as word recognition skills and spelling skills. Dufva and Vauras (2002) conclude that focusing on sound structure, phonological awareness and explicitly teaching grapheme-phoneme correspondences, as they did in the intervention program, develops English language proficiency and the ability to form correct representations of pseudowords.

Based on just these two studies, it seems fair to say that relatively little is known thus far about how it might be possible to influence phonological working memory or its connection to L2. The learning environment does not appear to play a great role, at least for adult learners, but a specific focus on sound structure may help.

4.2.5 Review of Previous Research on Phonological Working Memory and the L2

As obvious from the above, the relatively small number of studies focusing on the role of phonological working memory in L2 contexts show great variety in approach, which may make it difficult to form a comprehensive picture of all the factors that seem to affect the findings. Above, in an attempt to put in a nutshell what is currently known about the issue at hand, previous studies were discussed in light of their main findings. In the present section, a closer look is taken at two key constructs used in the studies, the L2 tasks and the phonological working memory measures. This should help in gaining a perspective on the sometimes conflicting results, as the phonological working memory measures used help to reveal what the researchers understand the memory structure to mean, while the L2 tasks reveal the underlying view of language learning or language knowledge, i.e. what is it exactly that phonological working memory is thought to be connected to.

4.2.5.1 L2 Tasks

It might be helpful to discuss the L2 tasks selected in the previous studies in categories of the language skills assessed. Various aspects of language knowledge have been studied in the phonological working memory context, and many studies have not looked at only a single skill; however, one target of study, vocabulary, stands above all others in popularity. Therefore, we start with vocabulary tasks, then move on to reading and other tasks through which language skills have been assessed. The details of the measures used in each study are not reproduced here. For this information, see sections 4.2.1–4.2.4.

By far the most widely used tasks to assess vocabulary knowledge in an L2 were translation (Service and Kohonen 1995; Cheung 1996; Masoura and Gathercole 1999; French 2003; Speciale et al. 2004) and learning word pairs, which although essentially the same task, is usually connected to an experimental setting, rather than used to measure pre-existing knowledge of an L2 (Papagno et al. 1991; Baddeley 1993; Papagno and Vallar 1995; Atkins and Baddeley 1998). Another very similar task was that used by Dufva and Voeten (1999). They had the participants produce lists of words and their translations on given topics. Other task types included the British Picture Vocabulary Scale (Thorn and Gathercole 1999), which measures receptive vocabulary, as did a lexical decision task where the words given were either made up or real ones in the L2 (Speciale et al. 2004). Finally, another vocabulary assessment method was a list of words, which the participants were asked to arrange into pairs of opposites (Speciale et al. 2004).

Among the studies interested in L2 reading, there was also a clear favorite: reading comprehension, in which the participants answered questions based on a text they read in the L2 (Service 1989; Service and Kohonen 1995; Speciale et al. 2004; Kormos and Sáfár 2008). Tasks where lists of words or nonwords were read were also used. These included word identification (Dufva and Vauras 2002; Swanson et al. 2004) and word attack (Swanson et al. 2004), and which indicate a view of reading as decoding.

Writing was most often assessed by having the participants write a composition or a passage of some sort (Service 1989; Service and Kohonen 1995; Speciale et al. 2004; Kormos and Sáfár 2008). One study also called for reproduction of structures, i.e. writing down what was dictated (Service and Kohonen 1995).

Listening comprehension seems to have been taken into consideration in several studies (Service 1989; Service and Kohonen 1995; Dufva and Voeten 1999; French 2003; Speciale et al. 2004; Kormos and Sáfár 2008), whereas in a few studies specific grammatical issues have been mentioned as assessed. Service and Kohonen (1995) mentioned testing for knowledge of phrases, recognition of adjective forms, correct word forms and past tense forms. French (2003) assessed grammatical knowledge with cloze tasks and Dufva and Vauras (2002) were interested in spelling. Even school grades have been of interest (Service 1989).

Oral production was of interest to Kormos and Sáfár (2008), whose participants took part in an oral exam that consisted of an interview, picture description and problem solving, not, however, described in any detail. Also in O'Brien et al. (2006) and O'Brien et al. (2007) an oral proficiency interview was used. O'Brien et al. (2006) used this method to assess lexical, narrative and grammatical skills, whereas O'Brien et al. (2007) used the same interviews to look into oral fluency.

Some of the studies discussed here have grouped several L2 tasks into batteries of tests under headings such as communicative or traditional tasks (Service and Kohonen 1995; Dufva and Voeten 1999) or overall proficiency (French 2003). Here, these batteries have been split into the different task categories listed in this section.

Vocabulary learning and knowledge were of interest in most of the studies discussed, as can be seen in the plethora of assessment measures used. The other areas of interest show the same variation on a smaller scale. However, while listings of the measures as presented above show that the measures were rather varied, of more interest here is whether the L2 tasks used somehow reflect the types of participants or whether the studies concern the participants' knowledge of an L2 already being studied or report an experiment where a previously unknown L2 is taught and then assessed. In fact, a clear majority of the studies focused on learners who came from the same monolingual background, i.e. had the same L1 and were learning the same L2 (Service 1989; Service and Kohonen 1995; Cheung 1996; Dufva and Voeten 1999; Masoura and Gathercole 1999; Dufva and Vauras 2002; French 2003; O'Brien et al. 2006; O'Brien et al. 2007; Kormos and Sáfár 2008). In the learning experiment studies – all except for Cheung (1996) – the participants did not study the L2 tested, or at least had no pre-participation experience of it. Because the participants had no prior knowledge of the L2 used in the study, their L2 background was evidently not of consequence, and thus not revealed. However, it could be argued that any previous L2 learning experience will have some effect on learning another L2, and might therefore have been worth at least a mention. Bearing in mind these general considerations, the more detailed comparisons below start with the most common language skill assessed, vocabulary.

Unsurprisingly, since very many different kinds of vocabulary tasks were used in these studies, it turned out that many of them were only featured in a single study. This could prompt the conclusion that certain tasks are used to assess knowledge and others to assess learning; however, but that is not necessarily the case. Of course, some of the tasks are more demanding and thus better suited to assessing knowledge: a learning experiment would have to be a fairly lengthy one if the participants were to be tested using, for example, a task involving the production of lists of words and their translations on different topics (Dufva and Voeten 1999).

Two vocabulary tasks were found that were very similar to each other and were used in several studies: translation and learning word pairs. Translation was used with both regular L2 learners and in learning experiments. Word

pairs were only used in studies involving learning experiments, possibly because it was the main method of teaching the participants something in a new language. In other words, the same task was first used for learning trials and then as a test of how well the pairs had been learnt. It is important to note that the words learnt in the word pair tasks did not have to be real language (Papagno et al. 1991). This makes very clear the fact that the ability to learn was of key importance, whereas what was learnt was not that relevant. The focus was on studying learning.

All of the experimental learning studies focused on vocabulary, probably because single words are perhaps the easiest thing to try and make somebody learn during an unavoidably short experiment. The second experiment by Speciale et al. (2004) could be seen as lying between an experiment and regular learning since the participants – novice learners of Spanish – were assessed at the very beginning and end of a ten-week Spanish course. Thus the participants were not taught by the researchers themselves, but it was nevertheless possible to observe and assess learning. This, of course, also enabled the use of other measures besides vocabulary. The same applies to French (2003).

With reading tasks, almost all of them were used to assess L2 *knowledge*. Only the above-mentioned study by Speciale et al. (2004) used a reading comprehension task to assess *learning*. Otherwise, the studies to do with reading were a very homogenous group, both with respect to the participants (L2 learners) and the goal of assessing knowledge. Clearly, reading is assessed after learners have some experience with the L2, although some task types such as word attack, where nonwords are read, could easily work in experimental settings as well (Swanson et al. 2004).

Except for writing and listening comprehension in Speciale et al. (2004) and the L2 tasks in French (2003), the tasks for the remaining language skills were all conducted in studies with regular L2 learners and used to assess knowledge of the L2s in question in each study, and therefore offer little to speculate about. In sum, then, vocabulary tasks appeared to attract the most varied kind of research interest, whereas the rest of the tasks were mostly used to assess the participants' knowledge in the language they had already been learning as an L2 prior to taking part in the studies in question. Overall, the issue of variety is present once again. It is good to see things from different perspectives, but with so few studies, it is not easy to determine whether the variety adds to or detracts from the mounting evidence about possible connections between phonological working memory and L2.

Some of the experimental studies appear very far from a typical or natural L2 learning situation, for example, when a nonsense language is being taught in a learning experiment. This obviously has its advantages: for example, there is seemingly no need to control for previous knowledge, since the language in question is not real and, therefore, knowing it is impossible. However, there may be aspects of it that resemble a language known to some of the participants and determining that would be a laborious task should the researchers decide to attempt it. At least some researchers seem to be aware of the artificiality of

nonsense word learning; for example, Papagno et al. (1991) comment on their CVC nonsense syllables possibly being unlike an actual foreign language enough to make the task of learning very different from those in their other experiments with a real foreign language.

By taking a closer look at how the researchers included here have justified the language measures used in their studies, their choice of a particular area of language knowledge or even their efforts to understand language or learning, interesting observations can be made.

Overall, the vocabulary studies seemed explicitly to consider the nature of language and learning less than the other studies. They often limited the theoretical background mostly to studies on the connections between phonological working memory and L2 vocabulary. This is, of course, understandable considering their focus and the fact that, in studies on phonological working memory and L2, vocabulary has dominated quite strongly. However, this gives the impression that the influence of phonological working memory takes place in near isolation, no other factors – linguistic, cognitive or those to do with the learning context – having any input. It is, of course, clear that an experiment has to focus on testing a particular hypothesis and it is thus limited in many ways: only a few things can be done and choices have to be made; nevertheless, the rationale behind those choices is often not made clear. From an SLA perspective, it would be valuable to know what beliefs about language and language learning have influenced the researchers' choice of methods.

There were two different approaches to vocabulary, namely learning and knowledge, which also meant different needs and possibilities regarding the L2 measures to be employed. The studies on L2 vocabulary learning almost exclusively used fairly simplistic learning trials that have very little to do with most natural language learning situations. To their credit, the researchers themselves did not claim otherwise. As they are interested in learning, it is understandable that a short learning trial is perhaps the quickest and most easily controlled way to do it, especially if the rate of learning (number of trials it takes to learn the vocabulary items presented) is of interest. Nevertheless, the researchers do not generally justify their basic choice of measures. They do not make evident their thinking behind the decision to teach paired associates, what advantages this might have over some other methods, et cetera. While they offer some reasons for using particular languages they do not justify the number of words to be learnt, the choice between visual or auditory presentation or between oral or written responses, raising (and leaving unanswered) a host of questions about their beliefs regarding the nature of language and learning.

Most insights about language seem to come during or after the experimental stage of the study. For example, Papagno et al. (1991) suspected that the different results between participants with different L1s might have to do with the differing association value of the L2 Russian words to them and with related systematic differences in the learning strategies applied. Aspects

that most, but not all, researchers pay attention to are the length of the words to be learnt (certain number of syllables), that they are perhaps concrete words (e.g. Papagno and Vallar 1995) and their frequency (in the case of L1 words). The foreign or nonsense vocabulary to be learnt is mentioned as pronounceable and, as the focus is on the learning *process*, the scores may reflect a slight disregard for the end product, the learnt word or nonword. For example, Speciale et al. (2004) mention making allowances in scoring their German (previously unknown L2) oral vocabulary learning task, as not all the phonemes exist in the participants' L1, English. This prompts the question about what the goal of the learning task is and, as learning is under scrutiny, what the participants are supposed to be learning, what is or is not relevant to their learning, and if they are aware of this while participating in the task.

Some researchers (e.g. Baddeley 1993) are explicit about their interest in what happens when participants are trying to learn novel vocabulary, such as when they have to resort to rote learning and how they try to benefit from visual imagery or semantic encoding. Also, an interest in a particular function of phonological working memory, for example, the effect of articulatory suppression during a vocabulary learning task in Papagno et al. (1991), makes it more apparent that a very controlled and easily replicable L2 measure has been the only option for obtaining differences in learning conditions.

It should be clear from what has been said above that in the studies on the connection between phonological working memory and L2 vocabulary learning so far, there is little happening that comes close to naturally occurring, socially situated language learning processes. The issue is, of course, very complex and only so many questions can be answered in a single study. However, there are already some studies that at least in part have made an attempt to simulate a natural language learning experience in their experiments. Atkins and Baddeley (1998), for example, used short sentences in their language tasks and encouraged the use of semantic encoding strategies and distributed learning in order to assist the participants to focus on the difficult items more during the learning stage of the trial. Then there are, of course, pioneers such as Service (1989), Speciale et al. (2004) and French (2003), who studied learning without conducting experiments as such and thus had to forgo control of the learning process.

The few studies looking at vocabulary *knowledge*, although no more explicit in their justifications for their language tasks, are, ironically, more so about their views of the learning process. Masoura and Gathercole (1999), for example, speculate about the teaching methods used in Greece following their results showing close links between the L1 and L2 vocabularies, which indicate that other factors are at play in vocabulary learning besides phonological memory. Masoura and Gathercole (1999: 387) cite previous research as they look for alternative possibilities and hypothesize that, at least, in the early stages of learning vocabulary in the formal context of a language class, it appears to be the norm to associate new L2 words with their equivalents in an already known language, usually the L1. Later, when proficiency in the L2 is at

a level where, for example, context is understood well enough for it to help in understanding new vocabulary items, the link to the L1 word is not necessary and the two words are not so closely associated in the learners' mental lexicon.

The researchers looking at several different areas of language knowledge and their relation to phonological working memory appear to be more aware of having to discuss their reasons for choosing their foci of interest, but do not always justify the measures used to assess L2 knowledge. O'Brien et al. (2006), for example, are very clear about their reasons for studying adult L2 oral production, but why they deemed it essential to focus on lexical, narrative and grammatical skills and used oral proficiency interviews as data is not explained. O'Brien et al. (2007), on the other hand, used the same data to look at fluency and go into considerable detail in describing fluency as a phenomenon and their measures of it.

Kormos and Sáfár (2008) used a ready-made test, the Cambridge First Certificate Exam, without giving any particular reason for their choice. French (2003), on the other hand, serves as a positive example. He used adapted versions of standardized tests, a very versatile battery of measures, and provides a whole section of justifications for them. He also explains the tasks in some detail. Indeed, standardized tests are a valuable tool for research, but there are always reasons for choosing a particular battery and these, in turn, could usefully be discussed and made explicit.

Dufva and Vauras (2002) had a logical backdrop to their tasks, as their studies were tied to the Finnish context of a communicative approach to L2 learning. While the understanding of communication in their tasks may not be shared by all, they describe their tasks thoroughly and justify them as measuring the key aspects of the beginning stages of L2 learning in the communicative approach.

In general, in these more varied studies, the researchers have clearly thought about and make explicit their views on language and learning. O'Brien et al. (2007) stress the influence that the learning context may have on L2 development while O'Brien et al. (2006) acknowledge that the processes in L2 oral production development may be quite different depending on the initial level (high/low) of the learners. Service and Kohonen (1995) explicitly state their understanding that, in the L2, several skill components are intertwined but that perhaps at some stages of L2 learning one component may stand out in importance to learning. Finally, French (2003) reports considering what learners actually do and how they are assessed and how that may be different from (and reflected in) what they were asked to do for his study.

Overall, then, it seems that the further away from mere vocabulary the researchers venture, the more explicit they are about their choices and understanding of the field. Being strictly interested in something like the rate of learning of three new vocabulary items, and only considering this in the working memory context is certainly acceptable and necessary for gaining knowledge about the learning process. However, more explicit ideas about language learning more broadly may lend credence to a study in the eyes of

those looking for information on why some of the students in their L2 class and outside the classroom learn more or faster than others. This is not to say that strictly theoretical studies are less valuable than those with more ties to the life of the L2 learner as we know it. It is more a question of who they are useful for and when. Sometimes ecological validity has to be sacrificed for a detailed study of a cognitive process which has the potential to lead to greater rewards in the end.

The following section sheds light on the other target of assessment common to all the studies in this chapter, phonological working memory and the tests used to measure it.

4.2.5.2 Phonological Working Memory Tasks

The choice of phonological working memory measure naturally depends on what is the core interest of the study. It is rarely just any kind of phonological working memory that researchers are interested in. Such aspects as active or passive phonological working memory or the alleged different parts of working memory or executive versus more specific working memory have been of interest in the studies included here. Due to this, the phonological working memory measures used in the studies are various as well. Moreover, many several different measures of phonological working memory have been used in a single study.

The present study was interested, in particular, in nonword repetition tests and previous studies using this task, so that, in the studies reviewed here, the single most frequently used measure of phonological working memory was, naturally, the nonword repetition test. More than half of the studies reviewed in this chapter used some version of it, and hence it definitely merits a closer look. Next, we review examples of how the nonwords used have been formed, which languages have been imitated, how the tests have been conducted, and how participants' productions have been scored.

Nonword repetition tests. What is perhaps most critical for the success of a nonword repetition test in accessing the target qualities is the choice of nonwords used in the test. As discussed in section 3.1.3, a number of features of nonwords have been shown to affect study findings. First of all, the nonwords chosen should conform to the phonotactics of a language known to the participants, so that unfamiliar prosodic patterns or articulatory difficulties do not cause repetition problems (Gathercole et al. 1994: 106). As nonwords are used in order to eliminate the influence of long-term memory as much as possible, low wordlikeness is also important (Gathercole and Baddeley 1993a: 48). Finally also the length of the nonwords is important, especially with respect to variety and sufficient length, so that differences between participants can be observed (e.g. Gathercole and Baddeley 1993a: 58).

Since the quality of nonwords is of the essence in the results gained with nonword repetition tests, we start by looking at how the studies reviewed here have ended up with their particular nonwords. In some cases, ready-made

nonwords from the Children's Test of Nonword repetition (CNRep, see e.g. Gathercole et al. 1994) have been used. French (2003) used CNRep in the L2 English nonword repetition test, and a second task with Arabic nonwords followed the same pattern. However, the source of the Arabic nonwords is not discussed. For all practical purposes, they may have been real Arabic, since the participants did not know or study that language. Thorn and Gathercole (1999) also used CNRep with English nonwords to assess their participants' phonological working memory, but in their case it was the L1, or one of two L1s in the case of French-English bilinguals. Their French nonword test (L2 or other L1) was constructed on the same principals as CNRep with regard to wordlikeness, for example; however, the exact details of how they went about creating the nonwords were not given. Similarly Kormos and Sáfár (2008) used L1 Hungarian nonwords from a previous study, but mention only that they were phonotactically legal.

Two studies, Baddeley (1993) and Papagno and Vallar (1995), gave no indication whatsoever as to how they came to use their particular nonwords, a point we address below, when we look at some other features of the nonwords and nonword repetition tests used. However, most researchers were fairly thorough in describing their nonwords. Speciale et al. (2004) did not go in for strictly phonotactically legal nonwords, but instead used randomly combined CV sequences in their nonwordlike nonword tests in experiments one and two, while the rest of the tests in the studies used real words as a basis for nonwords. In Service (1989), the Finnish (L1) nonwords were old Finnish words no longer in use, whereas the English (L2) nonwords were formed by interchanging syllables in existing English words. Service and Kohonen (1995) used the same English task. Dufva and Voeten (1999) and Dufva and Vauras (2002) similarly used the same material, and also formed their English (L2) nonwords by interchanging the first and last syllable of real English words. Speciale et al. (2004) conducted a Spanish (L2) nonword repetition test in their second experiment, where the nonwords were created by changing the letters of Spanish words at random.

As we have seen above in section 3.1.3, the wordlikeness of nonwords is a factor that can have a great influence on what is actually assessed in nonword repetition tasks, and thus what can be assumed on the basis of the results (e.g. Gathercole, Willis, Emslie and Baddeley 1991). It is therefore not irrelevant whether the nonwords used are much like real words in the language that is being imitated, provided that is a language which the participants are familiar with. The wordlikeness issue was brought up in most and also touched on in some earlier studies conducted before the importance of the issue was even really recognized. Baddeley (1993) mentions that the nonwords did not resemble English, which can be assumed to be Baddeley's research participants' L1. Service (1989) and Service and Kohonen (1995) said the nonwords conformed to Finnish and English phonotactics. Similarly Dufva and Voeten (1999) and Dufva and Vauras (2002) commented that their nonwords respected English phonology, but wordlikeness was not commented on.

Speciale et al. (2004) aimed at conducting nonwordlike nonword repetition tests in their two experiments, and described the nonwords used as not being wordlike. This is easy to accept, as they had produced the CV sequences randomly. For their Spanish nonword repetition test, Speciale et al. (2004) chose nonwords that rated highest in wordlikeness by Spanish native speaker raters. The researchers expected this to ensure that the participants' long-term knowledge of "the regularities of Spanish" was involved in the task.

Others have also assessed wordlikeness. Masoura and Gathercole's (1999) English (L2) nonwords were rated rather low and Greek (L1) nonwords were rated approximately in between high and low wordlikeness. Both were assessed by Greek native speakers. French (2003) used English nonwords in CNRep, but assessed their wordlikeness and found it relatively low. Thorn and Gathercole (1999) did not assess wordlikeness but used the same stimuli from CNRep.

In most studies, the stimuli, i.e. the nonwords to be repeated in the nonword repetition test, were taped to ensure that the same exact versions were presented to the participants. Baddeley (1993) and Papagno and Vallar (1999) did not comment on this, although the latter referred to earphones as the medium of presentation of the nonwords. The same two studies did not comment either on how the stimuli were produced. Masoura and Gathercole (1999) also did not comment on this issue, but as they followed the CNRep procedure, it is likely that a native speaker of the language spoke the nonword stimuli.

Speciale et al. (2004) used computer synthesized speech to present their nonwordlike nonwords, but for the Spanish nonwords a Spanish native speaker was used. Native speakers of the languages of the nonword repetition tests were also used in the remaining studies: a Finnish and an English native male in Service (1989) and Service and Kohonen (1995), an English male with a standard British English accent in Dufva and Voeten (1999) and Dufva and Vauras (2002), a female French-English bilingual in Thorn and Gathercole (1999), and finally an English native female and a native Arabic speaker in French (2003). The use of a native speaker with a neutral accent is familiar from the CNRep (e.g. Gathercole et al. 1994). While this may not be an absolute requirement, it is easy to see how a strong foreign accent could detract from the clarity of the task. As claims about phonological memory are made on the basis of nonwords in a certain L2, it is assumed that the nonwords chosen also reflect that L2 in pronunciation, appropriate word stress, et cetera.

The nonwords in all of the nonword repetition tests reviewed here were presented to the participants one by one, and the repetition attempt followed immediately after the nonword had been heard. In most of the studies this is stated clearly, in the remaining studies it can be safely assumed, for example, by looking at some other aspects of the studies such as scoring (Service 1989; Service and Kohonen 1995; Papagno and Vallar 1995; Baddeley 1993).

In terms of the exact numerical details of the nonword repetition tests used, the studies fall fairly neatly into three different categories. First, the

largest number of tests followed the CNRep paradigm. These nonword repetition tests had a total of 40 nonwords, the length of which varied from two to five syllables, with ten nonwords of each syllable length. The nonwords were presented one by one in ascending order of length. This format was used in Thorn and Gathercole (1999), Masoura and Gathercole (1999) (L2 English only), and French (2003). The Greek (L1) nonword repetition test in Masoura and Gathercole (1999) was similar, except that there were altogether 50 nonwords, as the test included six-syllable nonwords. Speciale et al. (2004) used a very similar format as well, only the numbers differed. In the nonwordlike nonword repetition tests a total of 32 nonwords were presented in ascending order ranging from one to eight syllables. The Spanish version also contained 32 nonwords, but these varied in length from two to eight syllables. Speciale et al. (2004) also scored the tests differently from CNRep. They gave a point for each correctly repeated nonword and correct syllable, whereas those following the CNRep procedure more closely only gave points for correctly reproduced nonwords.

The second category followed the model initially used by Service (1989). Two lists of nonwords were repeated, one with ten Finnish (L1) nonwords and one with ten English (L2) nonwords. Half of the nonwords in each language consisted of two syllables, half of four, and the order they were presented in was randomized. This version was used in Service (1989) and Service and Kohonen (1995). Dufva and Voeten (1999) and Dufva and Vauras (2002) adapted their nonword tests from Service (1989), but only used nonwords in English (L2). Points in all four of these studies were rewarded for correctly repeated syllables.

The third category includes the so-called adult versions of the nonword repetition test. Baddeley (1993) used three nonwords of each length, starting with one syllable and going on until the participant failed to correctly produce all the nonwords of one and the same length. Papagno and Vallar's (1995) version also consisted of three nonwords of each length, but varying from two to nine syllables. The nonwords were presented in ascending order of length and the test continued until three consecutive errors were made, even if they were not all in the same syllable-length category. The maximum number of nonwords to be repeated was 24. The score was the number of syllables in the longest correctly repeated nonword. Baddeley (1993) employed two ways of scoring. He too used the number of syllables in the longest correct nonword, but also the number of syllables correctly repeated for all three nonwords of the same length.

What could be problematic in the scoring methods in the adult versions is the use of syllable length as a score, as this means that the differences between scores are likely to be very small. The reason for this is that the range between the number of syllables in the shortest and longest nonwords is lower than the range between zero and the maximum number of nonwords. In a typical CNRep situation the score could vary from 0 to 40 points, whereas in Papagno and Vallar (1995) the maximum was nine and in Baddeley (1993) the maximum

was theoretically unlimited but turned out to be six. While this type of score is more informative about the maximum length it is possible for the participants to produce, the low range could complicate possible statistical analyses, and might render the scores fairly useless for many types of analyses, unless the number of participants is very big.

Going back to the whole range of studies with nonword repetition tests, what counted as an error was not specified in all of them, but usually the participants were allowed to have an accent, but could not cross phoneme boundaries without being penalized. Omitting sounds or rearranging or changing them in any way was also not allowed (e.g. French 2003). Often several raters were used and inter-rater reliabilities calculated (e.g. Service 1989), since rating nonword repetition tests requires hearing every single sound and judging whether they fall within an acceptable range from the original. Scoring was usually based on taped productions, especially with adult participants, where the capacity requirements for the scorer can be very high, making taping essential.

Other phonological working memory tasks with nonwords. Nonwords have not been used only in repetition tasks but also in other measures of phonological working memory. Cheung (1996) conducted an L2 English nonword span test in which the participants repeated sequences of 1–7 nonwords, each nonword having two syllables. There were two trials at each sequence length and the task continued until the participant made errors in both of the two sequences of a given length. The score was the length of the longest sequence (maximum seven) where at least one sequence was produced correctly. In O'Brien et al. (2006) and O'Brien et al. (2007), lists of L1 English nonwords were presented auditorily and only recognition was required. However, this time the participants did not pick out individual nonwords but had to indicate whether another sequence of nonwords corresponded to the original one. On this protocol, the demands the task sets for working memory were different, but no production was required.

While the nonwords in these studies may not have been repeated in the traditional fashion, the tasks or their description are not dramatically different from those in the more typical nonword repetition tests. The researchers do not go into much detail about their nonwords. The wordlikeness of the nonwords in Cheung (1996) was assessed as lying in between wordlike and not wordlike, but what counted as an error is not really discussed. The scoring in O'Brien et al. (2006) or O'Brien et al. (2007) was not problematic because participants either recognized the nonwords or nonword sequences heard or they did not.

To draw an analogy with the previous section on L2 tasks, the phonological working memory tasks used do not seem to behave in any particular pattern in relation to whether the researchers' interest has been in language learning or knowledge, or whether the participants were already learners of a particular L2 or taking part in a learning experiment. However, whether conscious or not, or whether explicit or not, different tasks inevitably

assess and reveal different things about phonological working memory. For example, whether only single nonwords or longer sequences of them are repeated could reflect the storage and rehearsal parts of phonological working memory to a varying degree. There are also findings (e.g. Majerus, Poncelet, Greffe and Van der Linden 2006) supporting the hypothesis that item and order memory tasks are independently related to vocabulary development. The capacity to retain order information or to retain item information thus appear to stem from separate cognitive processes. Additionally, in some studies a conscious attempt was made to prevent other factors from affecting the memory score for example through using a nonword recognition task instead of the participants having to produce speech (e.g. O'Brien et al. 2006; O'Brien et al. 2007). It seems safe to say, then, that in addition to choosing to use nonwords working memory researchers have to make a number of other decisions based on what part of phonological working memory they want to study with their nonword task and what else the task may require of the participants, possibly leading to other factors besides phonological working memory affecting the nonword task score.

Other phonological working memory tasks. A great number of the studies reviewed here used phonological working memory measures which did not include nonwords in any format. Digits were repeated in an auditory digit span test (Papagno and Vallar 1995; Baddeley 1993) and in digits forward and digits backward tests (e.g. Swanson et al. 2004; Kormos and Sáfár 2008). The verbal span tasks in Atkins and Baddeley (1998) included phonologically similar and dissimilar letter spans, both auditorily and visually presented, in addition to digits. Words had the main role in a rhyming task (Swanson et al. 2004) and in a vocabulary task done under articulatory suppression (Papagno et al. 1991). The details of the tasks are not repeated here, but can be found in sections 4.2.1–4.2.4, where each study is described in more detail. Naturally, there are numerous other working memory measures available in the literature in addition to those listed here; however, only the measures used in the L2-studies reviewed in the present chapter have been included in the summary.

To sum up, the range of phonological working memory measures is vast, but one or two things stand out. Nonwords and nonword repetition tests have been used more often than any other measure of phonological working memory. It has also become apparent that sometimes the line between phonological working memory and language measures is blurred: for example, phonological memory is one of Dufva and Vauras' (2002) English proficiency measures.

The choice of a phonological working memory measure could, and ideally should, reflect the researchers' understanding of phonological working memory. A particular phonological working memory measure may accord better with a certain view of phonological working memory or if the aim is to focus on a particular part of phonological working memory, for example only storage or both storage and processing, as in Kormos and Sáfár (2008).

Also, interest in a particular learner group may influence the choice. Among the studied reviewed here, those focusing on adult learners appear more heterogeneous compared to those with children. In the studies with child participants, the majority conducted some version of the nonword repetition test which most likely would not have been suitable for older participants. Nonword span tests were more popular with adult participants, but there was great variation in their assessment of phonological working memory.

The language of the phonological working memory task: language specificity. In the nonword tasks, the most commonly used target language was, perhaps not surprisingly, English. In Thorn and Gathercole (1999), it was included to get an impression of how well the participants could reproduce nonwords in their L1. Also in O'Brien et al. (2006) and O'Brien et al. (2007), L1 English nonwords were used specifically to prevent L2 knowledge from affecting the results. In Service (1989), Service and Kohonen (1995), Masoura and Gathercole (1999), Dufva and Voeten (1999) and Dufva and Vauras (2002) it was the L2 of the participants. Other languages imitated included Finnish (L1) (Service 1989; Service and Kohonen 1995), Greek (L1) (Masoura and Gathercole 1999), French (L1 or L2) (Thorn and Gathercole 1999), Spanish (L2) (Speciale et al. 2004), Hungarian (L1) (Kormos and Sáfár 2008) and Arabic as a control, not a language the participants knew or were learning (French 2003).

Baddeley (1993) and Papagno and Vallar (1995) did not mention any particular language that the nonwords might have been based on. They did say that they were pronounceable, but Baddeley (1993) stated that his did not resemble English, which was allegedly the language of his participants. Speciale et al. (2004) used nonwordlike CV sequences, in which case there was no language that they were meant to imitate. The language of the nonwords is not inconsequential, as at least the learning of phonologically familiar (based on L1 English) pseudowords has been found easier than that of phonologically unfamiliar ones (based on Finnish, a language unknown to the participants) (Service and Craik 1993).

Among the studies reviewed here, Masoura and Gathercole (1999) and Thorn and Gathercole (1999) studied the language specificity of phonological working memory most explicitly. In some other studies, the intentions were not as clear but through the choice of particular methods it was indicated that the researchers took a certain stand on the issue, e.g. O'Brien et al. (2006), O'Brien et al. (2007), and Kormos and Sáfár (2008), who all used L1 nonwords. Whether it was merely a question of wanting to obtain an assessment of phonological working memory that is as free from the interference of long-term knowledge as possible, or whether there was an assumption that some general phonological working memory could be tapped in this way as opposed to some language-specific phonological working memory is not immediately obvious from these studies.

Intuitively, using L1 in phonological memory tasks to prevent L2 knowledge from having an effect on the memory score and from rendering

comparisons of phonological working memory and L2 circular seems to make sense. If, as acknowledged early on, long-term knowledge of the sound structures of a language may affect the result of a nonword task, then surely the sensible thing is to use nonwords that level the playing field and do not depend on knowledge of the language that the memory scores are supposed to be compared with. The problem with using the L1 is, however, that then the memory score may be a reflection of the participants' L1 knowledge. If the logic is that participants are all so proficient in L1 that they cannot differ on a nonword repetition test on those grounds, it all becomes a bit more complex. After a certain age and within a very homogenous group of participants this may be true enough, but an all-embracing fact it is not. Furthermore, it can easily lead to limited research results. At least in the case of Kormos and Sáfár (2008), using L1 nonwords caused the participants (already age 15-16) to score fairly high on the task, and thus no participants with poor phonological working memory were found. Service (1989) had a similar but even more unfortunate experience, being unable to use the L1 nonword task scores at all, as all the participants scored so highly on them.

One solution may be to use a completely unknown language as the basis of nonwords, as in French (2003) and Farnia and Geva (2011). This choice is also a reflection of a particular understanding of phonological working memory and its relation to L2 knowledge. With an unknown language as the basis, it is obviously some general phonological working memory that the results are supposed to be about. However, it is likely that there will be sound combinations more or less familiar to the participants even then, especially if the L1s or L2s vary, and any of the languages within the participants' long-term knowledge may have varying effects on the results.

Without a doubt, phonological memory seems to respond differently to different language test stimuli, but whether this is because of knowledge of that language and to what extent remains unclear. The most suitable ways to get around the problem are equally unclear. If an unknown language is used, what are we actually measuring? Aspects of the unknown language could still resemble a known one, and that knowledge would end up having an effect regardless of any efforts to the contrary. In the present study scores on both L1 and L2 nonword repetition tests – but not in an unknown language – were used. This issue will be addressed and further speculated below.

Bearing these doubts or reservations in mind, we now summarize the previous studies on phonological working memory in the context of L2.

4.2.5.3 Summary and Conclusion

The present chapter has introduced and described studies focusing on phonological working memory in the context of L2 knowledge and learning from several points of view. What we know about the connection between phonological working memory and L2 knowledge and learning can be found in these studies, at least to some extent.

To sum up the results relating to phonological working memory and L2, there seems to be plenty of evidence for a connection between L2 vocabulary and phonological working memory, even for the one predicting the other. Vocabulary has even been credited the role of being behind other connections between phonological working memory and L2 knowledge. However, results to the contrary have also been reported. In some studies phonological working memory has not been found to be connected separately to any particular language skill. It also seems that phonological working memory affects learners of different levels in different ways, and it appears to function differently whether L1 or L2 is in question. Having better language knowledge or being a polyglot seems to lead to better phonological working memory, as does more exposure to language. All in all, there is a clear indication that phonological working memory is a factor in L2 learning, and while most of the evidence supports a connection to vocabulary learning and knowledge, phonological working memory can be said to be an important part of commanding other aspects of L2 as well.

The varying methods and results of the previous studies are somewhat baffling and definitely indicative of the fact that (phonological) working memory research is still very much work in progress and comprises a plethora of issues that are far from clear or agreed upon by the research community. The numerous different foci raise contradictory thoughts. Although there are relatively few studies on the connection between phonological working memory and L2, it is nevertheless interesting to learn about the issue from so many perspectives.

The present study is an attempt to provide some new insights on the connection between phonological working memory and L2 by focusing on Finnish children learning English at school. Hence, the present study follows the majority of the previous studies. The participants are so-called normal learners, children with one L1 learning one and the same L2 in school. However, many of the previous studies are really about studying phonological working memory, i.e. how it works and is constructed. The interest here is more on seeing whether inspecting the connection between phonological working memory and L2 can be helpful in identifying patterns that can be used in aiding L2 learning. Because the present study was part of a bigger longitudinal study, it was possible to include measures that were not specifically planned as part of the present study. This was done by including sixth-grade Finnish (L1) reading comprehension task results and fifth-grade data. In this way the present study is one of the more comprehensive studies on phonological working memory and L2, not a mere tilt at nonword repetition and vocabulary.

Having said that, there was also a definite hope that something new might possibly be discovered about phonological working memory. The aim was to look into the possible connections between phonological working memory and several L2 skills, as many studies have done before, only this time using less traditional language tasks. Vocabulary was naturally one of the L2 areas assessed, since it has been studied so much already. It was of interest whether

the connections would be found with the different kinds of vocabulary tasks used here. Otherwise the rationale behind some of the L2 tasks used was methodological rather than a specific interest in a particular language skill. The tasks used were fairly complex, nonword-repetition-like listening and repeating tasks. In these more experimental L2 tasks, many skills and many kinds of language knowledge are needed and are probably impossible to tell apart. Admittedly, the new kinds of tasks add to the variety of L2 tasks, the very phenomenon that the previous studies were criticized for above.

The present study also aimed to add to the evidence on the relationship between phonological working memory and different languages, the languages here being Finnish and English. English is probably the most widely studied language in this connection, whereas Finnish is represented in far fewer studies. In particular, there was a need for a Finnish nonword repetition test. Mainly in Service (1989) and in Service and Kohonen (1995) Finnish nonwords were used but the results could not be taken for analysis because of a ceiling effect. The suspicion was that it was not because of the Finnish language or Finnish learners but that the nonwords used were not suitable, at least not for the age group in question.

Before moving on to Chapter 5 and discussing the research questions, the following section shows the thinking behind the choice of employing some less traditional English tasks for the data collection.

4.2.6 Measuring L2 Knowledge

In addition to a fairly typical vocabulary task, the present study used two more unorthodox measures of L2 knowledge to gain as versatile data as possible as ecologically as possible. The philosophy of these two task types is recounted here but a more detailed description of the exact methods applied in the present study is provided later in Chapter 5.5.

4.2.6.1 Sentence Repetition

The sentence repetition tasks used in the present study were based on a task type used to gain information on learners' implicit language knowledge: elicited imitation. As this is a slightly less traditional measure, it is described here in some detail together with studies that have used it as an L2 measure. The actual written and spoken sentence repetition tasks used in the present study are described in Chapter 5.5.2.2.

While free production such as story telling might be an ideal measure of language knowledge, it could, according to Erlam (2006: 466–467), be considered impractical because of the challenges it presents for rating. Especially if there is an interest in a particular structure, there are no guarantees that the structure will appear in the participant's free production. For such purposes elicited imitation is a more practical method.

Elicited imitation is a measure used to assess language knowledge or proficiency in connection with both L1 and L2 learning and in

neuropsychological research. It involves the participants reproducing sentences presented to them by the experimenter (see e.g. Gass and Mackey 2007; Vinther 2002; Service and Kohonen 1995). The sentences are heard and repeated one at a time, usually orally but sometimes in written format (e.g. Scheibner-Herzig, Sauerbrey and Kokoschka 1991), although the justifications for this choice appear to be lacking. The assumption is that a participant who understands the meaning of a sentence is capable of reproducing it correctly. If either the semantic meaning or the syntax is not understood, correct reproduction will not be possible. Depending on the framework of the study, the sentences produced can be scored based on correct linguistic form or meaning.

Elicited imitation is thought to tap several aspects of language knowledge and use. It has been argued that correct repetition may give information about language proficiency because it requires the participant to process the heard model sentence correctly and to encode it by using grammar known to the participant (Vinther 2002). It has also been suggested that elicited imitation places a strain on listening comprehension and memory in addition to reproduction (Scheibner-Herzig et al. 1991). Participants have to identify the pronunciation of words, encode the sentences they hear into semantic units, and use their chunking capacity for words, phrases, and propositions, for example. Since elicited imitation is about assimilating the stimulus sentences in the participant's own internal grammar, repetition errors reflect linguistic deficiencies, errors due to grammar, not chance (Munnich, Flynn and Martohardjono 1994).

Some questions have been raised as to the validity of the elicited imitation method as a measure of L2 knowledge. First of all, it could be assumed that subjects in an elicited imitation task do not really understand but merely imitate the heard stimuli (see e.g. Erlam 2006). In such a case, information would be gained on their perceptual-motor skills, not their structural ability (Vinther 2002). It has also been speculated whether elicited imitation measures memory rather than language ability. It appears that both these issues can become problems; however, they can be controlled for by sufficient stimulus sentence length (see e.g. Gass and Mackey 2007). While a simple, short sentence might easily be repeated on the basis of mere imitation or memory, even by those subjects who do not completely understand it, a sufficiently long sentence is enough to force participants to actually understand and reconstruct the sentence upon repetition. Evidence for this reconstruction is discussed next.

It has been argued that elicited imitation cannot be only about rote repetition, since, in e.g. Munnich et al. (1994), ungrammatical stimulus sentences were converted into grammatical ones, revealing that the process is reconstructive and not mere rote repetition. This type of correction can also be viewed as an indicator of elicited imitation accessing subconscious grammatical knowledge instead of conscious prescribed rules, as is the case with some other measures. Further support for the reconstructive nature of elicited imitation is provided by Sachs (1967), who found that at least recognition memory of sentence form has been found to decline considerably more rapidly than

memory for the meaning of the stimulus sentence (Sachs 1967). Based on their sentence recall experiments, Potter and Lombardi (1990) also claim that sentence repetition is reconstructive and conceptually based, not mere repetition of the surface representation. The starting point for the repetition is a representation of the meaning of the stimulus sentence.

In sum, elicited imitation appears to involve many cognitive activities (listening, decoding, recalling, producing linguistic elements, et cetera), but it is currently not quite clear which of these activities is most relevant (Vinther 2002). Most of the elicited imitation work has been done in L1 contexts and with the Universal Grammar framework. Usually, in such cases, only one grammatical aspect is being looked at at a time, so that the stimulus sentences only vary in relation to the specific grammar construct under investigation (see e.g. Lust, Flynn and Foley 1996, Bley-Vroman and Chaudron 1994). As the focus of the present study is on children's L2 knowledge, three studies that have used an elicited imitation task in assessing learners' L2 knowledge will be summarized next.

Eisenstein, Bailey and Madden (1982) were interested in comparing elicited imitation and cued production as measures of L2 performance with adult learners. In particular, they looked for signs of progress in the present simple and progressive form in Wh-questions. The participants were 45 intermediate learners with varying language backgrounds, and were all taking part in an intensive ESL program. The elicited imitation task used in the study consisted of 16 English sentences, all approximately 15 syllables in length. Assessment was based on the tested structures – verb forms – being correct or incorrect. In the production task the participants had to produce questions based on a series of pictures shown to them. The exact scoring of the tasks is not explained in the article but significant similarities between the two types of tasks were found. However, when the tasks were analyzed separately for different levels of learners, great differences were found between them, the lower-level learners obtaining varying results in the production task. Eisenstein et al. (1982) suggest that while a production task is a fairly conservative measure of language knowledge, elicited imitation can show something of the hypotheses that the learner is testing. Imitation also forces the participant to use certain structures that the experimenter is interested in, and can thus reveal more about the developmental state of the learner.

Scheibner-Herzig et al. (1991) conducted a sentence repetition test with 37 German ninth-graders who were learning English as a foreign language for the fifth year. The researchers sought to find out how long the sentences could be for these participants to still be able to produce acceptable sentence repetitions, how their achievement could be evaluated with a repetition measure, and what dimensions of achievement affect the repetition results. The repetition test comprised 12 pairs of English sentences, all statements and varying from 3 to 14 words in length. Each pair contained a simple and a complex sentence of the same length. Only grammatical sentences were used. The sentences were read onto a tape by a native speaker of English and only presented to the

participants once. They were then asked to write the sentences down and allowed as long a time as needed. Scheibner-Herzig et al. 1991 consider global assessment of dictated sentences, especially long ones, important, and consequently the written sentences were scored on a scale from 1 to 5 points separately for grammatical accuracy and semantics. The final score (maximum 120 points) was the mean of the two scoring methods,. Based on the results it appeared that complex and longer sentences were more difficult to repeat, although the difference was not statistically significant. It was concluded that repetition taps chunking capacity, requires correct identification of word pronunciation, and forces the learner to encode the sentences in "meaningful and manageable units" (Scheibner-Herzig et al. 1991: 236). The results were also compared to those of a native English-speaking control group. The comparisons showed that the natives' performance on complex sentences did not differ from simple sentences as much as it did for the German participants.

Munnich et al. (1994) compared elicited imitation and grammaticality judgment tasks to find out what it was that the two tasks measured and how they related to each other. They were also interested in how performance factors might influence the results for the tasks. Two versions of both tasks were compared. In both versions of the elicited imitation task, the subject heard one sentence at a time and had 10 seconds in which to repeat it. In one version, the stimulus sentences were administered orally, and in the other, the subject heard them from a taped recording. Each sentence was only heard once. The two versions of the grammaticality judgment test were a timed read test and a timed taped test. The subject read or heard the stimulus sentences and recorded whether the sentence was grammatical or not. Fifteen seconds were allowed for every sentence and they were only read or heard once. There were 12 sentences in each of the four tasks, all complex sentences with relative clauses. Of the 48 sentences half were grammatical, half ungrammatical. All the sentences were 15 syllables long and had approximately the same number of words. The subjects were 12 Japanese adults who were advanced learners of English and had been trained in the vocabulary of the stimulus sentences before taking the tests. The scoring of the tasks is not described, but overall, the subjects performed better on the grammaticality judgment tasks than on elicited imitation. There was no difference found between taped and live stimuli, i.e. the two versions of the elicited imitation task, but the repetition of ungrammatical sentences was found to be more difficult than repeating grammatically correct sentences. Munnich et al. (1994) concluded that elicited imitation is not just rote repetition, since it works equally reliably with ungrammatical sentences: the subjects converted the ungrammatical stimuli sentences into grammatical ones. They concluded that elicited imitation is a good, sensitive measure of knowledge of constraints on grammar, but they also called for more cross-task comparisons to test the validity of elicited imitation as an L2 task.

In summary, studies on L2 using elicited imitation tasks are few and deal with quite varying topics. All the studies summarized here have been about learning English as L2, but other topics have also been studied. The focus has

been on particular aspects of grammar or then there has been a general interest in what such repetitions can be used to measure.

Elicited imitation has often been compared to other tasks, so the validity of the task itself has been one of the aims of research. It is safe to say that there is no one common research interest here, much less a common test design. The numbers of sentences vary as do assessments of the reply sentences produced. Quite a few of the elicited imitation studies use sentences which are approximately 15 syllables in length in order to exceed the short-term memory span, but there is a lot of variation in sentence length across studies (Vinther 2002). As for assessment, most studies seem to use a scale of categories where the two poles represent completely accurate and completely flawed reproduction (Vinther 2002). The scales seem to vary from study to study and the points or stages between the extremes may be somewhat difficult to tell apart exactly. Furthermore, in two of the three studies summarized above, it was not clear on what grounds the rating of the tasks was done. This appears somewhat ironic in light of Erlam's (2006) comment that an elicited imitation task might be superior to some other methods, especially in rating.

4.2.6.2 Story Retelling

A story task was included in the present study to elicit language beyond sentence length and to emulate free production but in a more limited manner as regards the vocabulary and structures used. A well-known task that comes close to the story task used here is the dictogloss task, which is more a tool for learning than assessment. It is also usually performed in collaboration with others, which obviously was not a possibility given the purposes of the present study. In a typical dictogloss task, the participants aim to reconstruct a text that has been dictated to them so that both the content and the linguistic form are as accurate as possible. Usually there are two readings and a possibility to take notes. For more details on the dictogloss task see Wajnryb (1990).

Besides the dictogloss-type format, other methods exist that aim to elicit longer, somewhat controlled language production. One widely used method is to have the participants narrate a story based on picture cues, as in the Frog Story (see e.g. Berman and Slobin 1994). However, the dictogloss worked better for the purposes of the present study as most of the other data collection methods used also required repetition. The repetition tasks guaranteed that all participants got the same language stimuli and were basically expected to produce the same exact stretch of language. This was the main criterion for selecting the task type. Since it was to be used for assessing language knowledge, it was practicable to have a more controlled task. This way all participants had the same goal, while it was clear what the perfect response was and what would give the maximum score on the task. In free production, or even elicitation based on a series of images, there are very few limits, and assigning scores becomes much more complicated. Moreover, many other factors besides language knowledge would have affected what and how much

the participants produced, and assessing meaning, at least, would have been impossible, in practice.

As the story was obviously much longer than the individual sentences reproduced in the sentence repetition tasks, it might be thought that the task was too demanding for the participants' memory capacity. However, the story was fairly short and simple, with not that many events for the participants to try and remember. As a whole, the story was a logically proceeding description of an event which in itself would have aided recall. Remembering one part would most likely have reminded the participant about what led to it and what came after. Because the story had to be reconstructed in writing, there was more time to plan the text and ways of expressing the meanings in the story. In spoken reproduction, in a one-on-one situation, there would have been much more pressure to recount a logical, grammatical story in the L2. As mentioned above, the spoken version would have also been too time-consuming to be conducted within the present data collection.

The research questions that were answered in part with the types of tasks described here are discussed next.

5 EMPIRICAL STUDY

5.1 Research Questions

While there is not as much research on the association between phonological working memory and L2 as there is on the corresponding relationship within L1, the pool of L2 studies is nevertheless significant. The need for further study does not arise from the number of studies as much as from their varied nature. Although connections between phonological working memory and L2 knowledge or learning have been found, the types of L2 and phonological working memory measures have had a great influence on the results, thus leaving open the question of what it is exactly in L2 that phonological working memory is connected to and what difference the choice of the language of the nonword repetition test makes to the connections found. These unsolved issues are at the root of the two general research questions addressed in the present study. The two main research questions are

> Is there a connection between children's knowledge of an L2 and their success in repeating L2 nonwords?
> Is phonological working memory language specific?

The first general research question is not new in itself. There is evidence of such a connection: success in one co-occurs with success in the other (e.g. Service 1989; Dufva and Voeten 1999; Service and Kohonen 1995; French 2003). However, from an SLA perspective it is interesting how language is assessed, i.e. what represents L2 knowledge in such studies. The present study differs somewhat in the approach taken to assessing L2 knowledge. The question at issue is, if a person has high or low phonological working memory capacity, is this reflected in their L2 ability when it is assessed separately for different levels of language using tasks requiring written or spoken production and focusing separately on language structures and meaning?

As discussed above, the methods used in previous studies are varied and sometimes also are the findings. While the nonword repetition tests used

appear to have remained fairly constant, there are great differences in language assessment. Hence, the findings on the relationship between phonological working memory and L2 depend greatly on the kinds of language tasks used. The goal in the present study is to answer in more detail the question about what aspects of L2 we are actually talking about when a connection is or is not found between L2 and phonological working memory. This is attempted by a systematic study of three different aspects of L2 tasks: level, response modality, and scoring. L2 nonword repetition results are compared to measures of L2 knowledge on three different levels: vocabulary, sentence and story. Two different modalities, spoken and written, are studied separately as are two assessment perspectives, scored on the basis of language structure or meaning. L2 nonword repetition was used to predict different combinations of these aspects to find out if any of them makes a difference in terms of there appearing to be a connection between nonword repetition and L2 knowledge.

Unlike previous studies, the present study seeks systematically to compare, first, different level data (vocabulary–sentence–story) from the same participants. Tasks on each of these levels separately have been used before; in particular, and there are a great number of studies on vocabulary. In fact, sometimes the finding has been that vocabulary is the key to all the other connections between phonological working memory and L2 (Service and Kohonen 1995).

Second, a comparison of spoken and written versions of similar tasks is made in the present study. Since nonword repetition tests are spoken, it is conceivable that the relationship between nonword repetition and a language task could vary due to modality, i.e. whether the language task is also spoken or whether it requires writing. The processing demands of writing and speaking tasks are different, and this could be reflected in dissimilar connections with nonword repetition. Mostly written tasks, but also some spoken ones, have been used before, but the two modalities certainly have not been compared before.

Third, the structure and meaning scores obtained for the same tasks are compared. While there are innumerable ways to use different tasks to assess language proficiency, there are also different approaches to scoring tasks. In a case where language data are elicited, there are decisions to be made as to the scoring of the data. In previous studies the same data have not been scored separately for structure and meaning. When the understanding and reproduction of sentences and a story are required, both detailed grammatical structures and the gist of the sentences/story is important. From a language assessment point of view, assessing form and function separately seems a more thorough analysis of the participants' language proficiency. While they are necessarily somewhat interconnected, there is the possibility that someone who is not very precise in their use of structures is nevertheless quite capable of conveying the meaning of the target sentence/story. It is possible that the connection between phonological working memory and L2, is, once again, seen differently when the focus is on structural details versus conveying the meaning.

Based on the previous research, it is hypothesized that English nonword repetition and L2 tasks are connected despite task type or assessment perspective. Most studies have found a connection, and while many of the aspects considered here have not been systematically studied before, there is no evidence to suggest they would not be connected. If this turns out to be the case, then the connection between phonological working memory and L2 knowledge may be assumed to be fairly strong, as it will be visible in tasks the demands of which go well beyond those of a typical written vocabulary task. In other words, the effect of phonological working memory on L2 knowledge applies to many kinds of language use situations, both written and spoken, and from word level to discourse level. If there are differences in the connections between nonword repetition and L2 according to the level, modality or focus on structure or meaning, then it will be possible to speculate on which aspects of language knowledge phonological working memory actually has an effect.

In addition to the manipulation of L2 assessment, possible changes with time in the connections between phonological working memory and L2 knowledge are investigated. While the main focus of the present study is on the participants' L2 and nonword repetition test performance in the sixth grade, the longitudinal design of the research project offers a chance to compare the participants' performance across different school years, providing an opportunity to track possible changes.

While the nonword repetition task was very similar in both grades, L2 assessment in grade five was different from that in grade six. The data from the two grades is compared to see whether there are major differences in connections. Research shows that phonological working memory develops as children get older (see e.g. Gathercole and Baddeley 1993a: 25–31; Gathercole, Willis, Emslie and Baddeley 1991: 365) and, in a similar vein, the obvious goal of L2 instruction is that learners achieve a higher level of performance as they proceed in their language studies over the years. Based on the data in the present study, it is difficult to say anything specific about the development of the phonological working memory of the research participants since the nonword repetition tests used in different years are not necessarily comparable. However, what can be inspected are the changes within the group. Based on correlations between tests from different years, conclusions can be drawn about whether the participants who succeeded well in the tests in earlier years remain better than the others, and vice versa, exhibiting a similar level of development across the group. If there are major changes in the connections between the two grades, this is likely to be due to L2, either to changes in knowledge or because of the difference in assessing it. If there are noticeable changes in English nonword repetition so that the two tests are not correlated, the situation is more interesting. Major developments in phonological working memory as such are no longer expected at this age, and therefore such a result would perhaps be accounted for by the effect of changes in L2 and the L2 knowledge on English nonword repetition.

The second general research question the present study aims to answer is whether phonological working memory is language specific. Based on previous studies, the language of the nonword repetition task appears to make a difference as to what kind of a connection is seen between phonological working memory and L2. For example, Thorn and Gathercole (1999) suggest that success in a nonword repetition test may be dependent on knowledge of the language which the nonwords are mimicking. However, not many studies have addressed this question directly, while the methods used as well as findings are somewhat varied. It is therefore not quite clear how language knowledge affects success in nonword repetition and if there is a way reach some kind of general phonological working memory by using L1-based nonwords as compared to those based on an L2 (e.g. Kormos and Sáfár 2008).

The present study includes both L1 and L2 nonword repetition tests to allow for comparisons of possible connections and to reveal possible effects of the source language of the nonwords in the phonological working memory test. For the same reason, measures of L1 Finnish reading comprehension are also included. Both L1- and L2-based nonword repetition tests are compared to L2 and L1 tasks to see if either or both are connected to success in the language tasks. If language knowledge affects nonword repetition, then there should be clear connections between the nonword repetition and language tasks relating to the same language, i.e. L1–L1 and L2–L2, and not such clear connections between nonword repetition in L2 and L1 language tasks or vice versa. The existence of possible connections is not assumed to depend on the type of language task (level or modality) as no such connection is indicated by previous research.

It may not be possible to say whether phonological working memory itself functions differently when different languages are in question or whether possible differences seen are a function of knowledge in those particular languages. Using nonword repetition and language measures in two languages that all the participants know to varying degrees may help shed light on the issue, depending on what kinds of connections are found. No previous studies on the successful testing of typically developing school-age speakers of Finnish as L1 with a Finnish nonword repetition test have been reported that would have been available or suitable for the present participants, and consequently there was clearly a need to consider the possible problems involved and attempt to create a functional test. Furthermore, only Thorn and Gathercole (1999) have successfully conducted a nonword repetition test in L1 and L2 for the same participants, and they only studied vocabulary knowledge. The present study thus attempts to extend this line of research.

As the present study is a part of a longitudinal research project, there is naturally an interest in the profiles of the individual research participants in terms of language tasks and nonword repetition tests. How the participants perform across all tasks could also shed light on the language-specificity issue, for example, whether there are profiles where the participant only excels at tasks specific to only one of the two languages, English or Finnish. It is

hypothesized that there are very different kinds of profiles as the two different languages do not have to be connected to each other. However, under the language-specificity assumption not all combinations are assumed possible. On the assumption that a nonword repetition test and language tasks in specific languages are connected, a participant will not be expected to perform well, for example, an L2 nonword repetition test and L1 language tasks and poorly on L1 nonword repetition and L2 language tasks.

If both English and Finnish nonword repetition tests are connected to both English and Finnish language tasks, then it can be assumed that they both tap a general phonological working memory. There are findings that the level/ability of language learners is a factor in whether language knowledge affects nonword repetition, which might also be seen in differences between tasks. If some of the L2 tasks are connected to English nonword repetition only and others also to Finnish nonword repetition, then conclusions could be drawn as to which kinds of tasks are language-specific, which are dependent on a more general phonological working memory factor. Finding that one or both of the nonword repetition tests tap general phonological working memory would be a step closer to a better understanding of the relationship between the phonological working memory system and long-term knowledge. In particular, if only one of the tests shows such connections, conclusions could be drawn as to what it is in the test that makes it a more general measure of phonological working memory as compared to the other one.

Answers to the research questions were sought by administering a series of tasks to the participants of the longitudinal research project. The participants were Finnish sixth-grade children in their fourth year of learning English as their L2. All the sixth-grade memory and language measures were taken within a 6-week period in the spring term, and the participants were 12 or 13 years old at the time. Tasks from the fifth grade were used in some of the analyses, as incorporating those results provided a longer time span to reflect on the relationship between phonological working memory and L2 knowledge; however, the main focus was on the sixth-grade data.

Both an English and a Finnish nonword repetition test were presented to obtain a broad assessment of the participants' phonological working memory and to investigate the claims that the language of the nonword repetition test may play a more important role in the results than originally suggested by the advocates of the task. In addition, a previous study (Service 1989) was unsuccessful in creating a Finnish nonword repetition task for school-aged children with sufficient variability to allow research on individual differences. The present study attempted to produce and apply a functional instrument using Finnish nonwords. The nonword repetition tasks are described in section 5.4.

Knowledge of English as L2 was assessed with vocabulary tasks, sentence tasks and a story task, all of which involved the participant hearing the stimuli and having to respond either orally or in writing. The Finnish language tasks used in answering some of the research questions were reading comprehension tasks based on two factual and two fictional texts. The description of the

English and Finnish language tasks are provided below in sections 5.5 and 5.6, respectively.

The participants as well as the tasks used, including their design and assessment, are described in more detail in sections 5.3–5.6 below.

5.2 Data Collection

A number of nonword repetition tests and tasks assessing language knowledge were carried out to answer the research questions discussed above. The fifth-grade tasks were administered in the autumn term, and the sixth-grade tasks another year and a half later in the spring term.

All the tasks in the fifth and sixth grade took place within a relatively short period of time to make sure that they assessed the properties of interest at approximately the same time. This was important since the aim was to compare, for example, repetition of English nonwords to L2 knowledge at a given point in time. At the rate the participants were studying English, a few weeks most likely did not bring about a very great difference in their language knowledge, but six months might have done, and that was to be avoided. All fifth-grade tasks were completed within a few months and the sixth-grade tasks within about a month. All the tasks used in data collection are listed in Table 3.

TABLE 3 Timing of data collection.

	5th grade (N = 16)	*6th grade (N = 15)*
Working Memory Tasks	English nonword repetition test 1	Finnish nonword repetition test English nonword repetition test 2
L2 Tasks	From an exam: Listening comprehension Reading comprehension Three verb form tasks (gaps) Sentence formation 1 Others: Sentence formation 2 Vocabulary (30+30)[1]	Vocabulary (spoken and written) Sentences (spoken and written) Story (written)
L1 Tasks		Reading comprehension[2]: 2 fictional texts 2 factual texts

1 Vocabulary task data collected and evaluated by the present author, other data collected by other members of the research team and evaluated by the present author.

2 Collected and evaluated by other members of the research team.

5.3 Participants

The participants were Finnish school children whose L1 was Finnish. All the participants started studying English as their first L2 in the third grade. They were eight or nine years old at the time.

Due to the longitudinal nature of the research project, the number of participants has changed over time. In the fifth grade when the first English nonword repetition test was presented, there were 16 participants left. They were divided into three different groups, each including other children, for English instruction. All the groups were taught by the same teacher.

In the sixth grade, the 15 remaining participants took part in all of the memory and language tasks. The participants were part of a bigger sixth-grade group which was then divided into two smaller groups for English instruction. Hence, the participants were divided into two separate English groups, each of which included several other students who were not participants in the present study. The two groups were taught at different times but had the same amount of instruction and shared the same English teacher.

5.4 Instruments and Procedure: Nonword Repetition Tests

The core of the data for the present study comes from grade six when two nonword repetition tests carried out: one Finnish and one English nonword repetition test. In addition, data from a previous, fifth-grade English nonword repetition test are reported (Miettinen 2003). A detailed description of the design and procedure of each nonword repetition test is provided in the three sections below. Over the course of the longitudinal study the participants had also taken part in two previous Finnish nonword repetition tests in the second and the fifth grade. The data from those tests are not reported in the present study, but a description is available in Miettinen (2003).

5.4.1 The Finnish Nonword Repetition Test

The Finnish nonword repetition test was administered in the sixth grade, in March, one year and five months after the corresponding fifth-grade task.

Design of the Test. The Finnish nonword repetition test was closely based on a similar test constructed in the previous school year (Miettinen 2003). Here the creation of the original test is described together with notes on the changes made for the sixth-grade version. Originally, a decision was made to base the test on the one conducted by Baddeley (1993). A very basic form of the L1 test was assumed to be too easy for the participants, and hence a more demanding test was needed in order for it to yield analyzable results. Baddeley's (1993) test

comprises three nonwords of each syllable length and, starting with the shortest ones, a participant tries to repeat them in the order of length until all three nonwords of a given length are reproduced incorrectly, at which point testing is stopped. Even if the participant is able to repeat fairly long nonwords, the test as a whole does not grow too long with only three nonwords representing each length.

In the fifth-grade test the nonwords did not go beyond eight syllables and this was not changed for the sixth-grade test. It is possible that some participants could have coped with even longer nonwords, but in practice, as a nonword, an eight-syllable one is already so difficult that there was no real fear that many of the participants would manage to repeat three of them correctly. As the shortest nonwords had two syllables and there were three of each length up to eight syllables the total number of Finnish nonwords was 21.

As no suitable set of Finnish nonwords was available, they were created with the help of guidelines for possible structures of words of different syllable lengths (Lapsen kielen kehitys ja dysleksia -projekti / The Jyväskylä Longitudinal Study of Dyslexia (JLD)). The goal was to construct nonwords that sounded like natural, phonotactically legal Finnish words. This entailed, for example, following the principles of Finnish vowel harmony, using typical consonants and their possible clusters while simultaneously aiming for enough variation in the numbers and types of diphthongs, et cetera. Special attention was paid to endings of words so that no tell-tale signs of any particular word class would be present to aid in the repetition. While technically, the aim was to have the nonwords resemble nominative nouns, the most important thing was to prevent familiar suffixes from being included. Such known endings would have shortened the new (or non-) part of the nonwords and made them easier to repeat.

To make the nonwords of different lengths, a set of nonwords of two, three and four syllables was created and then combined in various ways to obtain longer ones. Originally this was done at random and objectively, but in the end some selection was made based on the wordlikeness (as judged by the present author) of the created nonwords. In particular, it was difficult to make the longer nonwords sound natural in Finnish. After the experience of the fifth-grade test, some nonwords were revised for the sixth-grade version to make them even more wordlike. The final list used in the test is given in Appendix 1. As in Baddeley (1993), no practice items were included in the Finnish nonword repetition test.

A Finnish female native speaker read the nonwords on a minidisk several times and the most natural versions were then recorded on a CD. The stimuli were separated by an interval of six seconds to allow for repetition. The instructions, read by a Finnish native speaker male, were also recorded and inserted at the beginning of the test CD. In addition to this, instructions were also given live at the beginning of each individual test session and the participants asked if they understood what they were supposed to do. In the instructions, the participants were told that they were about to hear Finnish-

sounding words and to attempt to repeat them even though they had never heard them before. It was stressed that they should try to repeat the words even if they were not quite sure about what they had heard. In the sixth grade the participants were also reminded that they had taken part in a similar task the previous year. Before being administered to the children, the test was piloted with an adult native speaker female not familiar with the format. This was done to make sure the instructions were clear enough and that the interval between the nonwords was long enough for the repetition.

Procedure. In Baddeley's (1993) version the test was terminated once the participant missed all three nonwords of a specific nonword length. This was not done in the present study (or the fifth-grade version); instead, all the participants had the opportunity to hear and attempt to repeat all 21 Finnish nonwords. First of all, terminating the test would have required online scoring, and very carefully at that, as it would have led to ending the test. Furthermore, while completing the whole test enabled it to be scored in a traditional manner, one point for each correctly reproduced nonword regardless of how many items were missed in between these, it also allowed the stricter Baddeley (1993) scoring, i.e. the number of syllables in the last correctly repeated nonword before the length at which all were incorrect. In addition, it was known that a similar test would be conducted with L2 English nonwords and that the stricter scoring was not a possibility in that test. It was possible that participants might fail to reproduce all the nonwords of a certain length early on in the test but nevertheless manage to repeat some longer ones correctly. Terminating the test early would have led to very low scores and not enough variation between participants. As the aim was to obtain similar, comparable tests and scores on phonological working memory in both L1 and L2, going through the whole test in both languages was deemed the best option.

The participants were tested individually in an empty room where they heard the recorded instructions and the stimulus nonwords through headphones. The responses were spoken into a microphone and recorded on a minidisk and later downloaded onto a computer for scoring. All participants were tested within a week of each other.

The Baddeley score was found unsuitable in the fifth grade (see Miettinen 2003), and consequently was not even considered in the sixth-grade nonword repetition test. Therefore, the maximum score in the test was 21 as each nonword was worth one point. Even if only one phoneme was omitted, misplaced or replaced by another, the score for that nonword was zero. The criterion for a correct phoneme was that it should be close enough to the original so that if in a real word, it would not differentiate meaning. For example, Finnish has various s-sounds, but they are not used to differentiate meaning. Therefore, regardless of which s-sound the participant produced in place of the [s] in the original nonword, it would always be correct. Points were also not deducted if possible problems were due to a consistent deviation like an inability to produce the sound [r] in a conventional manner. Furthermore,

self-correction was accepted and the participants were given the benefit of the doubt and scored as repeating correctly if it was not absolutely clear that they omitted or mispronounced something.

5.4.2 The Fifth-Grade English Nonword Repetition Test

The first English nonword repetition test (Miettinen 2003) was conducted in October of the fifth grade. The 16 participants remaining at the time all took part in the test. While the test has a minor role in the present study, it is nevertheless used in some statistical analyses and merits a brief description here.

Design of the Test. The English nonword repetition test needed to be comparable with the Finnish version, which in turn was almost exactly like the sixth-grade Finnish nonword repetition test. In the English version there were fewer items, three of each syllable length from one to six syllables, making a total of 18 English nonwords. Six syllables was chosen to be the maximum length, as this was considered a challenging enough length for the participants, who had only studied English for less than five semesters.

Unlike the Finnish nonwords in the fifth grade, the English ones were not created specifically for the task. Instead, they were adopted from previous studies where English nonwords had been created by native speakers of English, and complied with English phonotactics (Gathercole, Willis, Emslie and Baddeley 1991; Gathercole 1995). The one-syllable nonwords came from Gathercole, Willis, Emslie and Baddeley (1991) and the ones with two to five syllables were taken from Gathercole (1995). The six-syllable nonwords were combinations of shorter nonwords. The stimuli are listed in Appendix 2.

A male native speaker of English read the stimulus nonwords onto tape. Since most of the material used in the participants' English class at school was spoken with a British rather than American accent, the native speaker employed here also had a British accent. As in the Finnish version, the instructions were taped and they were read in Finnish by the same person. The wording of the instruction was somewhat different, since the language of the items was different. The first English nonword repetition test was piloted together with the fifth-grade Finnish nonword repetition test.

Procedure. The procedure was exactly the same as for the Finnish nonword repetition test. The participants were tested within a week of each other. Since the task no longer dealt with the native language, the criteria according to which scoring was done differed somewhat. Phonemes clearly deviating from the stimulus were not accepted (e.g. /p/ for /b/), but participants were allowed to sound Finnish in pronouncing the nonwords. The maximum (traditional) score was 18. The Baddeley (1993) score was not included in the present study.

5.4.3 The Sixth-Grade English Nonword Repetition Test

The second English nonword repetition test was carried out in the sixth grade, in April, a year and a half after the corresponding fifth-grade test and about a month later than the sixth-grade Finnish nonword repetition test.

Administration of the Task. As in the case of the Finnish version, the sixth-grade English nonword repetition test very closely followed the procedure of the fifth-grade test. The same stimulus nonwords were reread by a native English-speaking female and recorded onto a CD to be used in the test situation. The participants were instructed the same way as the year before and all their responses were recorded on a minidisk. Assessment was revised from the previous year (Miettinen 2003) by scoring the tasks only for the total number of correctly produced nonwords. The maximum number of points for the second English nonword repetition test was thus 18.

5.4.4 Inter-Rater Reliability in Sixth-Grade Nonword Repetition Tests

In the present study, all the scores used in the statistical analyses came from the same scorer, the present author; however, the reliability of some of these scores was estimated by having additional scorers assess tasks and then measuring the consistency between scores. Checking inter-rater reliability in this way is important when there is any chance that the assessment may be subjective, for example, when there are no clear cut-off points, i.e. the participant's response is either correct or not, and the rater decides whether something is good enough or where the item falls within a range of possible scores or conditions. Naturally, even the simplest recording of measurements of the objective kind can include errors, but when the assessment itself is undeniably subjective, as in most language-related measures, reliability is always an issue. In general, inter-rater reliability is of especial importance when not all raters are assessing all items. Consistently stricter or more lenient raters are not necessarily a problem, but if some of the data are only assessed by a stricter scorer and some by a more lenient one, the scores are bound to vary, not only because of the quality of whatever is being assessed but also because of who assigned the scores, which impairs the reliability of the scores. In the present study, however, more than one scorer was used only in the nonword repetition tests, and all the scorers assessed all of the items produced in a given task.

Two additional raters besides the present author were asked to rate both nonword tasks: two native Finnish speakers for the Finnish nonword task and two native English speakers for the English nonword repetition task. The raters were given the sound files of the participants' nonword repetition tasks as well as the original nonwords as an audio file and in writing. The raters were informed of the assessment criteria orally and in writing and instructed to listen to the items as many times as needed. They were also asked to be consistent across all items and participants should they not be sure how successful the repetition of an item was but were nevertheless to lean one way or the other.

The original scores by the present author and the total of four additional raters were entered in a PASW file separately for each participant and item. Since the slightly longer Finnish nonword repetition test consisted of 21 nonwords, each Finnish rater scored 315 items altogether. In the English nonword repetition test there were 18 items, resulting in 270 separate scores by each of the three raters.

The measure of agreement used in the present study was Cohen's kappa (κ). This kappa statistic is used with two raters assessing the same data (or two ratings of the same data by a single rater) and it takes into consideration not only the observed proportional agreement between the raters but also that which could have been achieved simply by chance (see e.g. Sim and Wright 2005). The kappa thus indicates to what extent the agreement between the raters exceeds that which could be due to chance. $\kappa = 1$ indicates perfect agreement between raters whereas $\kappa = 0$ indicates that the agreement is no greater than could have occurred at random. Also negative kappa values are possible but not likely.

The cross-tabulations of the original rater and the additional ones (F_1 and F_2 for the Finnish tasks and E_1 and E_2 for the English tasks) are provided below in Tables 4 and 5. They show the numbers of items each rater scored as correct and incorrect in the total columns, and also the numbers of how many scores of each kind – correct and incorrect – the raters agreed and disagreed on. The numbers of occasions of agreement between the raters are marked in bold.

TABLE 4 Cross-tabulations of the Finnish nonword test ratings.

		Rater F_1			*Rater F_2*		
		Incorrect	Correct	Total	Incorrect	Correct	Total
Original rater	Incorrect	**102**	8	110	**100**	10	110
	Correct	6	**199**	205	5	**200**	205
	Total	108	207	315	105	210	315

TABLE 5 Cross-tabulations of the English nonword test ratings.

		Rater E_1			*Rater E_2*		
		Incorrect	Correct	Total	Incorrect	Correct	Total
Original rater	Incorrect	**86**	12	98	**86**	12	98
	Correct	13	**159**	172	7	**165**	172
	Total	99	171	270	93	177	270

Based on Cohen's kappa, the inter-rater reliability was very good across the tasks. The results were slightly better for the Finnish nonword repetition test: $\kappa = 0.902$, 95 % confidence interval (CI) (0.851, 0,953) between the original rater and F_1, and $\kappa = 0.894$, 95 % CI (0.841, 0,947) between the original rater and F_2 (N = 315). For the English nonword repetition test the corresponding coefficients were 0.800, 95 % CI (0.726, 0,874) for E_1 and 0.846, 95 % CI (0.779, 0,913) for E_2

(N = 270). There are no general acceptability levels for kappa that would work for all possible kinds of data, but the coefficients given here can be considered high and the agreement between the raters appears to be very high. Benchmarks for different strengths of agreement have been suggested but are considered somewhat random (Sim and Wright 2005: 264).

Cohen's kappa is widely used and was also considered the best option for the purposes of the present study. While the agreement between the authors was very good, it was not perfect and in some cases the original score differed from both of the additional raters' assessment. Such cases were fewer in the Finnish nonword repetition test than in the English one. In two of the three cases, both of the additional native Finnish raters had accepted something that to the original rater was clearly a case of an incorrect final vowel sound, *kitsamello* for *kitsamellu* and *vuoksattiisisuilasma* for *vuoksattiisisuilasmo*. In the third, case the two additional raters had not awarded a point for a nonword that the original rater had accepted.

The seven cases where the original assessment of the English nonwords differed from both of the native English raters were all cases where the natives had accepted nonwords with phonemes fairly clearly incorrect or missing, for example, *blonterstapin'* for *blonterstaping, skiticaught* for *skiticult, commorine* for *commerine, empriforvent* for *empliforvent*, and *fersatrationist* for *versatrationist*. From this perspective the present author was stricter in assessing the tasks; however, looking at the cross tabulations, the total numbers of incorrect and correct nonwords are roughly equal (Original: 63.7% correct, E_1: 63.3% correct, E_2: 65.5% correct).

Despite these few cases, it can be concluded that the inter-rater reliability in both Finnish and English nonword repetition tests was very high and that it was safe to proceed using the original scores in subsequent statistical analyses in order to uncover possible relationships between phonological working memory and L2 knowledge.

5.5 Instruments and Procedure: Measures of Knowledge of English

The present study looks at language from a cognitive perspective, and hence while it is understood that social aspects such as interaction have an impact on a learner's knowledge or proficiency in an L2 at any given moment, a more restricted view of language knowledge is adopted here. One reason for this is ease of assessment. The aim was to gain quantitative information of the participants' language knowledge in order to compare it to their phonological working memory ability. As resources were limited, it was also sensible to stick to what is considered to be the core of language proficiency: lexical items and grammatical construction (see e.g. Hulstijn 2007). Language is thus viewed as consisting of vocabulary and grammatical structures; however, the separation

of form and meaning is also recognized. The measures used to assess L2 knowledge were chosen accordingly.

The participants had begun studying English as their first L2 in the third grade. English language data were collected specifically for the present study in grade six, but some data from grade five used in a previous study (Miettinen 2003) were also included. By the sixth-grade data collection the participants had been studying English for almost 8 semesters.

The aim of the sixth-grade English measures was to gain as broad a picture of the participant knowledge of L2 English as possible. This entailed collecting data by methods that also allowed many approaches in the scoring stage. The levels on which language is looked at in the present study are at least the levels of vocabulary, clauses and text that goes beyond the length of a sentence. The morpheme level is also considered when tasks are scored for grammatical correctness. To be exact, even individual phonemes come into play when the oral tasks, at least the vocabulary task, are scored. This brings us to the other issue on broadening language tasks. In addition to looking at language on many levels, both the written and spoken modality were included as far as the participants' own language production is concerned. This was done for several reasons. First of all, there could be differences between the participants in that some do better on written tasks while others are more successful in speaking the L2. By employing both written and spoken production the effect of such, possibly systematic, differences can be controlled for in the statistical analysis. Furthermore, as the nonword repetition tests were auditory and oral only, it was of interest to see whether the shared modality is reflected in the connections found between success in the nonword tests and in the L2 tasks. Finally, the previous studies show a fairly clear bias in favor of written tasks, and thus the oral ones included in the present study are an attempt to bring something new to the table. All of the L2 tasks also required active production and were thus in line with the nonword repetition tests used. This links with the comprehensiveness of the tasks. Such tasks may be more demanding of the participants than, for example, recognition-based phonological working memory tests and L2 tasks, but at the same time many more conclusions can be drawn from, for example, a sentence written by a participant than from a yes/no answer.

It should be emphasized here that what the present study sought to measure was the participants' knowledge of English vocabulary and structures on different levels, and that in the attempt to gain this information the participants' knowledge of English pronunciation and spelling also played a part.

5.5.1 Fifth-Grade Tasks

These tasks and their scores come from a previous study (Miettinen 2003) with the same participants. They are described here briefly, as the same data are used in answering some of the present research questions (for more detailed information on the tasks see Miettinen 2003).

5.5.1.1 Vocabulary Task

A separate vocabulary task was included in order to gain a general picture of the participants' vocabulary knowledge. The participants took part in vocabulary quizzes as part of their regular school work, but these exams usually covered only a few text book chapters and the children studied for them beforehand. A specifically constructed vocabulary task guaranteed a wider array of vocabulary items for assessment, and in the absence of advance warning, it did not measure the participants' ability to memorize vocabulary for a specific purpose.

A written constructed answer format was employed in the vocabulary task as it has been found most suitable for this type of testing by Takala (1984). It was a fairly straightforward paper-and-pencil format, which the participants had plenty of previous experience with. It was also quite quick to implement, since the whole class could do it at the same time, and their familiarity with the format allowed the participants to get right to work without first having to grasp the workings of a new task type. All the participants had to do was write down Finnish words that corresponded to the English items presented to them in writing and vice versa.

Design of the Task. The test used isolated words so that no other kind of language knowledge could affect the results of the vocabulary task. A sample of 60 items was considered to provide a reliable estimate of the participants' knowledge of English vocabulary. Takala (1984) found no difference between active and passive English vocabularies with Finnish learners of English, but nonetheless both were tested in the present study. There was no reason to choose one language or the other, and so the 60 items were divided into half: for 30 English items their Finnish equivalent was to be given and for 30 Finnish items a corresponding English word had to be written.

The same guidelines were used in selecting the task items as in Takala (1984). The English textbook used in class, *Yes Adventures* (Westlake, Lintunen, Pitkänen and Satamo 1997), contained vocabulary lists that included the vocabulary from previous books and their accompanying workbooks. There were separate lists of words translated from English to Finnish and from Finnish to English. All the items were numbered by the present author and 30 were selected from each list at random using Research Randomizer (www.randomizer.org). It was decided beforehand which types of items would be rejected should they be included in the random list of 60 items. Those included adverbs regularly derived from adjectives, inflected forms of verbs, nouns, and adverbs, and words from texts that had not yet been studied. Such items as irregular verbs, proper names, phrases and idioms were included. Finally, the selected items were ordered alphabetically for the task.

A pilot test was conducted with an adult female native speaker of Finnish in order to make sure that the oral instructions were sufficient, that no problems would arise during the task or scoring, and also to estimate the time needed for the task.

Procedure. All of the 16 participants completed the vocabulary task during the same day in December in the fifth grade. The participants were divided into three different English classes. The task was administered in the beginning of the class. Moreover, all the other students in the groups took part in the task, even though they were not participating in the study.

Oral instructions were given before handing out the task sheets. The sheets themselves contained no instructions, only the lists of items with an empty line beside them on which to write down the equivalent word in Finnish or, on the flipside, in English. The students were instructed to take all the time they needed and to write down something if anything came to mind that could be the item asked for. The task took approximately ten minutes in each of the three groups.

Each item in the vocabulary task was worth one point, making the maximum for the whole task 60 points. Half points were not given, but the given answer did not have to be one of the alternatives offered in the textbook. In the Finnish responses inflection was considered acceptable, for example *seuraa* for *follow*. In the English responses to the Finnish items, exactly correct spellings were not required, since the participants were such young and novice learners of English. Minor misspellings (e.g. *woter* for *water*) were acceptable as were more serious ones, as long as it was obvious to the experimenter which word was aimed at. These cases included, for example, spellings showing that the participant knew roughly what the word sounds like but could not produce the correct spelling (e.g. *beasy* for *busy*).

5.5.1.2 Other Tasks

In addition to the vocabulary task described above, the participants' knowledge of English was assessed with six tasks taken from an English exam given by the teacher as part of their normal school work, and an additional sentence formation task. The exam tasks were Listening comprehension, Reading comprehension, Verb form task 1–3, and Sentence formation 1. All but the listening comprehension task were rescored by the present author for the purposes of the study, since the original teacher scores reflected the pedagogical aims of the teacher. As a rule, points were not deducted for spelling errors if they were not too obtrusive or did not change the meaning of the word. Below are short descriptions of the tasks and their scoring.

In the **listening comprehension task** the teacher read aloud ten English questions. The exam paper had 12 answers and the students had to write down the number of the question next to the answer that best suited the question. The two extra answers remained blank. The questions were not necessarily about the students or about any general facts, but the correctness of the answer depended on their grammatical or semantic fit, for example a question such as *What language do you speak?* had to be answered with *Spanish*. Each correct answer gave one point and the maximum for the task was ten points.

The **reading comprehension task** presented the participants with a short journal entry type text which recounted the events of a school day. There were ten questions on the text in Finnish and answers were also required in Finnish. One point per correct answer made the total of the task ten points.

The **three verb form tasks** were structurally very similar cloze tasks where verbs or parts of them had to be filled in, in their correct form. In **Verb form task 1**, there were ten sentences with the lexical verb missing in each. The participants had to choose the appropriate verbs from 12 alternatives. The verbs had to be inflected to agree with the subjects of the sentences. Half a point was awarded for selecting the correct verb, another half for the correct inflection. This brought the total of the task to ten points. **Verb form task 2** introduced another set of ten sentences with gaps to fill. This time the options were the third person *–s*, *do*, *does*, *don't*, and *doesn't*. One of the gaps had to be left blank, since the sentence in question was already grammatically correct. One point was given for each correct choice and the maximum for the task came to ten points. **Verb form task 3** consisted of a short story told in the first person. The ten gaps in the text had to be filled with one of the verb forms *am*, *is*, *are*, *have* or *has*. The maximum for the task was 10 points, since each correct choice was worth one point.

The last of the six exam tasks, **Sentence formation 1**, featured two pictures: one of a boy and one of a girl. The participants were asked to write about the two children in the pictures using Finnish cue words (e.g. *ei koskaan tuhma (Eng. never naughty)*, *tappelee siskonsa kanssa (Eng. fights with sister)*). Four sentences were expected for each picture, and each sentence was worth five points. Thus, the maximum score for the task was 40 points. A grammatically correct sentence which corresponded to the cue words gave the full five points. One grammatical error resulted in the deduction of one point so that a sentence like *He help his mom* was worth four points because of the missing third person *–s*. On the other hand, points were given if there was at least something correct about the answer. Thus, for example, *No naughti* for the expected *He is never naughty* was worth one point.

In addition to the exam tasks there was **Sentence formation task 2**, which was carried out for the purposes of the research project but partly constructed by the participants' English teacher. The task was originally used to test the participants' knowledge of prepositions as part of a practice exam. The task consisted of a picture and ten questions about the locations of certain objects in the picture. The participants had to answer the questions with sentences. Prepositions were naturally needed to give the appropriate locations (e.g. Question: *Where is the chair?* Acceptable answer: *The chair is behind the tree*). All of the questions were structurally the same as were the expected answers. The prepositions were not the only factor of interest in the present study, however. The sentences were also scored for word order, use of articles, and verb forms. Six points were given for each correct answer, and with ten sentences in all, the maximum was 60 points. Since an idiomatic answer to these types of questions is not necessarily a full sentence, communicativeness was considered in scoring

and as many as 4.5 points were given for mere prepositional phrases such as *Behind the tree*.

5.5.2 Sixth-Grade Tasks

Compared to the fifth-grade tasks, L2 assessment in the sixth grade was conducted very differently. There was an attempt to form a kind of assessment continuum starting with nonword repetition so that the language tasks also would be based on repetition of heard stimuli. The motivation behind the repetition-type tasks was to get more diverse and perhaps different kinds of information about the participants' language knowledge than would have been possible with more typically school-like paper-and-pencil tasks. On a more theoretical level, auditory stimuli were especially suitable for L2 tasks which were to be correlated with phonological working memory, since phonological working memory is thought to play a role in language learning by establishing long-term *phonological* representations of both lexical items of the language to be learnt and its syntactic constructions (e.g. Adams and Willis 2001: 92). Furthermore, as the tasks were not as clearly targeted, for example, to assess knowledge of certain grammatical features, more room was left for different ways of scoring the tasks.

Thus, the original idea was to have the participants hear and repeat (both in spoken and written form) stretches of language that would gain in length from single words through sentences to a whole little story. In the end, however, only the sentences followed the planned format exactly. Since the participants had been studying English as a foreign language for almost four school years at the time of testing, repeating single English words was considered too easy a task for them. Such a task would probably have led to a ceiling effect and not distinguished between the participants, rendering the task fairly uninformative or useless for further analyses. The vocabulary task was finally designed so that the participants did hear the stimuli, but instead of repeating the words they were required to produce translations, written or spoken, of the heard vocabulary stimuli.

As for the story, the participants were asked to write it down as well as they could after hearing it, but the planned spoken repetition was left out of the battery of tasks. Remembering and repeating a longer piece of English language was considered too demanding and stressful for the young participants. Conducting such a task would also have taken a lot of time, since all of the tasks eliciting oral data had to be carried out with each participant individually. As becomes clear in the task design description below, the story was first read to the students three times before they were asked to reproduce it (in writing). Doing this separately with each subject for the purposes of oral production would simply have required too much time. Since even the written task proved very demanding for some, a spoken version would probably have placed too much pressure on the participants. The designs and procedures of all of the sixth-grade English tasks are described in detail in section 5.5.2 and the scoring of each task is discussed in section 5.5.2.4.

5.5.2.1 Vocabulary Tasks

Because of the continuum concept of the sixth-grade L2 assessment and the fact that everything was based on auditory stimuli, the vocabulary task was somewhat different from that in the fifth grade (section 5.5.1.1). All 15 participants took part in the vocabulary tasks. The written version was conducted in two halves on two consecutive days in March and the spoken version within a few days in April.

Design of the Task. The vocabulary items were selected following the same procedure as in the fifth-grade vocabulary task, only this time more chapters of the textbook were included in the range of source material, since the sixth-grade test took place over a year later. The criteria for selecting the items were also the same but the number of items was different. Since there were now separate tests for written and oral production, in was decided that 20 items would be sufficient for translations from English to Finnish and another 20 items from Finnish to English. Since the same number of items was translated in both task types, there were altogether 80 items to be translated, 40 (20 + 20) in spoken form and 40 (20 + 20) in writing. The vocabulary lists are included in Appendix 3 (written) and Appendix 4 (spoken) in the order of testing.

The Finnish vocabulary items were read on a minidisk by a female native speaker of Finnish and the English words were read by a female native speaker of English. The words were presented in alphabetical order with an 8-second pause in between items to allow for translation. The complete lists were recorded on a CD.

Procedure. In the written vocabulary, task the participants heard the taped stimuli and wrote their Finnish or English translation responses on a sheet with lines and numbers corresponding to those on the CD. There was no written form of the stimuli given anywhere so the participants had to rely on what they heard and understood from the CD. The written vocabulary task was quite long, and hence it was divided between two testing days. In the sixth grade, the participants were not all in one and the same English class but in two different groups that met on the same day. Both groups did the English to Finnish translation all on the same day in March at the beginning of their respective English classes. All but two participants did the Finnish to English translation on the previous day, similarly at the beginning of an English class. The two students absent on that day did the task in an empty classroom two days later.

Before the task began, the students were given oral instructions on what was going to happen and what they were required to do. They were told that they would hear Finnish (or English) words one at a time from a CD and they would have to write down what they thought their equivalents were in English (or Finnish). They were told that there were many kinds of words on the CD, things like ‘nukkua’ (sleep), ‘talo’ (house), or ‘että’ (that). Terms like *adjective*, *verb*, et cetera, were avoided on purpose. The students were encouraged to

write the word down even if they were not sure of the spelling and told that they could also write down several words if they could think of more than one. They were told to listen carefully, and that they would hear the number of the word (corresponding to the numbered lines on their answer sheets) and then the same word twice, and that they should write down their answer after that. There was also one practice item before the actual task, both in the spoken and the written version.

The stimulus CD was played on a CD player in front of the classroom and the students wrote down their responses during the predetermined pauses. After the last item they were instructed to go over their answer sheets for possible spelling errors, et cetera, after which the sheets were collected. The procedure was exactly the same for both directions of translation.

Like all tasks requiring oral production, the spoken vocabulary task was conducted individually in a quiet room. The instructions for the spoken version of the vocabulary task were the same as for the written one, but here the participants were asked to say the corresponding English (or Finnish) word out loud and encouraged to guess, if they were not sure. The participants heard the stimuli from a portable CD player through earphones and the output was recorded on a minidisk. After each stimulus word was heard twice, the CD was paused and the participants were allowed as long as they needed to produce the translation. If they did not produce anything they were encouraged to do so and the task was continued after they either produced something or declined to do so. In the spoken version of the vocabulary task, both translation directions were completed in the same session. This took quite a long time, depending, of course, on how long the participants took to respond, but not as long as the written task, since the pauses in between the items were shorter. Spoken production was faster than written and longer pauses were only used when needed.

5.5.2.2 Sentence Repetition

Two elicited imitation tasks were carried out to assess the participants' English (L2) knowledge in the sixth grade. One task required written reproduction of heard stimulus sentences and the other spoken reproduction of similar sentences. The design and procedure of the elicited imitation tasks are described and discussed in the following two sections. The test sentences are provided in Appendices 5 and 6.

5.5.2.2.1 Written Elicited Imitation Task

The elicited imitation sentences were designed so that they covered the major grammatical issues of the participants' sixth-grade instruction up to the time of testing. These included verb inflection (person, tense, negative, progressive), pronouns (person, possessive), prepositions and adjectives (comparison). The same grammar points were repeated in several of the sentences so that vocabulary knowledge would have as little effect on understanding or

reproducing the syntax as possible, i.e. a particular grammatical aspect would not be repeated incorrectly solely because the vocabulary was unknown. In any case, the vocabulary of the sentences was also familiar to the participants from their textbooks. A total of 20 sentences were created: ten were needed for the written repetition and ten for spoken repetition. The sentences were divided into the two groups of ten as randomly as possible, with the exception that the same aspects of grammar had to appear several times not only in one of the lists but in both for the written and oral tasks to be comparable.

The length of the sentences was approximately 15 syllables to ensure, for example, that they would not be too short making the task a test of memory. All the sentences also had to be approximately of the same length to be comparable. The sentences were read onto a minidisk by a native English speaking female (the same as for the sixth-grade English nonwords and the English vocabulary task stimuli) and then transferred on to a CD to be played to the participants. Using prerecorded stimuli ensured at least some level of equality in experimental conditions for all the participants.

Each stimulus sentence was heard twice and was preceded by a (Finnish) number from one to ten to make the test design clear to the participants. The length of pauses in between different sentences was not predetermined. In other words, the participants were allowed all the time they needed to reproduce the sentences both in the written and the oral elicited imitation task. The pauses from the end of one sentence to the beginning of the next one ended up being approximately 20 seconds long in the written task.

Procedure. The written elicited imitation task was conducted as a group test. All 15 participants were available for this task, but since the participants were divided into two different groups for English instruction, they were given the task at different times. About half of the participants did the task one week later than the other half in April.

The participants were told that they would hear English sentences that were not connected to each other. They were encouraged to listen carefully and focus on understanding the sentences and to write each sentence down in English as well as they could remember, but only after having heard them twice. If they could not remember the exact words in the sentence, they could use their own words or even Finnish. This was to ensure that the participants would try to produce something if they understood what was said on the tape but could not remember the exact words. The participants were also reminded to read each sentence they had written to check for spelling errors, et cetera, and that there would be plenty of time to do so.

The participants were given a sheet of paper with lines and numbers corresponding to the numbers of the sentences on the tape. After the instructions, the stimulus sentences were played on a CD player in front of the classroom. The CD was paused after each sentence had been played twice to allow the students time to write down their responses. The answer sheets were collected after the participants had finished writing down the last sentence.

5.5.2.2.2 Spoken Elicited Imitation Task

The spoken elicited imitation task was conducted as an individual test one or two weeks after the written version. All 15 participants were available for testing.

The spoken version of the elicited imitation task was carried out individually in a quiet room. The recording of the stimulus sentences was similar to the one for the written version, so that each of the ten sentences was preceded by a number and all the sentences were played twice before allowing repetition. The participants heard the sentences through headphones from a portable CD player and their responses were recorded on a minidisk. The instructions were the same as in the written task, only this time the participants were encouraged to *say* something even if they were not sure of what they had heard. They were also given an example sentence (*I like ice cream.*) and asked what they would do if that was what they heard on the CD. This was done to make sure that they understood the instructions and would not try to translate the sentences into Finnish. The participants were also told that it was acceptable to start the sentence again and make corrections just as they could have done in the written version of the task. Again, the participants were allowed all the time they needed to repeat the sentence to the best of their ability. If nothing was produced, they were encouraged to say anything that came to mind about the sentence before moving on to the next sentence. No record was made of the length of pauses between sentences in the spoken repetition task.

5.5.2.3 The Story Retelling Task

The longest stretch of language the participants had to produce in the present study was the short story in the story task. Just as in the sentence repetition tasks described above, in the story task the participants had to first understand an L2 story and then apply the vocabulary and structures at their disposal to reconstruct the story. This way the same task was used to test both listening comprehension and written production. The story was only produced in writing.

Design of the Task. Since no suitable story was easily found, one had to be created. Knowing the type of material that the participants had worked with in their English classes, a suitable topic was decided on. The story was written and revised until the sentence structures and vocabulary of the story were suitable in light of the participants' age and experience with English. It ended up consisting of seven sentences, which was considered long enough for the sentences to make up a story but short enough to not be overwhelming for the participants. The story is provided below:

Mary's day at the zoo

It was Saturday morning. Mary and her brother Bill went to the zoo. There was a big lion in a cage. He looked hungry and a little sad.
"Can I give him my apple?" asked Mary.
"Sure, why not" said Bill.
So Mary gave her apple to the lion and it made the lion very happy.

Procedure. Both of the teaching groups had English during the same day, so all of the participants completed the story task at the beginning of their English class on the same day in March. Also, the children not participating in the research project but attending the same English classes took part in the task. The students were given papers with empty lines to write on and careful instructions were given orally on what was going to happen during the task and what they were required to do. The core of the instructions was also visible on a transparency. The students were told that they would hear a short story in English three times and that they should listen carefully and focus on understanding the story. They were told not to write anything down until given permission to do so after the third reading. They were advised to use their own words or even insert Finnish words if they could not remember how the story actually went. Finally, they were instructed to check their text once it was ready to make sure the English was as good as possible and to sit quietly until everyone was finished. There was no time limit for writing.

The story was then read out loud live three times in front of the classroom (by a native Finnish speaker proficient enough in English to be clear). Because the procedure was not likely to be familiar to the participants, they were not allowed to take notes as it was feared that they would try to write complete sentences and forget to listen. The prohibition on notes was compensated for by three repetition of the story instead of the usual two. The story was also considered short and simple enough for reconstruction to be possible without notes. Once the students had all finished writing, the papers were collected. The whole task took about ten minutes in both groups.

5.5.2.4 Scoring of the Sixth-Grade English Tasks

The story retelling task and the spoken and written sentence tasks were scored separately for meaning and structure. Scheibner-Herzig et al. (1991) also rated their elicited imitation tasks separately for meaning and grammar, although they used global scores on a scale from one to five. Meaning and structure are, of course, necessarily somewhat intertwined but the purpose of the double scoring was to obtain a single detailed, grammar-oriented, traditional assessment which could be compared with a score showing the participants'

understanding of the stimuli and their ability to reconstruct and convey the main idea in English. The division could be seen as a surface versus core idea type of classification. Looking at both structure and meaning is also in line with the aim of the present study, i.e. to obtain as broad a picture of the participants' L2 knowledge as possible. Furthermore, using phonological working memory capacity separately as a predictor of the meaning and structure scores of the same L2 task has not been done previously.

Below, the scoring criteria for the vocabulary tasks, sentence repetition tasks and the story task are explained. As the sentence and story tasks were scored separately for structure and meaning, there were naturally two different scoring criteria and maximum scores for those tasks. The maximum scores of all the tasks are presented in Table 6.

TABLE 6 The maximum scores of the sixth-grade L2 measures.

Vocabulary		*Sentences*		*Story*	
Written production	40	Structure (written production)	157	Structure	72
Spoken production	40	Meaning (written production)	91	Meaning	39
		Structure (spoken production)	149		
		Meaning (spoken production)	90		

Vocabulary Score. Written and spoken vocabulary were scored according to similar guidelines, even though the difference in modality resulted in some necessarily different considerations. The spoken vocabulary task was assessed on the basis of the audio recordings of the participants' productions. In all cases, pronunciation was clear enough to determine which words had been meant. If the utterances were considered acceptable responses to the stimulus words, a point was rewarded for each. In the case of written vocabulary, spelling errors were not considered a problem, if it was clear which word was meant, for example, *weard* for *weird* was considered acceptable. Understandably, spelling was only an issue when the participants were asked for the written English counterparts of Finnish stimuli. No participants had difficulties in spelling in L1 Finnish such as would have affected the scoring process.

The use of articles was not considered relevant in this task. In some cases this resulted in changes in the word class of an item, but that was not considered a problem, since these were words in isolation. For example, the word *ring* was allowed to be interpreted as a verb but also as a noun (a piece of jewelry), even though it was not preceded by article as were other nouns in the stimuli. Because of the focus on the meaning of the items, changes between singular and plural forms did not affect the score either (e.g. *vihannekset*, a plural in Finnish in response to an English singular *a vegetable*).

Each item in the vocabulary tasks was scored as correct or incorrect, and one or zero points were awarded accordingly. In the case of the written vocabulary task, it would have been feasible to subtract something for spelling

errors that were not detrimental to the meaning of the item. However, that would have changed the balance between the written and spoken task. In the spoken vocabulary task the words uttered were all clearly understandable, so awarding something between one and zero points would have been difficult. If the pronunciation of an English word was very Finnish, i.e. the word was pronounced as if it were a Finnish word, this was not considered an error. However, if the pronunciation was not consistent (e.g. [enivai] for *anyway*), points were not awarded. In such cases, it could not be known whether the participant could not recall the correct pronunciation or the correct spelling, so the line was drawn at having to be consistent within a lexical item. Again, this was only an issue when the participants produced English equivalents for Finnish stimuli. There were no problems in understanding the Finnish responses to English words.

In both the written and spoken vocabulary tasks, 20 items were heard in Finnish and 20 in English, and thus the maximum score for each task was 40.

Sentences: Spoken and Written Production. Before scoring could begin, the spoken sentences were first transcribed into written form. The problems related to that were mostly to do with pronunciation. Sounding Finnish was usually not a problem as it did not cause changes to the meanings of words, but general unclarity and occasional muttering were more common. This spoiled parts of clauses in some cases, and also meant that what could not be heard could not be assessed. Luckily, there were only a few clauses (0.8 % of the productions) of this kind and they were fraught with other problems as well.

Many pronunciation issues were not problematic in the end because of the context. Even if the meaning of a word in isolation was changed it was sometimes clear what was meant on the basis of the surrounding phrase, for example, *with some udder friends* and *with some utter friends* for *with some other friends*. In some cases it was not obvious whether another word was meant or if there was a pronunciation problem, for example, *French* vs. *friends*. The decision had to be made when transcribing, and the scoring was based on that decision.

When comparing spoken and written sentences it seemed that the spoken sentences contained more added words than the written ones. In normal speech, an extra word can accidentally slip out in a false start or otherwise; this is a natural part of spoken interaction and not a problem. If it happens in writing there is often the chance to take another look at the text and erase things put in mistakenly. Unfortunately, in the spoken sentence task, every little addition had to be taken into account, and it turned out that the transcribed responses were certainly more exotic than those written by the participants, for example, *he can't go home any yet*. On the other hand, spelling errors were obviously not an issue with the spoken sentences, but were a major problem with the written ones.

Below the scoring for meaning and structure of both the spoken and written sentences are described, and the descriptions apply to both tasks unless otherwise stated.

Sentences: The Meaning Score. Scoring the sentences for their meaning centered on the idea of semantic propositions and accompanying details. In particular, the focus was on the predicator and on the subject and complements linked to it, i.e. what happens, who does what: for example, someone (A) went somewhere (B). The division of scores was based on Feldman (1985), but the same method has been widely used. The procedure is described next.

Propositions and details were first identified in the (clauses of the) sentences. For example, the clause *Emma can speak French* contained the proposition that someone knows a language and the details were *Emma* and *French*. After that, the participants' productions were compared to the list of propositions and details and points were rewarded for matches. The propositions themselves could yield one or two points. If the gist of the proposition was present, for example, A doing B, the score was two points. If the proposition was not quite complete, one point was awarded. In the case of the example clause given above, the full two points would have been awarded for something like *Emma speaks French*, whereas *Emma likes French* would have yielded only one point. In addition to the proposition, each correct detail (or an appropriate substitution) was worth one point. In terms of details, the two previous examples would have been equally correct and worth two points.

Whereas Feldman (1985) scored details only if the gist or a fraction of the proposition was correct, in the present study the definition of a fragment was stricter, and points were given for correct details even if nothing was given for the proposition itself. For example, Feldman (1985) gave fraction points even for the subject alone, but in the present study the fraction point usually required the correct subject and perhaps even the correct verb in the predicator, otherwise a completely wrong idea might have been expressed, for which one point would have been awarded for a fraction of the proposition. Anything less was awarded no proposition points, but a point for each correct detail could still be given. For example, for the first clause of a sentence that was originally *He does his job very well* a response such as *Hi doesn't jaxt very well* received no points for the gist but two points for details (*hi* was considered a spelling error). These cases were very rare, however, as expressing the correct gist or fragment of the proposition usually included correct details as well and, vice versa, managing to produce correct details usually went together with at least a correct fragment of the proposition.

Compared to the story, a wider context was missing from the sentences, but in most sentences there were two clauses for the latter of which similar allowances were made as could be made in the story. For example, words could be left out because repeating something that is given in the previous clause was not always necessary for the sentence to convey the same meaning. For example, in a sentence like *Tom likes ice-cream but he doesn't eat it at school*, the pronoun *he* could be left out without any consequences for the meaning or the score of the sentence. Pronoun replacements, et cetera, were also acceptable where suitable. However, although some words were not necessary to be given points for the propositional meaning, their omission meant that the detail points were lost.

Furthermore, for the detail score, an internal logic was required. For example, if in the first clause of the sentence the plural noun *friends* was used, the pronoun *they* would be expected in the latter clause. If the participant used the singular noun *friend* instead of the plural, full points were only awarded in the latter clause for the use of the appropriate singular pronoun *he* or *she*.

The maximum meaning points per sentence varied from seven to ten in both the spoken and written sentence tasks. The maximum score for meaning was 90 in the spoken sentences and 91 in the written sentences.

Sentences: Structure Score. The structure scoring of the sentences was essentially based on lexical and grammatical morphemes being present and correct in the participants' productions. General guidelines (explained below) for the distribution of scores were made, after which all the different versions found in the participants' sentences were collected in grids according to word class or type of phrase. In this way all the possible problems in a certain word or phrase could be considered at the same time and different erroneous versions of a particular word or phrase could be assigned scores which were then considered in scoring the stories as a whole. This way it was made certain that all the participants were treated equally and the same problem in the same word was scored the same across participants. The scores were also similar across the word class or type of phrase so that a certain problem in one noun was equally serious in another noun.

Similarly to the meaning scores, predicators were considered central for the structure of clauses as well and worth two points. If there were auxiliaries in the verb phrase, they were worth two points each and the lexical verb was worth the normal two points as well. In the case of the progressive aspect, one point was given for the correct lexical verb and one for the progressive *–ing*, while in negation one point was given for the lexical verb and one additional point for the negative *-n't*. Points were subtracted for errors but never beyond the maximum for any word or phrase. One point was deducted for the incorrect tense or number or -1.5 points if both were wrong. If a grammatically incorrect element was added, a point was deducted and a missing apostrophe, for example, caused half a point to be deducted. Spelling errors were penalized with the deduction of 1 or 0.5 points depending on their severity, for example *woshing* for *washing* and *coing* for *going* the deduction was only half a point. Many of the issues mentioned here were naturally only relevant in scoring the written sentence tasks, not the spoken ones.

Each noun or pronoun was worth one point and an additional point was given for the use of an appropriate article. Possessive pronouns like *her* or *his* were worth two points since they were awarded for gender and possession separately. Plural forms were also worth two points, one for the noun itself and one for the use of a plural. Minor spelling errors caused half a point to be deducted, while for a more serious one a whole point was deducted, for example *jop* for *job* was considered only a half-point error.

Adjectives and possible modifiers were worth one point each. Acceptable substitutions as to meaning were awarded full points, for example, *three big brothers* instead of *three older brothers*. In the case of nouns, minor spelling errors in adjectives cost 0.5 points whereas serious spelling errors caused the deduction of one point.

As for prepositions, each correct one was worth two points, while knowledge of a preposition being required was rewarded such that an incorrect preposition was worth one point.

On a more general note, semantically acceptable replacements were awarded full points on the word level, and full points were also given for a response that did not aim for the same structure as the stimulus, but for the same meaning. Points were then deducted for errors in regard to the particular structure that the participant was aiming to reproduce, not in regard to the original stimulus. No points were given for Finnish substitutions even if they reflected understanding of the stimulus. For a missing lexical or grammatical morpheme, the maximum points for such morphemes were deducted.

The maximum structure points varied from 13 to 17 per sentence. The maximum structure score for the spoken sentence task was 149. In the written version it was possible to score 157 points.

Story: Meaning Score. The rationale for scoring the story for meaning was similar to that for scoring the sentences described above. The propositions and attached details were first identified and listed for each clause. As with the sentences, correct production of the gist or its fraction yielded two points or one point, respectively, whereas each correct or appropriately substituted detail was worth a point. For example, for the sentence *Mary and her brother Bill went to the zoo*, a response like *Mary and hes brother Bill goes to the zoo* yielded full points for the gist of proposition and details, whereas *There was a big lion* for *There was a big lion in a cage* yielded only a fragment point for the proposition and obviously not the full detail points since the cage was not mentioned.

Since it is the nature of a story to revolve around certain characters throughout and the events somewhat logically follow each other in the storyline, the repetition of every element was not required for the full two points for the gist of the proposition. In other words, the fact that this was a story was taken into consideration already at the proposition defining stage. For example, in the original story there is a line "*Sure, why not*", *said Bill* but because there are only two people in the story and this is a response to Mary's question, it is obvious that it is Bill who says this and only some formulation containing the idea of giving permission was required for the proposition points to be awarded.

The story context was also considered in scoring details, most obviously in accepting appropriate substituting pronouns for nouns. The numbers of errors and omissions were logged, although they were not relevant for the story meaning score. Similarly both logical and illogical additions were logged, but they did not affect the meaning score, i.e. if the participant produced the correct proposition and something extra, whether it was logical or not in the context of

the story, points were not deducted. Besides, such cases were rare in the data. One participant added description of Mary's brother Bill by writing *little brother Bill* and one participant was more prolific adding whole new parts to the story, for example, Mary saying that her knee is hurt and that she is hungry. Additions were not penalized because if they replaced something, then the points for the replaced parts had obviously been lost and the score was already affected. Also, the participants were encouraged to write as much as they could remember, which could have encouraged them to be creative at times. They were also not told to be careful about not adding anything of their own, so it is difficult to say whether the few who added anything really understood the story differently or were not sure if they had produced enough and wanted to add something.

Spelling errors did not result in minus points if the error did not cause changes in meaning. This was judged case by case. Even some very exotic spellings were accepted if it was obvious what word was meant and it was clear from the surrounding story that the participant knew the meaning but not the spelling of the lexical item. For example, *cate* for *cage* and *zoon* for *zoo* were considered acceptable forms.

The heading of the story (Mary's day at the zoo) was left out of the assessment because the participants were not explicitly asked to include it and it does not have propositional content similar to that of the rest of the assessed material. For the sake of comparability the title was not included in the structure score either.

The maximum scores (gist and details) varied from three to six between propositions and the maximum meaning score for the whole story was 39.

Story: Structure Score. Basically the same guidelines applied to the story as to the sentences, but because different structures were included, the scoring guidelines were partly different and are worth explaining here.

The predicators were once again an important part and points were not awarded for isolated words, even if correct, if a predicator was not present. The maximum score for each verb was two points and points were deducted if there were errors of tense or number (-1 point each or -1½ point if both wrong). Spelling errors and added elements in the verb phrases caused a point to be deducted (e.g. *chan* (for *can*), *make's*).

In the case of noun phrases, each noun was worth one point, but if more information was included, the noun could be worth two points, for example, in the possessive pronoun *her* there is both gender and possession, so both are worth a point. Using a suitable article was also worth a point. Minor spelling errors cost -½ point but two or more minor errors or a more serious error caused the deduction of a whole point. Examples of minor errors include *sadurday* for *Saturday* and *prother* for *brother*, whereas more serious errors were such as *zoom* for *zoo* and *leons* for *lion*. What is noteworthy here is that a spelling error considered minor when assigning meaning scores could be taken as serious when assessing structure.

Adjectives and quantifiers were worth one point each and a possible article was worth an additional point as was a modifier (*very happy*, 2 points; *a little sad*, 3 points). If anything was added, treatment depended on the grammaticality of the addition, and hence grammatically correct additions had no effect on the score. For example, *very hungry* instead of just *hungry* caused no score deductions because even though it is different from the original, there is nothing structurally wrong with it. As for nouns, minor spelling errors in adjectives were worth -½ point whereas serious spelling errors caused the deduction of one point.

Exactly as in the sentence score, propositions were worth two points, and for simply using a preposition, even if the wrong one, one point was awarded.

Since the maximum points for each lexical and grammatical morpheme were known, the maximum points for each clause and for the story as a whole could also be calculated. The maximum scores per clause varied from five to twelve and the maximum sum score was 72 points.

5.6 Instruments and Procedure: Measures of Knowledge of Finnish

Finnish language data already available in the longitudinal research project was used in some of the language specificity analyses. The test used was the Finnish reading test for grades 1–6 (ALLU), a standardized, valid and reliable reading test that can be used, for example, to assess reading level, to monitor year-to-year reading development and to diagnose reading difficulties (Lindeman 1998: vii, 44). It is based on how reading is understood in cognitive psychology and measures three components of reading: language awareness, technical reading skills and reading comprehension. However, in the sixth grade only the last two are assessed.

Lindeman (1998: 27–33) describes the sixth-grade tasks and their assessment in great detail. Reading comprehension, the part of the test used in the present study, is assessed with five types of questions (detail/fact, cause-effect/order, conclusion/interpretation, word/phrase and main idea / meaning) that measure both literal and interpretative understanding of different levels of information. The texts used are authentic excerpts from magazines and children's literature and while the topics are various, two of the texts in the test are factual and two are fictional.

According to Lindeman (1998), the emphasis in the ALLU test is on activating existing knowledge, connecting new information with existing knowledge and using reading strategies. The majority of the multiple-choice questions measure making inferences based on the text together with existing previous knowledge, while the rest measure understanding of information mentioned directly in the text. The extent of the stretch of text that provides the information needed in answering a question varies greatly. It could be based on

one, two or several sentences, a paragraph, or even the whole text and previous knowledge.

In the ALLU test, working independently, the participant reads four texts, two of which are factual and two fictional, and answers altogether 48 multiple-choice questions. The texts can be used and reread through the entire test and there is no time limit. In the sixth-grade test there are four answer options to choose from in each question and a point is awarded for each correct choice. No points are given or deducted for an incorrect or missing answer. The maximum score for each text is 12 points.

Since the Finnish measure used was a reading comprehension task, it was obviously different from the English measures but as a standardized, widely used measure it was definitely a solid measure of part of the participants' L1 knowledge. Indeed, it is difficult to imagine L1 tasks that would have been very much like the L2 tasks and still suitable for participants who have only been studying the L2 for under four school years. Thus, this was not a realistic goal for the L1 tasks in the present study. Furthermore, as previous findings support the association between phonological working memory and reading, there was no reason to suspect that a reading comprehension task would not be suitable for the purposes of the present study, i.e. for having a kind of L1 control task in the language specificity analyses. The main focus of the present study was in studying the connections between the L2 tasks and nonword repetition tests both in the L1 and the L2, but if the Finnish (L1) nonword repetition test could not be shown to be connected to a Finnish language task and thus be shown to function as expected based on previous research, further language specificity inquiries would not be very sensible.

5.7 Data Analyses

The first research question dealt with a possible connection between nonword repetition and language knowledge in English. Correlational analysis was deemed a suitable method for studying the possibility of a connection in that a positive correlation between the English nonword repetition test scores and the L2 English language measures would point to one depending on the other, or possibly a third variable that the two have in common. In any case a connection would be revealed by the correlational analysis, even if the reasons for it would not.

The second research question regarding the possible language specificity of phonological working memory was also broken down into studying connections. If nonword repetition tests in different languages are similarly connected to measures of different languages, then the phonological working memory construct is not likely to be language-specific. If there are differences in the connections then the language of the tasks could be considered to be the cause of the differences and phonological working memory function could be viewed as somehow dependent on the language in question and thus language-specific. The existence of a connection was again translated as a positive

correlation between scores. Other methods were then used to study such connections further.

In order to answer the research questions, data from the nonword repetition tests and language tasks were entered into PASW Statistics 18 for statistical analyses. A composite score of L2 knowledge in the sixth grade was formed by adding together the scaled scores of each of the different L2 tasks, including both modalities (written and spoken) and scoring methods (structure and meaning). The scores had to be adjusted to the same scale because of the great variation in maximum scores on the language tasks. As both the written and spoken sentence tasks were scored for both structure and meaning, the sentence tasks yielded four scores, whereas the vocabulary tasks were represented by two different scores (written and spoken), as was the story task (structure and meaning).

Correlational analyses were performed on nonword repetition and language task data to investigate the relationships between English nonword repetition and English language knowledge, as well as the different possible connections between the English and Finnish nonword repetition tests and English and Finnish language tasks. Both Spearman rank-order correlations and Pearson correlations were calculated for the nonword repetition tests and English composite scores and for each of the English tasks separately, as well as the Finnish language tasks. The different modalities (spoken and written) and the two assessment methods (structure and meaning) were also investigated separately. Spearman rank-order correlations were used owing to the relatively small number of participants, but because most of the variables came from a bivariate normal distribution, calculating Pearson correlations for them was also possible (e.g. Alkula, Pöntinen and Ylöstalo 1994: 241).

Cluster analysis was used to further study the connections between the nonword repetition tests and the language tasks. The groups that resulted from the cluster analysis were studied to obtain a better understanding of the connections between phonological working memory and language knowledge as the clustering was based on the standardized scores of both the L1 and L2 nonword and language tasks. The cluster solution(s) also served to identify individuals who, by their performance on the memory and language tasks, somehow differed from their cluster or the whole group, or whose profile seemed otherwise interesting.

Cluster analysis is a form of multivariate statistics used to group cases (or variables) into previously unknown groups so that the cases are similar to others within the group, i.e. the groups are homogenous, and as different from the other groups as possible, with respect to the variables included in the cluster analysis (Heikkilä 1998: 240). In the present study, this meant grouping the participants according to their scores on the memory and language tasks to see if any such homogenous groups exist within the whole group of participants, and, if so, whether it is possible to draw information on learner profiles from this.

Cluster analysis is not a unitary method in the sense that choices as to the way the clustering is carried out and numbers of clusters can be made at several stages and that ultimately it is the researcher who makes the choice between the clusters solutions by determining what makes sense content-wise. In carrying out cluster analysis, a number of decisions about the analytical process have to be made.

To begin with, there are several families of clustering methods (see e.g. Aldenderfer and Blashfield 1984: 35). The most commonly used method is hierarchical agglomerative cluster analysis, which produces the very helpful graph of the analytical process, the dendrogram. In this clustering method, the cases (or variables) start out as separate and are then joined stepwise into clusters based on similarity until all the cases form one big group (Everitt and Dunn 1991: 101). There are many possible techniques for this clustering procedure since there are several ways of determining the distance or similarities of the clusters.

There are four kinds of similarity measures, meaning calculating which clusters should be joined at which stage. According to Aldenderfer and Blashfield (1984: 22–25), the two that are used more in the social sciences are correlation coefficients and distance measures. Among the first group, Pearson's product moment correlation coefficient is the most popular. The distance measures could also be called dissimilarity measures, and two widely used ones are the squared Euclidian distance and Manhattan distance. With the basis for calculating the similarity or dissimilarity or distance between the clusters decided, several options remain regarding what to do with that information, i.e. how to use the similarity of distance between clusters as a tool in forming further clusters. According to Aldenderfer and Blashfield (1984: 38), the four most popular linkage methods are single linkage, complete linkage, average linkage and Ward's method. The single linkage method always joins the two clusters closest to each other in each step. If there are several cases in each cluster, the distance between the two closest cases between the clusters is evaluated (Everitt and Dunn 1991: 102; Aldenderfer and Blashfield 1984: 38). In complete linkage, the candidate to join the cluster has to be similar enough to all the existing members, not just one (Aldenderfer and Blashfield 1984: 40). In the case of average linkage, there are several ways of calculating it, but basically linkage in this method is based on the average of the similarity of a potential addition to a cluster and the variables already in the cluster (Aldenderfer and Blashfield 1984: 40). Finally, Ward's method aims to produce clusters within which the variance is as small as possible, meaning that the changes in the variance within each cluster after the potential addition are calculated and the addition causing the smallest variation is selected (Aldenderfer and Blashfield 1984: 43).

In the cluster analysis carried out in the present study, a hierarchical procedure and squared Euclidian distance were used, whereas several linkage methods were used and the most suitable one chosen on the basis of the sensibility of the attained solution. Also, several different numbers of clusters were inspected before choosing the one that produced the most logical division

of clusters. Aldenderfer and Blashfield (1984: 21) state that it is not absolutely clear whether data should always be standardized for cluster analysis, but it should be done at least when using units of measurement that differ from each other significantly. Without standardization, variables differing greatly in size and standard deviations could drown out variables smaller in size and standard deviation (Aldenderfer and Blashfield 1984: 26). In the present study, both standardized and raw scores produced the same cluster solution.

The results of the analyses are presented in the next chapter but before moving on to those, the data and the steps leading to its analysis are summarized below in Table 7.

TABLE 7 Stages of data processing.

Data collection
L2 (English) measures: vocabulary (written and spoken production) sentences (written and spoken production) story (written production)
L1 (Finnish) measures: reading comprehension (two factual texts) reading comprehension (two fictional texts)
English nonword repetition test
Finnish nonword repetition test
Preparation
Scoring L2 measures, meaning and structure separately for sentences and story
Scoring nonword repetition tests (four additional raters, inter-rater reliability calculated)
Input
Scaling L2 scores
Forming various composite L2 scores based on scaled scores
Processing
Calculating Spearman and Pearson correlations between L2, L1 and nonword repetition scores
Conducting cluster analysis of L2, L1 and nonword repetition scores

6 RESULTS AND DISCUSSION

The aim of the present study was to investigate the relationship between phonological working memory and L2 knowledge, with particular focus on the effects of less traditional language tasks, task modality and task scoring. The two main research questions were: is there a connection between children's knowledge of an L2 and their success in repeating L2 or L1 nonwords? Is phonological working memory language specific?

In this chapter the results of the statistical analyses are reported and discussed. First, an overview of the participants' scores on the sixth-grade memory and language tasks is provided in the form of descriptive statistics.

6.1 Descriptive Statistics

A summary of the descriptive statistics of the sixth-grade data is provided in Table 8. As apparent from the table, all 15 participants took part in all of the sixth-grade language and memory tasks. The table also serves to clarify the division of the language tasks into different modalities and scores, which in turn give a rough idea of the differences between the tasks. The scores between tasks varied widely because of the wide variation in the content of the items (e.g. word vs. sentence with several clauses), leading to very different maximum scores. The sentence tasks yielded twice the data compared to the vocabulary and story tasks, since both modalities and both assessment methods were applied to this task type.

TABLE 8 Descriptive statistics of the sixth-grade language and memory measures.

	Modality (production)	*Scoring method*	*N*	*k*	*min*	*max*	*mean*	*median*	*s.d.*	*skewness*	*kurtosis*
Vocabulary	*Written (Max = 40)*		15	40	11	36	24.67	25.00	7.41	-0.056	-0.905
	Spoken (Max = 40)		15	40	16	37	25.13	27.00	5.90	-0.063	-0.289
Sentences	*Written*	*Structure (Max = 157)*	15	20	73	154	109.03	104.50	28.12	0.322	-1.327
	Written	*Meaning (Max = 91)*	15	20	49	91	74.60	72.00	13.51	-0.359	-0.986
	Spoken	*Structure (Max = 149)*	15	20	62	143	110.20	112.00	28.96	-0.412	-1.367
	Spoken	*Meaning (Max = 90)*	15	20	45	90	73.60	79.00	15.23	-0.755	-0.847
Story	*Written*	*Structure (Max = 72)*	15		13.5	64	40.43	43.50	16.77	-0.371	-0.932
	Written	*Meaning (Max = 39)*	15		9.0	39	31.20	35.00	8.69	-1.514	1.755
English nonword repetition	*(Max = 18)*		15	18	7	14	11.47	12.00	2.48	-0.623	-1.102
Finnish nonword repetition	*(Max = 21)*		15	21	9	18	13.93	15.00	2.96	-0.210	-1.144

N = number of participants
k = number of items in the measure
min = the lowest number of points scored
max = the highest number of points scored
s.d. = standard deviation

These raw scores were used in assessing the connections of the tasks separately. In addition, various composite scores to be used in the analyses were calculated by scaling the scores and taking means.

The normality of the sixth-grade data was assessed with the Shapiro-Wilk test of normality and by inspecting the skewness and kurtosis of distributions. The only variable where the distribution was clearly not normal was the meaning score of the story task ($p < 0.01$). While there was no obvious ceiling effect, the distribution was clearly negatively skewed (-1.514). Two participants received the maximum score of 39 points and 11 of the 15 participants scored between 31 and 39 points. The task had thus been relatively easy which, in addition to the meaning score being less fine-tuned than the corresponding structure one, led to there being relatively little variation in the scores. As the number of participants in the present study was fairly small, mainly non-parametric statistical tools were used. However, the normality of the majority of the variables allowed the use of some parametric tests as well.

Below, the research questions are addressed in separate sections and the specifics of the statistical analyses used and the results thereof are provided and discussed. A more general discussion of the findings and their relation to previous research as well as their impact on the view of the link between phonological working memory and L2 knowledge is suspended until Chapter 7.

6.2 English Nonword Repetition and Knowledge of English

The first research question concerned the connection between the English nonword repetition test and the English language tasks. In addition to the question of whether such an overall connection exists, more detailed connections were also of interest. The data enabled, for example, investigations into the effect of the modality used in the language tasks and of the effect of the criteria for scoring the language tasks. This complex question is divided up and discussed in the four following sections (6.2.1–6.2.4).

6.2.1 General Connection

The overall connection between phonological working memory as measured with English (L2) nonwords and L2 knowledge is considered first. As there existed a separate meaning and structure score for many of the language tasks, it was reasonable to consider the connection of the nonword repetition test and the structure and meaning scores separately as well.

In order to study overall L2 knowledge, a composite English score had to be formed. This was done by first scaling each of the scores for all of the variations of the L2 tasks (see e.g. Table 8, Descriptive statistics) to the same maximum score and taking their mean. Similarly, the scaled scores were used to calculate the totals of all of the structure and meaning scores to be used in the analyses. These applied to the sentence and story tasks only, since vocabulary

was not scored separately for structure and meaning. Furthermore, the meaning score of the story task was left out of the composite scores used in calculating Pearson r correlations as the distribution was not normal.

In order to investigate the possible connections, Spearman rank-order correlations were calculated. While the non-parametric test is the most suitable for the present number of participants, a lot of information is lost in using ranks instead of the actual scores on the different measures. For that reason, also Pearson r correlations were calculated where possible. The scores on the English nonword repetition test were compared to the composite English score and to the total scores for structure and meaning of the sentence and story tasks. Such comparisons would reveal if there is a connection between nonword repetition and L2 knowledge overall or separately depending on the scoring method. The Spearman rho and Pearson r correlation coefficients are presented in Table 9 below.

TABLE 9 Correlation coefficients for the English nonword repetition test and the composite L2 scores.

	English nonword repetition test	
	Spearman ρ	*Pearson r*
Composite English score	0.804**	0.783**
Structure total	0.772**	0.776**
Meaning total	0.790**	0.793**

** Correlation is significant at the 0.01 level (2-tailed).

The English nonword repetition test and the composite English score were significantly correlated, lending support to the previous findings (e.g. Service and Kohonen 1995 and Dufva and Voeten 1999) that nonword repetition and L2 measures are connected, and indicating a relationship between phonological working memory and language knowledge.

The separate analysis of the structure and meaning scores for some of the language tasks revealed no major differences in the correlation coefficients between nonword repetition and the composite, structure and meaning scores of the English tasks. The correlation between English nonword repetition and the mean of the structure scores on the sentence and story tasks was statistically significant, as was the correlation between the corresponding meaning score and nonword repetition. These results indicate that with respect to the relationship between language knowledge and nonword repetition, the method used to score the language tasks (structure vs. meaning) did not play a major role in whether or not a connection was detected. In other words, repeating L2 nonwords was connected to knowledge of structure and the ability to derive meaning, or command of form and function, of L2.

6.2.2 Different Level Tasks

Since language tasks of different types were used, it was, naturally, of interest to see if the vocabulary, sentence or story tasks were individually connected to nonword repetition. Each task was also investigated in more detail, with separate analyses of the written and spoken performance and the structure and meaning scores, where appropriate.

For this second round of the correlation analyses, three composite scores had to be formed. A composite sentence structure score was formed from the structure scores for the written and spoken sentences, and similarly a composite sentence meaning score was formed from the meaning scores for the written and spoken sentences. Finally, a composite vocabulary score was calculated from the written and spoken vocabulary tasks. Scaled scores were again used in tasks where scores with different maxima were combined, i.e. all but the vocabulary tasks. Hence, the differences did not result in one modality or one scoring method weighing more in the composite score. Since there was no such difference between the maximum scores of the written and spoken components of the vocabulary task, these scores and their sum score were used as such.

The Spearman rank-order correlations and Pearson r correlations for the English nonword repetition test and the different L2 scores are presented in Table 10. The correlations were statistically significant across all the tasks, except for the spoken vocabulary task, which did not show a significant correlation with nonword repetition.

TABLE 10 Correlation coefficients for the English nonword repetition test and the L2 English measures.

	English nonword repetition test	
	Spearman ρ	*Pearson r*
Written vocabulary	0.814**	0.803**
Spoken vocabulary	0.496	0.436
Composite vocabulary	0.731**	0.693**
Spoken sentences, structure score	0.822**	0.839**
Spoken sentences, meaning score	0.724**	0.794**
Written sentences, structure score	0.705**	0.694**
Written sentences, meaning score	0.784**	0.756**
Composite sentence structure score	0.794**	0.798**
Composite sentence meaning score	0.803**	0.793**
Story, structure score	0.723**	0.706**
Story, meaning score	0.626*	-

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

To give an idea of how different the connection between nonword repetition and spoken vocabulary was as compared to the other measures, the difference in the strength of the connection between nonword repetition and written

vocabulary and nonword repetition and spoken vocabulary is shown in the form of scatterplots in Figure 3.

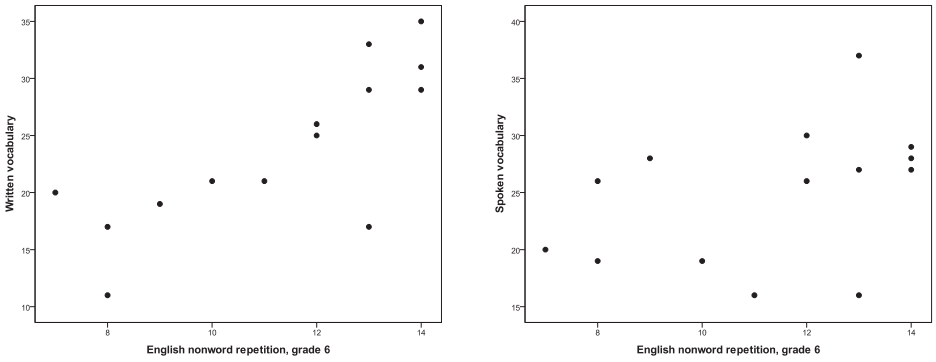

FIGURE 3 Scatterplots of the English nonword repetition test and the written and spoken vocabulary tasks.

Although, on account of the small number of participants, the differences between statistically significant and nonsignificant correlations is not necessarily very large, it appears evident from comparing the scatter plots of the two different vocabulary tasks that the correlation between nonword repetition and spoken vocabulary is clearly weaker than that between written vocabulary and nonword repetition.

The scatterplot with spoken vocabulary also reveals one potential outlier, a participant with one of the lowest spoken vocabulary scores and a relatively high English nonword repetition test score. Removing the outlier brings the correlation between the English nonword repetition test and the spoken vocabulary task to a moderate level: $\rho = 0.610$ and $r = 0.572$ ($p < 0.05$).

For the rest of the correlations (Table 10), the results of the correlation analyses indicated no major differences between the task types, the English nonword repetition test showing a significant correlation with all three task types (vocabulary, sentence and story). The scoring method (structure vs. meaning) produced no variation in the level of detection of a statistically significant connection, while the modality of the task only appeared to play a role in the case of the vocabulary task. After removing one outlier, the difference was diminished, although even before that step was taken, the correlation between nonword repetition and the composite vocabulary score was statistically significant.

In conclusion, there appeared to be a fairly strong positive connection between English nonword repetition and the different measures of English, much as was expected in light of previous studies. Numerous studies have found phonological working memory as measured by nonword repetition tests to be associated with vocabulary knowledge or learning (e.g. Service and

Kohonen 1995). While L2 tasks identical to the sentence and story tasks in the present study have not been used in previous studies, various language tasks have been used in a number of studies and statistically significant connections have been found between many task types and phonological working memory (e.g. Dufva and Voeten 1999; Kormos and Sáfár 2008). It was therefore expected (and found) that this would be the case in the present study as well.

6.2.3 Modality

The role of the modality (spoken vs. written) of the language tasks in their connection to nonword repetition was studied further by forming composite scores of the (scaled) scores of all of the tasks carried out in the written and spoken modalities separately. Fewer scores were involved in forming the composite spoken modality score, since the story task was only conducted in the written modality. The Spearman rank-order correlations and Pearson r correlations for these variables are given in Table 11.

TABLE 11 Correlation coefficients for the English nonword repetition test and the spoken and written L2 measures.

	English nonword repetition test	
	Spearman ρ	*Pearson r*
Spoken English tasks	0.712**	0.780**
Written English tasks	0.772**	0.768**

** Correlation is significant at the 0.01 level (2-tailed).

The nonword repetition test scores showed a significant correlation with both the spoken and written English task scores. This means that the tasks in both modalities were strongly connected to English nonword repetition.

The overall connections do not come as a surprise, as it has already been reported in section 6.2.2 above that the English nonword repetition task was significantly correlated with almost all of the tasks in isolation. What was somewhat surprising, however, is the similarity of the connections between nonword repetition and the two different modalities. It would have made sense if the additional factor common to the nonword repetition and the spoken tasks – spoken production – had shown a stronger association than that between the nonword repetition and the written tasks. The reasons for the similarity can only be speculated on. Possibly, the written modality was more familiar or felt safer to the participants so that they were better able to perform well in writing, thereby cancelling out the otherwise expected advantage of the spoken tasks. The correlation coefficient between the spoken tasks and English nonword repetition changes somewhat if the same outlier, mentioned earlier, is again removed from the analyses, although the participant was not a clear outlier in the other tasks included in the composite spoken task score ($\rho = 0.834$ and $r = 0.893$, $p < 0.01$).

6.2.4 Comparing Grades Five and Six

Studying the actual development of language knowledge or phonological working memory would have required a carefully planned research design where the measures at times 1 and 2 would have enabled the calculation of change in language knowledge or phonological working memory. Without knowing exactly what has changed and how much, it is impossible to estimate such development, much less what might have contributed to or detracted from it.

Therefore, while the present study did not attempt to track the development of phonological working memory or L2 knowledge over time, the existing data were used for simple correlation analyses to compare the fifth- and sixth-grade memory and language tasks. This was done to reveal any possible major shifts in the participants' memory and language scores between the two grades. The descriptive statistics of the fifth-grade tasks are presented in Table 12 and the Spearman rank-order correlations and Pearson r correlations between the L2 tasks and fifth and sixth-grade memory tests are listed in Table 13. Only the composite language scores were used in the analyses. One participant was not available for one of the fifth-grade tasks (Sentence formation 2) and another participant did not take part in the sixth-grade measures, and hence the analyses using the fifth-grade English composite score are based on the results of fourteen participants.

TABLE 12 Descriptive statistics of the fifth-grade measures.

	N	*k*	*min*	*max*	*mean*	*median*	*s.d.*	*skewness*	*kurtosis*
Vocabulary *(Max = 60)*	16	60	18.00	52.00	32.25	30.00	10.34	0.461	-1.016
Listening comprehension *(Max = 10)*	16	10	2.00	10.00	6.56	7.50	2.34	-0.502	-0.747
Reading comprehension (Max = 10)	16	10	0.50	10.00	6.34	6.88	2.90	-0.700	--0.387
Verb form task 1 *(Max = 10)*	16	10	3.50	10.00	6.88	7.00	1.75	-0.351	-0.060
Verb form task 2 *(Max = 10)*	16	10	1.00	9.50	5.56	6.00	2.41	-0.489	0.077
Verb form task 3 *(Max = 10)*	16	10	4.50	10.00	7.41	7.50	1.85	-0.109	-1.348
Sentence formation 1 *(Max = 40)*	16	8	5.50	37.50	22.25	22.50	9.37	-0.322	-0.406
Sentence formation 2 *(Max = 60)*	15	10	18.00	58.00	42.47	45.00	12.85	-0.541	-0.763
English nonword repetition *(Max = 18)*	16	18	2.00	9.00	5.75	6.00	2.08	-0.127	-0.639

N = number of participants
k = number of items in the measure
min = the lowest number of points scored
max = the highest number of points scored
s.d. = standard deviation

TABLE 13 Correlation coefficients for the English nonword repetition tests and English (L2) measures in grades five and six.

	English nonword repetition test, grade five		*L2 tasks, grade five*	
	Spearman ρ	*Pearson r*	*Spearman ρ*	*Pearson r*
English nonword repetition test, grade six	0.623*	0.503	0.722**	0.643*
L2 tasks, grade six	0.723**	0.677**	0.881**	0.883**

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Based on the correlations in Table 13 it is apparent that no dramatic changes had taken place. The only coefficient that stands out is the Pearson r correlation for the two English nonword repetition tests, which only approaches significance ($r = 0.503$, $p > 0.05$). This is interesting because the two nonword repetition tests were almost identical and the participants were at an age where major changes in phonological working memory capacity were no longer expected. One possibility is that changes in long-term knowledge of English are reflected in the absence of a strong correlation with the nonword repetition scores. If there has been uneven progress in English proficiency within the group of participants during the year between the tests, and if their language knowledge affects their success in the English nonword repetition test, this could be reflected as a lower correlation between the two nonword tests. The strong correlation between the language scores in the two years does not support this hypothesis, but this evidence is not conclusive as the language measures used in the two years were very different. Nevertheless, the Spearman rank-order correlation for the two memory tasks is statistically significant and, again, outliers are revealed in the corresponding scatterplot (Figure 4). Removing the most extreme one, however, yields a correlation between the two English nonword repetition tests of $\rho = 0.684$ ($p < 0.01$) and $r = 0.612$ ($p < 0.05$).

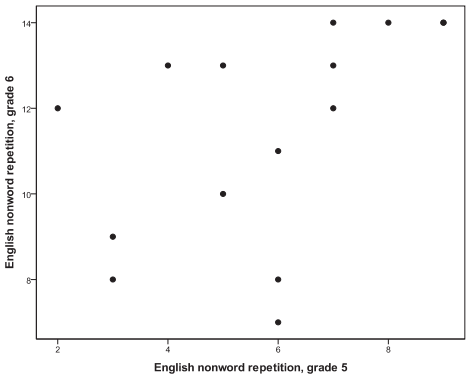

FIGURE 4 Scatterplot of the English nonword repetition tests in grades five and six.

All the participants received a higher score in the test in the sixth grade, but the outlier presents the greatest change, from two points in grade five to 12 points in grade six out of the 18-point maximum in each test. The reasons for such a difference can only be speculated on. Most likely, the participant had problems with the fifth-grade nonword repetition test that were not due to phonological working memory, as such an improvement over one year is not likely to be due to the development of the phonological working memory system only.

It bears repeating that the measures used in the present study did not allow for the detection of development in phonological working memory or L2 knowledge from grade five to grade six; however, based on the correlations in Table 13, it can be said fairly safely that whatever development there was in either, it was similar enough across all participants not to cause major changes in the significance of the correlations between the different tasks.

It should also be noted here that with the composite scores such as the total of the sixth-grade L2 tasks given above, it is perhaps wise to proceed with caution. It can be argued that the composite scores entail repetition. For example, while technically the separate scores for the structure and meaning of sentences could be very different, this is not the case in the present study. The correlations between the meaning and structure scores in the sentence tasks and the story are very high, reflecting the fact that those performing well on structure also did so on meaning and vice versa. A similar relationship exists between the written and spoken sentences. In the present data, there are no great differences overall between the participants' performance on the written and spoken versions of the sentence task, although the correlations are not as high as between the structure and meaning scores for the exact same task.

6.3 Language Specificity of Phonological Working Memory

The second research question dealt with the possible language specificity of phonological working memory and the language-dependency of the nonword repetition test. In order to answer the second research question, the nonword repetition test was also carried out in Finnish in the sixth grade. As evident from Table 8 (Descriptive statistics) the test yielded interpretable results, unlike in some previous studies, and could be used for statistical analyses.

First and foremost, the two sixth-grade nonword repetition tests, the Finnish and English one, were not statistically significantly correlated ($\rho = 0.385$ and $r = 0.414$, $p > 0.05$), indicating that success in one only moderately predicts success in the other. This could be considered as indication that the language of the nonword repetition tests makes a difference in the assessment of phonological working memory. In addition, there was no statistically significant correlation between the Finnish reading tasks and the English language measures.

The first step in studying language specificity is a general enquiry into a possible connection between a phonological memory task based on the L1 of

the participants and measures of L2. If the nonword repetition test is very closely tied to the participant's knowledge in that particular language it should not necessarily have connections to a completely different language. To study whether the Finnish nonword repetition test was connected to the English language tasks, in essence the same correlation analyses were carried out using the Finnish test as its English counterpart. The Finnish nonword repetition test was compared to the composite English score, the different task types separately (vocabulary, sentences and story), the different modalities (written and spoken tasks), and the different scoring methods (structure and meaning). All these Spearman rank-order correlation coefficients and Pearson r correlations are provided in Table 14 along with the corresponding English nonword repetition test correlation coefficients for easy comparison.

TABLE 14 Correlation coefficients for the Finnish and the English nonword repetition tests and L2 language measures in grade six.

L2 task	*Finnish nonword repetition test*		*English nonword repetition test*
	Spearman ρ	*Pearson r*	*Spearman ρ*
Written vocabulary	0.642**	0.639*	0.814**
Spoken vocabulary	0.247	0.307	0.496
Composite vocabulary	0.510	0.533*	0.731**
Spoken sentences, structure score	0.645**	0.661**	0.822**
Spoken sentences, meaning score	0.513	0.596*	0.724**
Written sentences, structure score	0.547*	0.590*	0.705**
Written sentences, meaning score	0.531*	0.578*	0.784**
Composite sentence structure score	0.664**	0.650**	0.794**
Composite sentence meaning score	0.586*	0.600*	0.803**
Story, structure score	0.583*	0.652**	0.723**
Story, meaning score	0.419	-	0.626*
Structure total (sentences and story)	0.656**	0.663**	0.772**
Meaning total (sentences and story)	0.511	-	0.790**
Spoken English tasks	0.642**	0.593*	0.712**
Written English tasks	0.571*	0.646**	0.772**
Composite English score	0.613*	0.632*	0.804**

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

As can be seen in Table 14, the Spearman rank-order correlation was not statistically significant between the Finnish nonword repetition test and several of the English language measures, namely spoken vocabulary, the composite vocabulary score, and the meaning scores for spoken sentences and the story, as well as the composite meaning score. Using Pearson r correlations, the results were somewhat different, as all the correlations, except that for the spoken vocabulary task, were at least moderate. Even the correlation (Pearson r only) between Finnish nonword repetition and the spoken vocabulary task was

statistically significant ($r = 0.561$, $p < 0.05$), when yet another outlier was removed (Figure 5). The outlier had one of the highest scores on the Finnish nonword repetition test and one of the lowest scores on the spoken vocabulary task. Removing the participant also changed the correlation coefficients for the composite vocabulary score to $\rho = 0.721$ and $r = 0.725$ ($p < 0.01$). For the other measures there were no outliers that would have made a significant difference to the correlation coefficients.

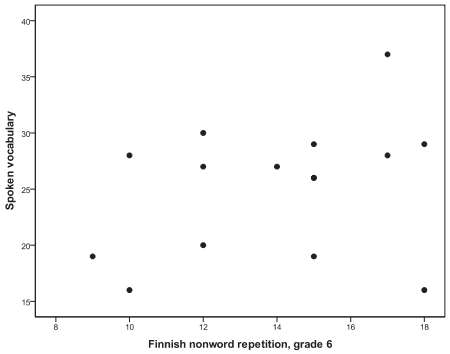

FIGURE 5 Scatterplot of the Finnish nonword repetition test and the spoken vocabulary task.

The persistently weak or moderate connection between Finnish nonword repetition and the spoken vocabulary task is an interesting finding because – as has been made abundantly clear in section 4.2 above – vocabulary is the area of language knowledge that has been studied the most in connection with phonological working memory and which has most often been found to have a connection to phonological working memory, so much so that some studies indicate that even connections to other areas of language knowledge are also, at their core, due to a link between vocabulary and memory (e.g. Service and Kohonen 1995). This link is not highly evident in the results of the present study. The link between written vocabulary and Finnish nonword repetition is strong, but it is difficult to say much about it owing to the spoken vocabulary task behaving so differently in this respect. Whatever the reason for this result, it would be difficult to say on the basis of the Finnish nonword repetition test how well participants would be expected to perform in a spoken vocabulary task such as the one used here.

Clearer positive correlations between the Finnish nonword repetition test and English language measures were also found. As measured with both the Spearman rho and Pearson r correlations, the Finnish nonword repetition test correlated significantly ($p < 0.01$) with the structure score of the spoken sentence task and with the composite structure scores for sentences and all tasks combined. The Pearson r correlation between the Finnish nonword repetition test and the structure score of the story was also fairly strong. In fact,

the connection between the Finnish nonword repetition test and the story task appears very different, depending on what the score is based on. There is either a moderate or stronger correlation for the structure score ($\rho = 0.583$, $p < 0.05$ and $r = 652$, $p < 0.01$), but no significant correlation for the meaning score ($\rho = 0.419$, $p > 0.05$).

Judging from these results, it seems that the scoring method may have some influence: the strongest correlations were connected to the structure scores. However, not all the individual structure scores stand out like this. The structure score for the written sentences shows only a moderate link to Finnish nonword repetition ($\rho = 0.547$ and $r = 0.590$, $p < 0.05$).

The rest of the tasks – with different modalities (written and spoken) and scoring perspectives (structure and meaning) taken into account – showed mostly a moderate correlation with the Finnish nonword repetition test. It is difficult to say whether the correlations are strong enough to have any real bearing on the issue, given the number of participants. For the majority of the tasks, however, it appears that the ability to repeat Finnish (L1) nonwords has something to do with performance in English (L2) tasks.

In conclusion, what is obvious in Table 14 is that the correlation coefficients between the Finnish nonword repetition test and the L2 tasks, even though statistically significant, are clearly weaker than those for the English nonword repetition test. It seems thus likely that the language of the nonword repetition test may make a difference to whether strong or weak connections are found between phonological working memory and language knowledge; however, phonological working memory is not clearly language-specific.

6.3.1 The Finnish Reading Task

So far the Finnish nonword repetition test scores have only been subjected to correlation analyses with the English (L2) task scores. The investigation into the language specificity of phonological working memory would not be complete without bringing in the measures of the Finnish (L1) language as well. In the present study, the Finnish tasks were composed of four reading comprehension tasks; these are described in more detail in section 5.6 above. The correlation coefficients used to study the connection between Finnish nonword repetition and Finnish reading are given below in Table 15.

TABLE 15 Correlation coefficients for the Finnish nonword repetition test and the L1 measures.

Finnish reading test, grade 6	*Finnish nonword repetition test, grade 6, all*		*Finnish nonword repetition test, grade 6, correlations without the outlier*	
	Spearman ρ	*Pearson r*	*Spearman ρ*	*Pearson r*
total	0.538*	0.578*	0.791**	0.779**
factual1	0.445	0.462	0.711**	0.723**
factual2	0.464	0.458	0.571*	0.545*
factual total	0.487	0.498	0.643*	0.657*
story1	0.649**	0.674**	0.822**	0.807**
story2	0.406	0.426	0.692**	0.674**
story total	0.578*	0.627*	0.868**	0.861**

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Looking at the regular correlation coefficients in the first two columns, it first seems that the connections between the Finnish nonword repetition test and the Finnish language tasks are positive but mostly moderate to low. The clearest statistically significant connection is that of the first story task ($\rho = 0.649$ and $r = 0.647$, $p < 0.01$). However, after excluding one extreme outlier from the analysis, the situation changes considerably, as can be seen in the two rightmost columns of Table 15. Now, a statistically significant correlation emerges between Finnish nonword repetition and all of the Finnish language measures.

This difference in correlations is a good reminder of the effect of the small number of participants in the present task and how one extreme case can prevent an otherwise solid connection from appearing. It should be noted, however, that the outlier removed here was indeed very extreme. This participant had the highest score on the Finnish nonword repetition and the lowest score on the Finnish language tasks (the total score, the first factual text and the second story text). Due to the extreme nature of this one case, it was deemed appropriate to run the analyses without the oddity and to consider the latter correlations a more accurate description of the situation.

It can thus be concluded that the overall connection between Finnish nonword repetition and the Finnish reading tasks in grade six was fairly strong and, for the most part, the participants who were successful at repeating L1 nonwords also did well in L1 reading comprehension tasks, while, similarly, a poor result on an L1 phonological working memory measure went together with poor reading comprehension in L1. Taken together with the fact that the Finnish nonword repetition also correlated with the L2 tasks, this indicates that success in an L1 nonword repetition test is not merely a reflection of knowledge of the language on which the nonwords are based but possibly reflects a more general processing ability which is connected to knowledge of both the L1 and L2. Whether the same is true for the English nonword repetition test was studied next.

To complete the correlative analyses, the English (L2) nonword repetition test scores were correlated with the Finnish (L1) reading comprehension scores. These correlation coefficients are presented in Table 16.

TABLE 16 Correlation coefficients for the English nonword repetition test and the L1 measures.

Finnish reading test, grade 6	*English nonword repetition test, grade 6*		*English nonword repetition test, grade 6, correlations without the outlier*	
	Spearman ρ	*Pearson r*	*Spearman ρ*	*Pearson r*
total	0.376	0.358	0.551*	0.481
factual1	0.276	0.317	0.458	0.488
factual2	0.494	0.413	0.564*	0.467
factual total	0.462	0.408	0.551*	0.512
story1	0.149	0.234	0.249	0.297
story2	0.222	0.212	0.420	0.366
story total	0.258	0.257	0.436	0.387

* Correlation is significant at the 0.05 level (2-tailed).

The rightmost columns of the table again show the coefficients calculated with only 14 participants. The same exceptional participant, the oddity discussed above needed to be excluded from this set of analyses as well due to a top score on the English nonword repetition test paired with a very low score on the Finnish tasks (Figure 6). There was another possible outlier but removing that participant from the analysis did not make such a significant difference to the correlations obtained.

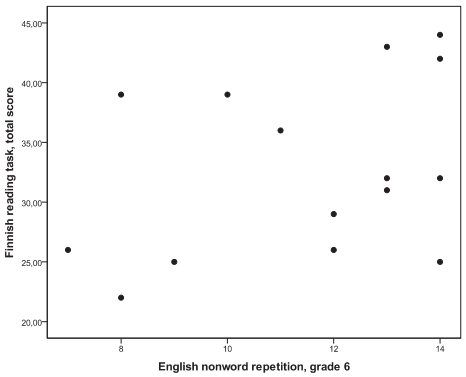

FIGURE 6 Scatterplot of the English nonword repetition test and the Finnish reading task.

This time the effect of disregarding the outlier was not so dramatic. While originally no significant correlations were found between English nonword repetition and the Finnish reading measures, the correlation coefficient of the second factual text was moderately significant ($\rho = 0.564$, $p < 0.05$) when the

deviant participant's scores were excluded from the analysis. This task is then also reflected moderate correlation between the English nonword repetition test and the total score for the factual tasks and the total for all of the four measures ($\rho = 0.551$ and $\rho = 0.551$, $p < 0.05$ respectively). The Pearson r correlation coefficients did not reach statistical significance for any of the Finnish reading comprehension tasks.

In sum, when the most exceptional participant was left out of the analysis, the connection between English nonword repetition and the Finnish reading comprehension tasks was moderate at best and weak for the majority of the tasks. This means that based on the results of the present study nothing very certain can be said about success on an L1 reading comprehension task based on one's ability to repeat L2 nonwords. The tasks appear somewhat related, but to make predictions would be risky.

6.3.2 Individual Contributions of the Nonword Repetition Tests

Based on the findings reported above, it seems evident that there are some differences in the relationships found between nonword repetition and language knowledge depending on the language of the nonword repetition test and the language assessed. Both nonword repetition tests correlated – albeit on somewhat different levels – with the L2 measures and while the Finnish nonword repetition test was strongly correlated with Finnish reading tasks, the English nonword repetition test was not. To further study the contributions of the two tests, the correlations between each test and the L2 and L1 tasks were calculated again with the influence of the other nonword repetition test partialled out. The statistically significant correlations between English nonword repetition and L2 knowledge did not disappear after partialling out the Finnish nonword repetition test scores. All the English measures remained at least moderately correlated with English nonword repetition after Finnish nonword repetition was partialled out. Similarly, partialling out English nonword repetition test scores from the correlation between Finnish nonword repetition and the L2 English measures caused no major changes in the correlation coefficients. Only the correlation coefficient of spoken vocabulary was diminished enough to be no longer of statistical significance when compared to the moderate correlation before elimination of the influence of English nonword repetition. All the other English measures remained at least moderately correlated with Finnish nonword repetition after partialling out English nonword repetition.

The statistically significant correlation between Finnish nonword repetition and Finnish reading remained very similar, also after English nonword repetition was partialled out. Only the correlation coefficient of the second factual text diminished enough for the correlation to become nonsignificant, but this was in any case the smallest correlation coefficient between Finnish nonword repetition and the Finnish reading tasks to begin with ($r = 0.545$, $p < 0.05$). Both the total score of the factual texts and the total reading score remained statistically significantly correlated with Finnish

nonword repetition. English nonword repetition and Finnish reading were not sufficiently significantly correlated enough to warrant calculating partial correlations ($p > 0.05$ for all Pearson r correlations).

These results would indicate that there are independent contributions from each nonword repetition test and that they do not tap the same exact construct, or at least are not interchangeable measures of it. If they both measured the same construct they would have correlated with each other and, most likely, partialling out one would have eliminated the correlations between the language measures and the other nonword repetition test as well. The fact that the nonword repetition test and the language measures in each language remain correlated after partialling out the nonword repetition test in the other language could be taken as evidence of language specificity. The language-specific effect of phonological working memory on L2 knowledge remains and is statistically significant in both cases. However, as Finnish nonword repetition also remains correlated with English measures even after English nonword repetition is partialled out, it would appear that the Finnish nonword repetition test has to access a language-general phonological working memory capacity. Still, the small sample size has to be taken into account and conclusions should therefore be drawn with caution.

6.3.3 Systematic and Language-Specific Individual Differences

So far the language specificity of phonological working memory has been investigated only by correlating pairs of measures. In search of a clearer answer to the second research question regarding the language specificity of phonological working memory, a more comprehensive picture of all the sixth-grade measures was needed. The measures were thus entered into a cluster analysis of cases in an attempt to see if any such groups of participants would be formed that could reveal something of the language specificity issue. Basically this would mean the formation of clusters around the two languages, English (L2) and Finnish (L1). The clustering revealed some interesting, somewhat exceptional cases among the participants. These will be discussed in their own section below, but first the groups and the findings on language specificity are discussed.

Cluster analysis was another way of studying the connections between the four major elements here: L2 English and L1 Finnish nonword repetition and the L1 and L2 language measures. The focus thus broadened to include the complete range of measures at once while at the same time enabling the consideration of each individual participant and their profiles through all the sixth-grade tasks.

6.3.3.1 The Language-Memory Groups

All the sixth-grade measures were used in the cluster analysis: the English and Finnish nonword repetition tests, written and spoken English vocabulary, written and spoken English sentence tasks (both structure and meaning scores

separately) and the English story task (structure and meaning scores separately), and the four different Finnish reading comprehension tasks. The raw scores on each measure were used in the analysis.

The aim was to investigate the whole range of scores for each participant to find possible patterns in links between tasks and gain a broader view of the interaction between phonological working memory and language in both of the two languages by going beyond correlating pairs of measures. The solutions of the cluster analyses also served to pinpoint interesting individuals whose profiles were then looked at more closely in an attempt to deepen the understanding of the relationship between phonological working memory and L2 knowledge.

The linkage method that produced the best possible cluster solution on the sixth-grade tasks was average linkage (the between-groups method). The within-groups method produced the exact same solution. Further, the most logical and reasonable option was that of four clusters. As discussed in section 5.7, cluster analysis does not produce a single correct solution but the usefulness of a cluster solution has to be assessed based on the content of the clusters being interpretable.

The clusters can be seen in Table 17, where the figures are the participants' rankings for the two phonological working memory tasks, the composite score of the English language tasks and the total score of the four Finnish language tasks. The rankings were not used in the analysis itself but are only used here to facilitate the interpretation and comparison of the clusters. The same table with the participants' scores instead of rankings is given in Appendix 7.

Four groups for such a small sample may seem excessive, and it certainly would be if the purpose was to make general claims solely based on the groups found. However, the cluster analysis here serves as a descriptive tool to further inspect the language specificity of phonological working memory, an issue already studied using other means. In addition, it was a tool used to form participant profiles and aided in pinpointing somehow interesting individuals for further investigation.

TABLE 17 The language-memory clusters described with participants' rankings for the tasks.

	Participant	*English nonword repetition*	*Finnish nonword repetition*	*English language tasks*	*Finnish language tasks*
Cluster A	1	2.5	3.5	3	1
	2	2.5	6.5	4	3
	3	6	3.5	1	2
Cluster B	4	13.5	15	15	15
	5	12	13.5	13	13.5
	6	15	11	11	11.5
	7	6	13.5	14	9
Cluster C	8	10	1.5	10	6
	9	11	6.5	9	4.5
	10	13.5	6.5	12	4.5
Cluster D	11	8.5	6.5	8	11.5
	12	8.5	11	7	10
	13	2.5	11	5	7.5
	14	6	9	6	7.5
	15	2.5	1.5	2	13.5

As there were only 15 participants to be divided into clusters they were understandably fairly small: two clusters of three, one cluster of four, and one cluster of five participants. The clusters are described in detail below.

Cluster analysis was also carried out using the composite or total scores only. It is technically possible that having more measures of one construct leads to it having more weight in the analysis over other constructs with fewer measures and eventually to a skewed cluster solution (see e.g. Aldenderfer and Blashfield 1984: 21). In the present study, however, the cluster solution was the same regardless of whether all the individual language task scores were used or whether only the composite/total scores were used. Apparently, then, the difference in numbers of variables was not significant, at least not sufficiently to have affected the cluster solution.

Cluster A. The three participants in this cluster scored relatively well on all the measures. They were among the best third on almost every measure. In the English nonword repetition test, they scored highest or second highest (several participants had the same score) and in the corresponding Finnish task they scored second and third highest. Consideration of the rankings based on the raw composite score of the English tasks revealed that the participants in this cluster held positions one, three and four. With respect to the total score for the four Finnish language tasks, they ranked highest, occupying positions one, two and three. Overall, then, the participants in this cluster were the top group which succeeded well in all the tasks. Only two single scores were closer to the

average performance and those are the meaning score of participant 1 for the spoken sentence task (ranked 7/15) and participant 2's score on the first Finnish factual reading task (ly6t1) (also ranked 7/15).

Cluster B. The opposite of Cluster A is a cluster of four participants who did relatively poorly on all the measures. Their scores were among those in the poorest third for almost all the measures. On the English tasks (composite score), they ranked 11, 13, 14 and 15 and on the Finnish nonword repetition test they also obtained the three lowest scores (shared positions for some of the scores). Three out of the four participants were also placed low on the English nonword repetition test and the Finnish language tasks (total score), ranking, respectively, at 12, 13 and 15 and 11, 13 and 15. The exception is participant 7, who scored fairly well on these tasks ranking sixth (shared) on the English nonword repetition test and ninth on the Finnish language tasks. On the Finnish language tasks his ranking varied from 5 to 13.

On some single tasks the participants performed on a level closer to the average, but not as exceptionally as in the case of participant 7. One case that perhaps deserves to be mentioned here is participant 5, who ranked fifth, in the top third (shared position), on the spoken vocabulary task. This is noteworthy as this participant's overall ranking on the English tasks was 13/15.

Cluster C. This could be named the Finnish cluster, as the three participants obtained scores that were among the highest five scores on both the Finnish nonword repetition test and the Finnish language tasks. In English they were less successful. All their scores on the English language tasks and the English nonword repetition test were among the bottom third or below average at best. The best ranking is eighth place on the meaning score for written sentences occupied by participant 9. There are no particularly unusual cases in this group; instead, all could be described as performing well on Finnish tasks, no matter whether memory or language, and not very well on the English ones.

Cluster D. The fourth cluster is the largest one and also the least homogenous of the four; however, the participants scored fairly well on the English measures. They either ranked in the top third or were about average on the English measures, both memory and language tasks. In the Finnish tasks, however, the participants' performance varied greatly. Based on the total scores of the Finnish language tasks, three of the five participants in this cluster were about average while two were located in the poorest third of the whole group. Two scores on individual tasks were in the top third, however: participant 13 on the first factual text and participant 14 on the second factual text. The Finnish nonword repetition task showed even more variation. Participants 12 and 13 obtained the same score and were ranked 11th. Participant 14 ranked in the middle range, in ninth position. Participant 11 obtained the third best score but shares this with several other people, and participant 15 obtained the top score and also shares that with another person. In other words, while the participants

in Cluster D performed successfully on the English measures, they are not very alike when it comes to the Finnish nonword repetition test scores. Whereas the participants in Cluster C were good at Finnish nonword repetition and poor at the corresponding English task, there is no such neat distribution in Cluster D.

In sum, based on the cluster analysis, the participants seem to fall into four fairly distinct categories. There are participants who excel at everything or nothing, and those whose success appears to depend on the language of the task. Theoretically, mixed clusters would have been possible so that the division would not have appeared to be so clearly language-related. For example, there were numerous English language measures, and if the clusters had been formed around a division in them, the participants within a cluster might have varied greatly in their success in all the other measures.

As it is, we can at least say that Cluster C could be seen as an indicator of language specificity. The participants in this cluster are systematically (in both memory and language tasks) good in one and poor in the other language, in this case good in Finnish and poor in English. This is not something that Clusters A and B can help us estimate, as the participants in them were either good or poor in all tasks in both languages. Based on their performance it is impossible to say anything about possible connections with phonological working memory because there is no difference between the two languages. Cluster D falls somewhere in between. It seems to support the idea of language specificity in the sense that the participants do well on both the English nonword repetition task and English language tasks but poorly on the Finnish language tasks. However, in this cluster, the scores of the participants on the Finnish nonword repetition test are so varied that they could either support or disprove any indication of language specificity.

To further elaborate on the complex connection between success on the nonword repetition test and L2 measures, two written tasks by three participants are given in Table 19 below. The first two participants, 3 and 7 obtained the same score for the English nonword repetition test (13/18), but in L2 knowledge participant 3 is the best of the whole group (ranks first) while participant 7 is very poor (ranks 14th). This is clear from their responses in the written sentence repetition task and the story task (Table 19). The third example, participant 4, on the other hand, is roughly on the same level in L2 knowledge as participant 7 (ranks 15th), but is considerably poorer on English nonword repetition (8/18 points). The comparison of these three participants demonstrates well the fact that, while on a more general level the connection between scores on the nonword repetition test and L2 measures is clear, on the level of individuals, similar success on a nonword repetition test does not always guarantee similar success on the L2 measures and, vice versa, corresponding L2 knowledge is not necessarily a sign of a similar nonword repetition outcome. The participants' exact scores on the example tasks are given in Table 18. Because these cases are participants from the ends of the L2 knowledge spectrum, two further examples are given in Table 19 to show what

kind of language the mid-range participants produced. The participants included in the example are participants 12 and 11, who ranked seventh and eight out of the 15 participants on the L2 measures, and both are in Cluster D.

The individuals who differed from their cluster-mates are worth a closer look as they can provide valuable information about the nature of the connection between phonological working memory and language knowledge. We already know that a statistically significant connection exists here so that in most cases the profiles of the participants can be predicted in that good nonword repetition success will generally be accompanied by good language knowledge and vice versa, but that does not have to mean that this is the case for every individual, not even in a fairly small sample such as the present one. Below the atypical or exceptional individuals in the above cluster analysis are discussed in more detail.

TABLE 18 Examples of the written sentence repetition task and the story task for three participants.

Participant 3 (L2 overall ranking 1)	**Participant 7 (L2 overall ranking 14)**	**Participant 4 (L2 overall ranking 15)**
Sentences (written production) structure score: 154/157 points, ranking 1, propositional score : 91/91 points, ranking 1 (shared)	**Sentences (written production)** structure score: 73/157 points, ranking 14 (shared), propositional score : 64/91 points, ranking 12	**Sentences (written production)** structure score: 73/157 points, ranking 14 (shared) propositional score : 49/91 points, ranking 15
Tom likes ice cream a lot, but he doesn't eat it at school. They are washing the dishes, but they aren't cleaning the whole house. The boys in this class are going to cinema tomorrow. He does his job very well, and he doesn't waste any time. I have got two younger sisters, so I'm the oldest. The girls in my group walked to school yesterday. I am reading a book and my friend is watching TV. She has got three older brothers, so she is the youngest. They do their homework, but they don't enjoy it very much. I love movies, but I didn't see movie about the dog.	Tom like ice sgreeam alot put He dont et it a school The are whashing adishes put heyaren ot cleaning a house to Boy is this glas going asinema tomorrov Shes Does shes jop I have got g two young sisten put im a oldest the gorbs in my groop walking a shool yesterday im Freeding a Boog and my Friend watcing tv she has got tree old Brothers shes a youngest they Do the homevork But they enjoy that -	Tom likes a icecream really must but his didn't likes gomes a school. Theyre a washing windows but theyre didn't cleans a car. A boys in a school and gomes a kino tomorrow. Hi doesn't jaxt very well but she doesn't likes this. I have got two sisters and then i am a old A carols im my coorps. And I'm a school yesterday. I'm reading a book and my Frend looked a tv She is got three brothers and she is a young Their two a homework but she didn't have good that. I love a moves. But i didn't see a dog
Story structure score: 64/72 points, ranking 1 (shared) propositional score : 39/39 points, ranking 1 (shared)	**Story** structure score: 13.5/72 points, ranking 15 propositional score : 20/39 points, ranking 13 (shared)	**Story** structure score: 16/72 points, ranking 13 propositional score : 20/39 points, ranking 13 (shared)
It was Saturday morning. Mary and her brother Bill went to the zoo. There was a big lion in a cage. He looked very hungry and sad. "Can I give him my apple?" asked Mary. "Sure, why not?" said Bill. So, Mary gave her apple to lion. That make lion very happy.	its wat sator day in a out he lähtivät ther a Big laion in a Häkissä its a hungry and litle sad karina ask voinko antaa omenani sille Yes its a happy noww	Mary go a zoo Saturday. And she brothers whit Mary is hungry. Then you go to zoo. An animal is very big and strange. Mary cats a lion apple and Maryn brother was a tiger whit. Then Mary say: my knee is hurt. Mary eats a lion and lion is very habby.

TABLE 19 Examples of the written sentence repetition task and the story task for two mid-range participants.

Participant 12 (L2 overall ranking 7)	**Participant 11 (L2 overall ranking 8)**
Sentences (written production) structure score: 104.5/157 points, ranking 8, propositional score : 79/91 points, ranking 7	**Sentences (written production)** structure score: 113.5/157 points, ranking 6, propositional score : 85/91 points, ranking 6
Tom likes ice gream alot. But he dose'nt eat that in sckool The are woshing the dicis. but they arent clean the hol haus They boy's an the glass are going in the seminen He das hes job very vel and he dosent vastin time I have got two jang sister so I am the oldest The girls in grup walk in the school yesterday I'am reading a book and my friend wach TV. She has got trhe old prothers so she ist the jangest They do the homework but the dont joy that very mutch I love moves. But I didint saw the move ver is the dog	Tom's likes Ice-gream a lot, but he doesen't eat it at school. They cleaning the dishesh, but they don't all hause? The boys in they class go cinema tomorrow? he does he job very well and he don't weist time? I have got two young sisters, so I am oldest The girl's in my groub walk's school yesterday? I am reading a book and my friend watchin tv She has got three oldest brother so she ist the youngest They do they homeworks but they don't enjou that very mutch I love movies but i did see movie about dog
Story structure score: 43.5/72 points, ranking 8 propositional score : 36/39 points, ranking 5 (shared)	**Story** structure score: 46/72 points, ranking 7 propositional score : 36/39 points, ranking 5 (shared)
It was a sadurday morning. Mery and her prother Bill went the Zoo. There was a big lion and its looks hungry and sad. -Chan I give my aple for lion? ask Mary her prother. -Sure. Why not. So Mary give a aple to lion and make's lion happy.	it was Saturday morning. Mary and her brother Bill go zoo. There was big lion in gage. it looks little hungry and sad. -can I give my apple, Mary ask -sure why not, says Bill. Her give an apple and lion was happy.

6.3.3.2 Interesting Individuals

The perhaps most strikingly exceptional case and thus a very interesting individual, is participant 15, who was also identified as an outlier in section 6.3.1. He belongs to Cluster D, where he fits very well because of his success in the English tasks while his performance on the Finnish tasks is rather mixed. While he has the highest score (shared) in the Finnish nonword repetition test, his poor performance on the Finnish language tasks earned him the second to worst total score. Participant 15 is thus consistently excellent at everything else – ranking second both on the English nonword repetition test and the composite score of the English language tasks – but the Finnish reading comprehension tasks were clearly problematic for him.

Participant 7 is another interesting, contradictory case and also an outlier, discussed in section 6.2.2. He was placed in Cluster B where the participants did not do well on any particular task, although the performance of participant 7 on the different tasks was actually quite varied. He differs from the other participants in his cluster in that he performed relatively well on the English nonword repetition test. He is, however, second to last on the English language tasks. Just about the opposite is true for the Finnish tasks: his score on the Finnish nonword repetition test is the second to lowest while his performance on the Finnish language tasks is about average. Granted, the differences between being poor or average or good or average are not that great in such a small sample, but it is nevertheless interesting to see such a clear example of how the language of the memory task undeniably makes a difference but does not necessarily go hand in hand with the corresponding performance on the language tasks in the same language.

In particular, the fairly good English nonword repetition test score paired with very poor performance on the English language tasks raises the question whether there is perhaps something peculiar in the skills of participant 7 or his attitude towards language tasks. Very wide-reaching conclusions obviously cannot be made on the basis of the present study but, at any rate, the success or lack thereof of participant 7 in the English language tasks most likely does not hinge upon the ability to repeat English nonwords.

The other interesting individuals showed more minor discrepancies in their scores on the measures. Participants 5 and 12 stand out because a sizable difference between their scores on the written and spoken vocabulary tasks. Both got a fairly good score on the spoken vocabulary task – participant 12 ranked second and participant 5 fifth – but did not do as well on the written task, with participant 12 landing exactly in the middle by being ranked eighth and participant 5 being ranked 13th. The other participants, no matter what cluster they belong to, do not vary as much in their scores on the two vocabulary tasks.

A closer look at the vocabulary tasks shows that there was no real difference between them, the average score on the written task being 24.7 and the corresponding figure for the spoken task being 25.1. However, two more

participants obtained a better score on the written than the spoken vocabulary task, contrary to what would have been expected based on the scores of participants 12 and 5, and two other people scored exactly the same on both vocabulary tasks. There are also no major differences between clusters: all the clusters contained participants who did better on the written vocabulary task and those who scored higher on the spoken version. However, participant 12 is the only one in Cluster D who scored better on the spoken task. All in all, the surprising scores of participants 12 and 5 do not seem to be indicative of any systematic difference between the two vocabulary tasks; however, on the individual level, both are clearly better on the spoken vocabulary task, and in the case of participant 5, this task stands out in his otherwise poor English performance.

In the other language tasks, such great differences did not emerge between either the spoken and written version or the meaning and structure scores. In the phonological memory measure, however, two participants stand out. As already discussed above, the three members of Cluster C did considerably better in the Finnish nonword test than the English one, as that was a factor in their assignment to that particular cluster. However, one of the members in that cluster, participant 8, is a very extreme case. Also noted as an outlier in section 6.3, participant 8 has the top score in Finnish nonword repetition but ranks 10th on the English nonword repetition test. Participant 13, a member of Cluster D, similarly shows a large difference between his nonword repetition scores, but in the opposite direction from that of participant 8. He ranked second on the English nonword test and 11th on the Finnish one.

So both participants 8 and 13 fit very nicely into their respective clusters, but are also the extreme cases in them, especially when it comes to their memory task scores. None of the other participants shows that great a difference in the phonological working memory assessment, depending on the language of the nonword test. Even with the other evidence already presented, participants 8 and 13 seem to be the shining examples of the language specificity of phonological working memory. With them there is no question regarding the effect of the language of the nonword repetition test on their phonological working memory assessment.

7 GENERAL DISCUSSION AND CONCLUSIONS

The present study set out to investigate the previously reported (e.g. Service and Kohonen 1995; Baddeley 1993) connection between phonological working memory and L2 knowledge with language tasks that differed from previous studies and allowed comparisons of modality and form vs. meaning scoring. The other locus of interest was the effect of the language of the nonword repetition tests on the possible associations found between phonological working memory and language knowledge. In this chapter the findings of the present study are summarized and compared to those of previous studies. Possible directions for further study are also suggested.

As described in Chapter 4, phonological working memory has been found to be associated with L2 vocabulary learning and knowledge (e.g. Service and Kohonen 1995), various other measures included in task batteries (e.g. Speciale et al. 2004) and such more comprehensive measures as school grades (Service 1989). Furthermore, there have been indications of phonological working memory functioning in a language-specific manner, meaning that the language of the nonword repetition test seems to make a difference to the kinds of connections that are found between phonological working memory and aspects of L2 knowledge (e.g. Masoura and Gathercole 1999; Thorn and Gathercole 1999; French 2003). Overall, the main findings of the present study support the previous findings: success in English (L2) nonword repetition was found to be strongly associated with knowledge of L2 English while the Finnish (L1) nonword repetition test showed a less strong connection. The Finnish nonword repetition test was, in turn, strongly associated with the Finnish reading comprehension measures, indicating that the test itself is functional. English nonword repetition was not connected to the Finnish measures.

7.1 Connection between Children's Knowledge of an L2 and their Success in Repeating L2 Nonwords

The first general research question dealt with the overall connection between L2 knowledge and repetition of nonwords based on that same L2. As stated above, the results offer support for such an association: a statistically significant connection emerged between phonological working memory and L2 knowledge as fairly strong positive correlations were found between the English nonword repetition test and the English language measures in grade six (see sections 6.2.1 and 6.2.2 for details). In addition to this main result, the type of L2 task, i.e. the level of language investigated, did not seem to make much difference to the strength of the connection nor were there clear indications of the modality of the task (spoken or written production) having any effect. Among all of the L2 measures, the spoken vocabulary task shared the weakest connection with English nonword repetition, but it was nevertheless moderately significant. Furthermore, the findings showed no dramatic changes in the connection between phonological working memory and L2 knowledge between grades five and six.

It should be noted that the language material used in the tasks employing the two different modalities were not controlled for in the way that each word and sentence list would have been presented to half of the participants in the spoken task and to another half in the written task. The small number of participants combined with the relatively small number of vocabulary items, and especially sentences, was the reason for the decision to present every participant with exactly the same tasks in the present study. The risk of ending up with two sets of tasks that varied greatly in difficulty and also with two groups of participants who varied in proficiency was considered too great. For fear of ending up with such result-skewing combinations, it was decided that having 15 participants with the same exact tasks was the better option. Obviously, this was done at the expense of being able to tell the effect of the language material apart from that of the modality.

As it is, the different modalities were assessed with different language materials (different word lists and different sentences) and it is impossible to say for certain to which extent the scores in the different modalities actually reflect similarities or differences in the modalities or the language of the tasks. Therefore, the result of the present study that there were no differences between the modalities has to be viewed with caution. The fact that the modalities do not appear to be very differently correlated with phonological working memory may be due to the language content skewing the demands of the tasks. In other words, it may be possible that the differences between the modalities are greater or smaller than is suggested by the present results.

Nevertheless, these results are in agreement with the majority of the previous studies reviewed in section 4.2. That connections were found overall was expected on the basis of prior knowledge. For the different task types or

levels of language there was also evidence of connections from previous studies. In the present study, knowledge of L2 vocabulary was found to be statistically significantly connected to phonological working memory, as assessed with an English nonword repetition test. This result was expected and is similar to those of previous studies on the connection. Thorn and Gathercole (1999) did not compare vocabulary knowledge and phonological working memory directly, but found that the profiles of L2 learners, nonnative bilinguals and native bilinguals on vocabulary tasks and nonword repetition tests were distinct from each other, so that those with better receptive or productive knowledge of vocabulary in a given language were also more successful at repeating nonwords in that particular language. Masoura and Gathercole (1999) found nonword repetition to be linked to L2 English vocabulary scores similarly to Speciale et al. (2004), who found L2 Spanish nonword repetition test to be connected to success on productive and receptive L2 vocabulary tasks.

The other English measures in the present study, sentence and story repetition, were unlike any found in previous studies on phonological working memory in the L2 context, making direct comparisons impossible. However, the written sentence task used in the present study is similar to the dictation task in Service and Kohonen (1995), in which the participants were dictated paragraphs of a story and required to produce long phrases which were then scored for structures. The task was part of a communicative test found to correlate with nonword repetition.

Stories, in general, have been used mainly for reading comprehension tasks (e.g. Service and Kohonen 1995), but in the study by Dufva and Voeten (1999) the participants had to write some parts of a Finnish story in English. This study was also one of the few that assessed L2 tasks based on meaning, as they scored the communicative story task based on propositional information in addition to structures. Dufva and Voeten did not, however, analyze the meaning and structure scores separately, as was done in the present study. In the present case, however, looking at structure and meaning separately did not reveal much, as the scoring perspective did not appear to make a clear difference to the connections found, even though there was some indication of a stronger connection between nonword repetition and the task scores for structure. This does not have to mean that there could be no differences, since assessment always requires a plethora of decisions to be made, and there are therefore countless different ways to assess language tasks whether the target be structure or meaning. It is therefore entirely possible that differences in connections to phonological working memory would have emerged had the scoring been different, not to mention if different kinds of language tasks had been used.

Previous studies (as in assessment at school), reviewed in sections 4.2.1–4.2.4, exhibit a strong written task bias, which does not allow many comparisons of the present findings from the spoken tasks with previous studies in this respect. All of the L2 tasks in the present study involved stimuli presented auditorily, and almost half of the measures required spoken

production. In previous studies, spoken tasks have been occasionally used, but they are very different from the ones in the present study. Auditory stimuli are fairly common, but only a few studies have made use of spoken production. Some vocabulary tasks, where the participants are required to name objects presented to them, have been used (e.g. Thorn and Gathercole 1999), and oral proficiency interviews have been used to study, for example, narrative abilities and fluency (O'Brien et al. 2006; O'Brien et al. 2007). While some such examples exist, L2 tasks that go beyond vocabulary knowledge and require spoken production are a rarity and to take this a step further, it appears that no previous research has sought to compare the spoken and written modalities In fact, it appears almost as if written tasks are preferred immediately the participants' writing is at a level that is at all tolerable. This, of course, makes sense from a research ecological standpoint: data collection and assessment are likely to be less time-consuming when written tasks are used instead of spoken ones.

7.2 Evidence for and against the Language Specificity of Phonological Working Memory in the Repetition of English and Finnish Nonwords

The second general research question aimed at clarifying whether phonological working memory is a part of general language aptitude or somehow language-specific. In brief, the main finding was that phonological memory in itself does not appear to be language-specific. The connections found between phonological working memory and the language tasks varied somewhat according to the language of the phonological working memory measure, but the L1 nonword repetition test correlated with language tasks in both the L1 and the L2.

This result differs from the findings of Masoura and Gathercole (1999) who found a close link between L1 and L2 vocabularies but that only L2 vocabulary was associated with a combination of L1 and L2 nonword repetition. In the present study, in the correlation analyses, the Finnish nonword repetition test was found to be strongly connected to the Finnish language tasks and also to the English language tasks although not as strongly. The situation was not the same for the English nonword repetition test which was strongly connected to the English language measures but not to the Finnish reading tasks.

There is some indication of language specificity in these dissimilar connections between nonword repetition and language knowledge for the two languages. The relationship between English (L2) language knowledge and phonological working memory as assessed with English nonword repetition is rather strong, as is the one between knowledge of Finnish (reading comprehension) and phonological working memory as assessed with a Finnish (L1) nonword repetition test. However, when the language tasks are reversed,

the correlations differ. Finnish nonword repetition still shows a connection with the English language tasks, but the correlation between English nonword repetition and the Finnish language tasks is basically nonexistent. Only the second factual text showed a moderate correlation with the English nonword repetition test.

Based on these findings it seems safe to say that the language of the phonological working memory task does matter, but how exactly remains unclear. It would appear that the Finnish (L1) nonword repetition test taps some more general phonological working memory process, as it showed a connection to tasks in both languages even though the language measures were not correlated with each other. Furthermore, it seems probable that the English (L2) nonword repetition test scores reflect the participants' long-term knowledge of English. It was strongly connected to the English (L2) measures but not the Finnish measures, despite the fact that previous research supports the connection between phonological working memory and reading. It also remained strongly connected to the English measures even after the Finnish nonword repetition test was partialled out, suggesting that the language, English, contributes to the link fairly strongly. Furthermore, the nonword repetition tests in the two languages were not statistically significantly correlated, suggesting that they do not tap the exact same construct or that one of the tasks is influenced by something else as well. In previous studies, some researchers have used L1 nonwords to study the connection between phonological working memory and L2 (e.g. Kormos and Sáfár 2008). Based on the present study, this appears a viable option as the L1-based test showed connections to both languages assessed.

The wordlikeness of the Finnish nonwords used in the present study is unknown because they have not been subjected to evaluation in that sense. However, they could be thought to be fairly low in wordlikeness, as they are based on Finnish syllable patterns and are not merely versions of real words. Accordingly, they should tap phonological working memory rather than knowledge of Finnish to the extent that this is ever possible, as some language influence is necessarily always present, even if not enough to make the nonword repetition test a language task.

By now it has been fairly clearly established that the language of the nonwords in a repetition test makes a difference to the kinds of language connections that are found. This prompts the question whether the correlations found between the L1 nonword repetition test and the language measures in both the L1 and the L2 could be interpreted as the use of L1 nonwords being a way to reach some general phonological working memory. Could the use of L1 nonwords lead to an assessment of a *general* phonological working memory instead of a language-specific one, as seems to be the case with the L2-based nonword repetition test? Considering the role given to working memory in language aptitude, it would be an ideal situation to find that there is a general phonological working memory that influences all L2 learning, and equally importantly to find a task that taps it. A task tapping some general phonological

working memory should show connections to measures in different languages, in this case both L2 English and L1 Finnish, instead of just one language. In the present study this was much more obvious with the L1 Finnish nonword repetition test. Whether using different L1 measures would have shown connections between L1 and English nonword repetition, and whether using L1 reading measures only is the reason for the absence of a connection between English nonword repetition and L1 could of course be speculated on. However, if the English nonword repetition test had tapped a general phonological working memory capacity, the limitation of the reading comprehension tasks would not have been an issue, as numerous previous studies have found phonological working memory ability to predict later L1 reading skills (Gathercole and Baddeley 1993b; Gathercole 1995; Muter 1998).

On the other hand, if it is true that phonological working memory is more important in the beginning stages of language learning and that later it is more about word associations, et cetera, that come from the long-term knowledge of the language, as suggested by, for example, Speciale et al. (2004) and Masoura and Gathercole (2005), then the conclusions could be completely reversed. If we forget for a moment that we are dealing with L1 and L2, but instead think of these as any two languages, one being a language acquired or studied for well over a decade and the other a language studied for a few hours a week for less than four school years, then we could justifiably consider the role of phonological working memory in language knowledge from the time-of-study perspective referred to above.

On the assumption that phonological working memory is more essential early on in language learning, the test most strongly correlated with the L2 measures would be more likely to tap some more general phonological working memory. In the present study, the Finnish nonword repetition test was, in two respects, less strongly connected to the early stage L2 knowledge, i.e. the Finnish nonword repetition test was more clearly connected to the Finnish tasks than English tasks, and the correlations between the Finnish nonword repetition test and L2 English were not as high as the correlations between the English nonword repetition test and L2 English. If phonological working memory is most important in the beginning stages of language learning, how is it possible for the L1 nonword repetition test to tap phonological working memory in general better than the L2 nonword repetition test? These results can, however, be seen as consistent with the English nonword repetition tapping more general working memory, as this test was strongly predictive of the more beginner-level language, L2 English, and less predictive of the participants' L1, in which they have a lifetime's worth of experience and plentiful long-term knowledge.

Gathercole and Masoura (1999) found in their vocabulary study that a more general nonword repetition capacity (i.e. a combination of L1 and L2 nonword repetition tests) was related to L2 but not L1 tasks. If this is applied to the present findings then it could be concluded that the L2 nonword repetition test taps into a general phonological working memory better than the L1 test. In

the study by Gathercole and Masoura (1999), everything correlated with everything, though, and thus they did not have the asymmetry of connections found in the present study. Although this asymmetry may be partly due to the fact that only one dimension of L1 knowledge, reading, was tested, conclusions should nevertheless be drawn with caution. Moreover, work by, for example, French and O'Brien (2008), suggests that phonological working memory may be important for grammar learning after it has stopped predicting word learning. In other words, the influence of phonological working memory on language learning does not necessarily disappear or even weaken upon acquiring more language knowledge, but changes shape.

The fact that the nonword repetition tests were conducted in both the L1 and the L2 in the present study and the correlations between those tests and the language measures in the L1 and L2 varied provides some basis for language specificity/generality speculations such as those presented above. An alternative approach to the problem is to study nonword repetition in a language completely unknown to the participants, such as the Arabic-based nonword test that French (2003) used. If the main source of error in phonological working memory assessment is the participants' varying knowledge of the language that the nonword repetition tests are based on, it is conceivable that nonwords based on an unknown language could solve some of these issues. Indeed, such a test would have made an invaluable addition to the current data. At present, there are no studies comparing the relation of a nonword repetition test based on an unknown language with both L1 and L2 language measures. French (2003) used Arabic nonwords because he did not want L2 development to affect phonological working memory assessment at the beginning and end of an intensive L2 program. French found English (L2) and Arabic nonword repetition tasks to be strongly correlated and equally good predictors of English learning. Similarly, Farnia and Geva (2011) used Hebrew-based nonwords in addition to English ones to control for the familiarity of nonword material. Both tests were found to predict vocabulary development but appeared to tap into different cognitive processes.

Speciale et al. (2004) used nonwordlike nonwords that were not based on any real language but were different length combinations of syllables made by randomly combining certain consonants and vowels. This test predicted the learning of productive L2 vocabulary. Finally, Thorn and Gathercole (1999) were interested in the language specificity of phonological working memory. Their monolingual English speakers performed a French nonword repetition test, but it was only used to show that they did better in an English nonword repetition test and that language knowledge is, thus, linked to success in nonword repetition.

The results from the cluster analysis reported in section 6.3.3 also reflect the unbalanced connections between the nonword repetition and language measures in the two languages. The strongest evidence for language specificity is provided by Cluster C, in which the participants succeeded very differently in the tasks in the two languages (well in the Finnish nonword repetition test

and the Finnish reading tasks, but poorly in the English nonword repetition test and the English language measures), but mostly the data is not very helpful. Many of the participants were either good or poor at everything, which means no difference exists between languages. Perhaps this could be interpreted as general connection between phonological working memory and language knowledge, as poor phonological working memory regardless of the language of the nonword repetition test seems to accompany poor language knowledge in both languages assessed. The same is true of good phonological working memory.

It is the mixed groups – Clusters C and D – that are most interesting. First of all, it is interesting in itself that language-centered combinations were found even in a group of learners as small as the one in the present study. Furthermore, it could be argued that there is less to say about people who are good or poor at everything. Some of the results, especially for the participants seemingly poor at everything, could, of course, stem from willingness to take part in the tasks or even test anxiety, et cetera, but when (systematically) mixed results are found, it gives one something to sink one's metaphorical teeth into. Something must be going on if on the same memory task performance differs clearly depending on the language of the task or if all the other measures indicate level A but all the L1 Finnish tasks (both memory and language) are on level B. Such differences cannot be attributed to a slip-up of some kind when there are several measures for the same variable, as in the case of the language tasks in the present study.

Exploring the participants' individual profiles across the nonword repetition tests and language measures in English and in Finnish, brought to the fore a few issues related to the language specificity of phonological working memory. Already on the level of the whole group, clear differences within participants could be seen in their success in the nonword repetition tasks in the two languages. This was made very explicit by the two extreme cases, participants 8 and 13. If a participant does well on a nonword repetition task in language A and poorly on a similar task in language B, this could be seen as a sign of phonological working memory functioning differently in those two languages. Establishing that there is a difference does not, however, help in understanding what the difference is a sign of. However, together with the other findings it has been argued here that the English nonword repetition test reflects language knowledge while the Finnish nonword repetition test taps into more general phonological working memory. It is probable that both nonword repetition tests are somewhat affected by lexical knowledge, but because the differences between the participants' L1 vocabularies are not likely to be as great as in the L2, it is the English nonword repetition test that appears more dependent on language knowledge. The data in the present study is certainly clear enough to induce speculations but not extensive enough to actually resolve the issue.

What can be said on the basis of participants 8 and 13, and some less extreme cases, is that the language of the phonological working memory test

does matter. If a nonword repetition test is done in only one language – which has been the case in most previous research – this can yield a considerable difference in the impression of phonological working memory of these "mixed" individuals depending on what language is chosen. However, this discussion is based on the results for a mere handful of people. Including the two extreme cases, there are three such contradictory individuals in Cluster C and two in Cluster D. So while the clusters are fairly distinct, this phenomenon clearly needs to be studied further with more participants.

7.3 Evaluation of the Study Conducted

There are still fairly few studies specifically on the relation between phonological working memory and L2 knowledge. At its simplest, the biggest strength or contribution of the present study to the field was the development of a more sensitive L1 nonword repetition test in Finnish and adding to the arsenal of L1 and L2 criterion tasks.

A great amount of effort went into the designing of the L2 test battery, as the aim was to use more complex language tasks, both the written and spoken modalities and dual assessment. While, given the data and the methods of analysis of the present study, modality (written or spoken production) and assessment perspective (structure or meaning) did not ultimately appear to play a great part in the connections, these are nevertheless new results that the present study was able to produce due to the meticulous planning of data gathering.

Another strength of the present study is the twofold assessment of phonological working memory, as the nonword repetition tests were conducted in both L1 and L2. This was mostly possible because of the new Finnish nonword repetition test. As a Finnish nonword repetition test had not successfully discriminated among Finnish 9- to 12-year-old L1 speakers before, employing an L1 Finnish nonword repetition test that managed to discriminate between participants who were already at this age can be considered a minor success in itself. Even more importantly, however, this made it possible to compare the results of phonological working memory measured in two languages in which the participants' knowledge was on very different levels. It further allowed study of the language specificity of phonological working memory and whether one of the tests was sensitive to more general phonological working memory, if such a construct exists. There is still relatively little research of this kind, and although the present study is a fairly small-scale one, it adds to the existing knowledge.

This brings us to the most obvious limitation of the present study. The number of participants was quite small considering the statistical methods that were intended to be used to answer the research questions. The analytical options were certainly limited because of this. For example, dividing the participants into groups according to their level of knowledge in L2 and then

comparing them using various statistical methods, as done by Kormos and Sáfár (2008) and O'Brien et al. (2006), was not feasible with the present data because of the low sample size. However, it was possible to use cluster analysis to form groups of participants distinguished by both phonological working memory and language knowledge in both the L1 and the L2.

The number of participants was due to the fact that the present study was part of a longitudinal research project and these particular participants were the only ones of interest, having been followed throughout their entire school career. However, because the data were longitudinal, the participants were fairly well known, contact had been maintained with their teachers, et cetera. Much was known about the participants outside of what is included in the present study. The situation was thus very different from one where large groups of unknown participants are recruited. On the other hand, the small number of participants also made it feasible to collect data with tasks such as the ones used in the present study, most of which required testing the participants individually and multiple laborious scoring rounds.

Because the longitudinal research project was not very psychometrically inclined and there were no psychologists involved, typical background factors such as intelligence were not assessed or taken into account. That would have been a fairly standard procedure in psychometric studies; however, it has frequently been stated in the phonological working memory literature that intelligence does not have an effect on phonological working memory capacity: for example, low intelligence does not cause low working memory capacity (Gathercole and Alloway 2008: 48). Some connections between phonological working memory and intellectual disability have been found, nevertheless. Numminen, Service and Ruoppila (2002) found that adults with intellectual disability differed from children with matched fluid intelligence on measures of phonological working memory (nonword span and nonword repetition). In working memory tasks that were based on familiar information there were no differences between the groups. Numminen et al. (2002) suggested that the children were able to use working memory more efficiently, whereas the adults with intellectual disability had to rely on long-term knowledge.

The participants in the present study were from culturally and linguistically similar backgrounds, forming in that sense a very homogenous group. Therefore, their experience going of performing a task such as Finnish or English nonword repetition was very much the same, and thus differences in their task performance are not readily attributable to differences in background variables of this kind. Of course, there were considerable differences in the English language knowledge of the 15 participants, but they all had been learning and using English in the same institutional environment, the same school, for an equal amount of time, so the differences in their knowledge of English were within very reasonable bounds.

Since the L2 tasks used in the present study were partly novel, it is of interest to ponder how successful and valid or reliable they are. First of all, the tasks were successful in the sense that they were useful in answering the

research questions. All the participants produced analyzable responses, which is to say that the tasks were neither too easy nor too difficult for the target group. There were, however, a few individual items, or parts of items that were too easy; namely, all the participants scored full meaning points on some of the clauses in the sentence repetition tasks. This never happened when the score was based on structure, as the latter was scored in more detail. The difference between the two scores can be seen as one clear advantage of scoring tasks using two different methods. Similar problems existed with the story task, where the participants scored fairly high meaning scores and variance of the scores was fairly small. An additional strong point of the L2 measures, which has practical relevance, is that the tasks were not too time-consuming. The whole battery could be administered in a few days and none of the individual tasks took so long that the results would have been affected by fatigue.

With perhaps the exception of the meaning score of the story task, none of the tasks stand out very clearly: none of the tasks seemed to be easy or difficult compared to the others or behave otherwise oddly in relation to the averages, and so it could be argued that overall they all (including both scoring methods) reflected the participants' knowledge of English at the time of testing and were valid measures of it. The task with the weakest (although statistically significant) correlations with the other L2 tasks was the spoken vocabulary task. It also shared the weakest correlation with both the Finnish and the English nonword repetition tests. This was surprising, as essentially it is a very commonly performed task. It was only the modality that differed from that most usually applied. However, it should not have been widely different from test situations that the participants have experienced in their language studies. The participants were familiar with being required to orally produce translations of single lexical items in class. Based on the present data, it is difficult to say what should be concluded about the task. It could be an unsuccessful measure or it may have revealed something worth another look as regards language assessment. It may also be that some of the participants felt this test intimidating – precisely the reason why a spoken story task was left out of the battery – and were more reluctant to answer than in the written version or incapable of focusing and answering correctly.

The reliability of the assessment of L2 knowledge was increased by using several measures for this variable. All the measures correlated with each other fairly strongly, which can be seen as indicating that they all tap into the same underlying construct of L2 knowledge. The validity of the L2 assessment rests on two decisions made before the measures were constructed. First, the goal was to inspect L2 knowledge in many different ways to make the assessment as broad as possible: different levels of language were included, different response modalities were employed and different aspects of knowledge were considered in differentiating between structure and meaning. This all was done to ensure that it is, in fact, possible to talk about L2 knowledge, and not, for example, merely vocabulary. Secondly, none of the tasks were entirely made up, but all have a long history in language learning and assessment and could therefore be

expected to reflect L2 knowledge. The vocabulary task was a very common one, only its modality, having to produce translation equivalents in writing and orally based on auditory stimuli, was novel. The sentence repetition tasks were modeled after elicited imitation tasks, and the story retelling task was a modification of the dictogloss, a fairly common task in language classrooms. A spoken version of the story reproduction task was excluded from the battery, partly because of validity considerations. It was considered to be possibly too anxiety-inducing, and causing the participants' productions to tell more about other issues than L2 knowledge.

The tasks, then, were not entirely new, but the idea of studying them all in the phonological working memory context was, and this is where the added value of bringing expertise in applied linguistics to the table in addition to psychology is perhaps clearest. However, it has to be admitted that while the present study criticized previous studies for employing a plethora of measures, it has itself added to this variety by conceiving of new ways to assess language knowledge in connection with phonological working memory. Nevertheless, the goal was to attempt something new, and that was done.

With respect to the language measures overall, the English and Finnish language tasks were very different. For the sake of comparability, similar sets of tasks would have been better, but obviously this would not have been entirely possible because the two languages were L1 and L2. Even if the L1 data had been collected specifically for the present study, a reading comprehension task might still have been considered a valid test for assessing sixth-graders' L1 knowledge. Had the languages under scrutiny all been L2s, the language tasks might have been intentionally made more comparable. As it is, the Finnish language measures used have the benefit of being part of a standardized test battery already deemed a reliable and valid measure of L1 Finnish reading. Nevertheless, more Finnish measures would have helped greatly in the language specificity investigations, and hence ultimately the greatest problem was the narrow focus on L1 reading comprehension.

On a more critical note, the language test battery of the present study was fairly limited. However, there were many demands on the participants' time and they could not be monopolized for the benefit of this research. Probably, there can never be too much data in a study such as the present one, but one phonological working memory measure in two languages, a standardized test of L1 reading comprehension and L2 English tasks covering several levels of language knowledge was more than sufficient to provide satisfactory information on the issues targeted.

The reliability of the nonword repetition tests was already discussed in section 5.4.4, but a few words about the validity of the tests are perhaps appropriate. Again the present study took a methodological perspective, as the aim was to first use functional nonword repetition tests in the L1 and the L2, and secondly to see how they relate to language knowledge. The validity of the nonword repetition test as a measure of phonological working memory was thus a matter under scrutiny here. Based on previous studies, it would seem

that nonword repetition tests have something to say about phonological working memory. Whether the nonwords should be based on a particular language (L1, L2 or an unknown L3) was one of the questions prompting the present study, and the results suggest that at least both the L1 and the L2 nonwords showed a correlation with L2 knowledge. There are a number of reasons why the connection between English nonwords and English language measures appears to be stronger than that between the English language measures and Finnish nonwords. It has been proposed here that the underlying issue is the long-term knowledge of English, which could have affected the English nonword repetition test performance through, for example, the wordlikeness of the English nonwords. Certainly there were more familiar morphemes in the English nonword repetition test than in the Finnish one (e.g. *penneri**ful***, *versatration**ist***, *prindlefenneris**er***). Nonword repetition can never completely escape the effect of long-term language knowledge, but eliminating the most obvious links such as familiar morphemes would be a step in the right direction.

The present study extends the conclusions from previous studies that there are strong links between phonological working memory and L2 knowledge and that it is of consequence which language is used as a basis for a nonword repetition test targeted at phonological working memory. Some suggestions as to what directions future research could take are provided below.

7.4 Suggestions for Future Research

The present study seems to have created more questions than it has answered, which is not necessarily a weakness. First of all, the study gave some indication of an effect of the modality (written and spoken production) of the L2 tasks, but intriguingly the modality did not have the expected effect on the association between phonological working memory and L2 knowledge as assessed with measures requiring spoken production like the nonword repetition test used to assess phonological working memory did. Obviously, there might have been some effect of modality, but it could have easily been masked by the overall strong connection between phonological working memory and L2 knowledge or the fact that the language materials in the spoken and written versions of the tasks were different. This would be one interesting issue to look into in future studies with more data and different methods of analysis.

Another suggestion has to do with the L2 measures used. In phonological working memory investigations in general, because L2 knowledge is a very complex phenomenon, it is important to employ different kinds of L2 tasks. When we use one particular measure to assess it, we can really only talk about what the language learner does or is able to do in the L2 within the confines of that particular task. It is unjust and misleading to then talk about their wider knowledge of language or their ability to learn foreign languages on the basis of a short and simple task. Language knowledge is, of course, a useful umbrella

term in studies where several different kinds of tasks are used, so long as it is clear what kind of language knowledge the study is limited to. The more, and in more varied ways, L2 knowledge is assessed, the better our understanding of the role of phonological working memory in it, be it one language task at a time or larger batteries.

A further aspect in need of more research is the influence of the language of nonword repetition tests. As discussed at length above in section 7.2, the language of phonological working memory tasks remains a dilemma. It makes a difference, but is it possible to get at an assessment of some general phonological working memory by choosing a particular language to base the nonwords of a nonword repetition test on? Based on the results of the present study, the L1 rather than the L2 should be chosen for this purpose, but more research aimed explicitly at this issue would be welcome. Of course, decisions about starting language studies in the first place are not likely to be based on some general phonological working memory capacity, unless it is a very exceptional deficiency we are talking about, as in Baddeley (1993). Nevertheless, knowledge of possible problems would help in taking pre-emptive measures in planning teaching, for example.

In fact, as it is beginning to look increasingly clear that there are strong associations between phonological working memory and L2 knowledge, the first step should be to provide aid for those with poor phonological working memory capacity so that they will not be doomed to gaining only a poor knowledge of an L2. As has been discussed above in section 3.1.2, phonological working memory appears to be fairly fixed once fully developed. The current understanding is that little can be done to improve poor phonological working memory but that there are ways to make it as small an obstacle to L2 learning as possible. In institutional language learning, in particular, this can be achieved by the teacher favoring methods and techniques that do not require learners to rely so heavily on their phonological working memory. Learners themselves can, and probably naturally do, opt for learning strategies that take the processing load off of phonological working memory.

It is therefore suggested that since L2 learning is something that people need to do regardless of their phonological working memory capacity, more research should be directed at the exact methods, techniques and strategies to be used in teaching and learning L2 in cases where the learner has poor phonological working memory capacity, or in general to make the most out of this memory component that is of limited capacity even when functioning perfectly. This should be the ultimate goal of all studies on phonological working memory in the context of L2 knowledge and learning. What we know about the memory construct or its connection to L2 knowledge is of no consequence, if we do not know what to do about it in real-life L2 learning situations.

The research (e.g. French 2003) so far suggests that phonological working memory has the greatest role at the initial stages of L2 learning. After this the accumulating phonological knowledge removes some of the pressure on

phonological working memory capacity. It therefore seems that it is to get through this initial stage that learners with poor phonological working memory need the most help. Knowledge of the role of phonological working memory in L1 acquisition might be helpful in figuring out ways to accomplish this.

The effect of phonological working memory on L1 acquisition is thought to diminish partly because strategies are learnt with age and applied more efficiently to maximize the functioning of working memory (Gathercole and Baddeley 1993a: 25–26). Where L2 learning starts in the early school years, the learners would be old enough to be taught such strategies if their exact nature were known. It has also been proposed that highly redundant exposure to vocabulary is one reason why the effect of an early phonological working memory deficit does not stop children achieving age-appropriate L1 vocabulary levels during the early school years (Gathercole et al. 2005, 2008). Increasing exposure might thus be fruitful in an L2 context as well, but, in most countries, this would require curriculum changes in order to guarantee enough exposure to the L2. Levels of exposure comparable to that in L1 will never be reached in a typical institutional educational setting, although there are at least ways to make considerable increases during the crucial initial stages. If it is not possible to increase total exposure much, quality instead of quantity may be a solution. Focusing more on typical phonemes and phoneme combinations has been found helpful, at least by Dufva and Vauras (2002). Extra teaching of phonology, i.e. more exposure to and rehearsal of the phonology of L2, might be one step towards forming a stronger knowledge base through which long-term knowledge could aid working memory (e.g. Ellis and Sinclair 1996).

Moving to a more theoretical level, it should be kept in mind that the underlying issue here is the connection between working memory and L2 knowledge, not merely the validating of one particular model. Therefore it might be beneficial to give the other working memory models discussed in section 3.1 another look in future research in order to avoid imposing model-induced bias upon the understanding of the relationship. The present study obviously did not take that route. The reasons why the Baddeley and Hitch (1974) model is the most widely applied one are discussed in section 3.2 and while there are fairly understandable reasons for the dominance of the model, it is still only one of several possible explanations for the role memory plays in L2 learning.

Finally, future research on phonological working memory and L2 knowledge might also make a more drastic change of direction in the way suggested by Dörnyei (2009) in his reconsideration of individual differences as best understood in the dynamic systems theory paradigm. The main idea is that individual differences should not be thought of as highly stable and independent of context; instead, individual variation is due to all the pertinent elements operating together. Dörnyei (2009: 209) quotes DeKeyser and Juffs (2005) as having suggested that aptitude or components of it should be studied in context, meaning that the dynamic interrelationship between aptitude and aspects such as learner age, stage of acquisition and teaching is considered.

Dörnyei sees dynamic systems theory as perfectly fitting this kind of an approach.

From the point of view of phonological working memory this may not be as groundbreaking as it first seems. As discussed above, phonological working memory researchers have been aware of the fact that the influence of the memory system on L2 learning varies, in that, for example, it appears to have more of a role early on in language learning, and that teaching methods can be used to make its influence stronger or weaker. Nevertheless, it is intriguing to think of the possibilities and challenges for L2 and phonological working memory research in a context where cognition is considered together with both of the other two areas of mental functioning suggested by Dörnyei (2009: 202), motivation and affect, not to mention the environment.

The present study set out to investigate the cognitive side of things, but even short tasks such as those used here to assess phonological working memory and language knowledge did not appear immune to the effects of motivation and affect, as indicated by the varied performance across tasks by some participants. It is therefore easy to understand that while phonological working memory may be an integral part of the cognitive structures essential to L2 learning, there are a number of other aspects that have to come together for L2 learners to be successful. The phenomenon is truly complex and so, inevitably, is studying it.

YHTEENVETO

FONOLOGINEN TYÖMUISTI JA VIERAAN KIELEN TAITO: SUOMALAISET LAPSET ENGLANNIN OPPIJOINA

Kielenoppijoiden yksilöllisten erojen tutkimuksen keskeinen tavoite on selvittää, mikä tekee joistain ihmisistä muita parempia kielenoppijoita. Muistin, ja etenkin työmuistin, rooli on tärkeä ainakin kognitiivisissa oppimisstrategioissa, mutta myös oppimistyyli voi olla muistisuuntautunut (O'Malley ja Chamot 1990; Dörnyei 2005). Työmuistin on myös havaittu olevan yhteydessä kyvykkyyteen oppia kieliä (engl. *aptitude*) (esim. Skehan 2002).

Työmuistilla on keskeinen rooli tiedon prosessoinnissa ylipäätään mutta myös toisen kielen prosessoinnissa ja oppimisessa (esim. Robinson 2005). Tämä tutkimus keskittyy fonologiseen työmuistiin, joka on osa Baddeleyn (1986) työmuistimallia. Kielellinen informaatio säilyy fonologisessa työmuistissa vain hetken, jos sitä ei pidetä yllä harjoittelemalla, äänettömällä artikuloinnilla. Ongelmat kielen äänteellisen aineksen lyhytaikaisessa varastoimisessa tai harjoittelussa johtavat ongelmiin kielen oppimisessa, koska aines ei siirry pitkäkestoiseen muistiin tai siirtyy sinne väärässä muodossa. Fonologisella työmuistilla siis oletetaan olevan hyvin keskeinen rooli kielenoppimisessa. Se onkin todennäköisesti yksi tärkeä yksilöllisten erojen aiheuttaja kielenoppimisessa.

Vaikka nykytiedon varassa ei vaikuta todennäköiseltä, että fonologiseen työmuistiin voitaisiin suuremmin vaikuttaa, on olemassa keinoja, joilla oppijoita voidaan auttaa käyttämään olemassa olevaa fonologista työmuistiaan mahdollisimman tehokkaasti. Heikon fonologisen työmuistin negatiivisia vaikutuksia kielenoppimiselle voidaan myös pyrkiä ehkäisemään erilaisten opetusmenetelmien avulla. Kirjallisten materiaalien käyttäminen voi lievittää fonologiseen työmuistiin kohdistuvaa painetta, ja ainakin kielen fonologian opettamisen lisääminen, äänneyhdistelmiin keskittyminen ja kirjain-äännevastaavuuksien opettaminen voivat helpottaa uusien sanojen oppimista (Dufva ja Vauras 2002; Ellis ja Sinclair 1996). Fonologisen työmuistin ja vieraan kielen oppimisen välistä yhteyttä on kuitenkin tutkittu vielä suhteellisen vähän, minkä lisäksi näkemykset sopivimmista tavoista arvioida fonologista työmuistia eroavat.

Fonologista työmuistia on tutkittu eniten suhteessa äidinkielen oppimiseen, mutta sen roolista etenkin vieraan kielen sanaston oppimisessa on myös näyttöä. Servicen (1989) tutkimus on ensimmäisiä vieraan kielen oppimiseen liittyviä tutkimuksia, ja yksi sen johtopäätöksistä on, että alkuvaiheen englannin oppimista on mahdollista ennustaa epäsanantoiston perusteella. Service ja Kohonen (1995) puolestaan totesivat, että yhteys epäsanantoiston ja kielitaidon eri mittareiden välillä riippuu luultavasti sanastosta. Sanaston oppimiseen näyttäisi vaikuttavan etenkin äänetön artikulointi fonologisessa työmuistissa (esim. Kaushanskaya ja Yoo 2011). Tutkimushenkilöiden mahdollinen monikielisyys ja vieraan kielen oppimisen aloitusikä näyttävät vaikuttavan fonologisesta työ-

muistista saatavaan arvioon (esim. Papagno ja Vallar 1995). On myös näyttöä, että fonologisen työmuistin rooli on erilainen kielenoppimisen eri vaiheissa (esim. O'Brien, Segalowitz, Collentine ja Freed 2006). Kaiken kaikkiaan aiempi tutkimus vaihtelee suuresti sen suhteen, millaisia kielitaidon arvioinnin menetelmiä ja fonologisen työmuistin mittareita on käytetty.

Epäsanantoistotesti on yksi yleisimmistä fonologisen työmuistin arviointimenetelmistä. Epäsanoilla tarkoitetaan merkityksettömiä mutta olemassa olevalta kieleltä kuulostavia sanoja, jotka testissä tyypillisesti kuunnellaan ja toistetaan yksi kerrallaan. Aiemmassa tutkimuksessa epäsanantoistotestien pohjana ollut kieli on vaihdellut äidinkielestä opiskeltavaan tai jopa täysin tuntemattomaan vieraaseen kieleen.

Tämän tutkimuksen tavoitteena on laajentaa kielitaidon arviointi kattamaan kielen eri tasoja ja verrata aiempaa systemaattisemmin kirjoitettua ja puhuttua kieltä sekä rakenteiden ja merkityksen pohjalta tehtyjä arviointeja. Tutkimuksessa tarkastellaan fonologisen työmuistin ja vieraan kielen taidon välistä yhteyttä äidin- ja vieraskielisen epäsanantoistotestin avulla. Molempia epäsanantoistotestejä verrataan myös äidinkielen lukutaitoon fonologisen työmuistin kielispesifisyyden arvioimiseksi. Tiivistetysti tämä tutkimus siis käsittelee fonologisen työmuistin ja vieraan kielen taidon suhdetta sekä epäsanantoistotestiä fonologisen työmuistin arvioinnin menetelmänä. Tutkimuskysymykset ovat:

1. Onko lasten vieraan kielen taidon ja vieraskielisten epäsanojen toiston välillä yhteyttä?
2. Onko fonologinen työmuisti kielispesifi?

Tutkimus on luonteeltaan pääasiassa määrällinen: muuttujien välisiä suhteita on tarkasteltu korrelaatio- ja ryhmittelyanalyysin keinoin. Tutkimushenkilöt ovat suomalaisia lapsia, jotka ovat opiskelleet englantia ensimmäisenä vieraana kielenään kolmannelta luokalta asti. Tämän tutkimuksen aineisto on kerätty pääosin kuudennella luokalla, mutta myös viidennen luokan aineistoa käytetään osassa analyysejä.

Tutkimus on osa laajempaa pitkittäistutkimusta (*Tilanteinen kielellinen tietoisuus ja vieraan kielen oppiminen*), joten siinä on käytetty myös valmiina ollutta aineistoa, mutta pääosa aineistosta on kerätty erityisesti tätä tutkimusta varten. Englannin kielen taitoa arvioitiin kuudennella luokalla viidellä eri tehtävällä: suullisella ja kirjallisella sanastotehtävällä, suullisella ja kirjallisella virketehtävällä ja kirjallisella tarinatehtävällä. Kaikkien tehtävien pohjana oli kuuloärsykkeisiin reagoiminen. Sanastotehtävissä tutkimushenkilöt tuottivat suullisen tai kirjallisen käännösvastineen kuulemilleen suomen- ja englanninkielisille sanoille. Virketehtävissä kuullut englanninkieliset virkkeet toistettiin suullisesti tai kirjallisesti yksi kerrallaan, kun taas tarinatehtävässä ainoastaan kirjoitettiin lyhyt saneltu tarina.

Suomen kielen arviointi kuudennella luokalla perustuu neljään Ala-asteen lukutestin osioon (kaksi tieto- ja kertomustekstiä) (Lindeman 1998). Viidennellä luokalla suomen kieltä ei arvioitu, mutta englannin kielen taitoa arvioitiin kaik-

kiaan kahdeksalla tehtävällä (sanasto, kuullun ymmärtäminen, luetun ymmärtäminen, verbimuototehtävät, lauseenmuodostustehtävät).

Fonologisen työmuistin arvioimiseksi suomen- ja englanninkieliset epäsanantoistotestit tehtiin sekä viidennellä että kuudennella luokalla. Koska soveltuvaa suomenkielistä epäsanantoistotehtävää ei ollut saatavilla, sellainen kehitettiin tätä tutkimusta varten. Englanninkielinen epäsanantoistotesti on muokattu olemassa olevasta testistä (Gathercole, Willis, Emslie ja Baddeley 1991; Gathercole 1995). Suomalaisia epäsanoja toistettiin molemmilla kerroilla kaikkiaan 21 (2–8 tavua) ja englanninkielisiä 18 (1–6 tavua).

Tutkimuksessa löydettiin yhteys lasten fonologisen työmuistin ja vieraan kielen taidon välillä: sekä suomen- että englanninkieliset epäsanantoistotehtävien tulokset korreloivat positiivisesti englannin kielen taidon kanssa. Kielitehtävien modaliteetilla ei näyttänyt olevan juurikaan merkitystä yhteyden ilmenemiselle, toisin sanoen sekä suulliset että kirjalliset tehtävät korreloivat epäsanantoistotehtävien kanssa. Vain suullinen sanastotehtävä osoittautui hieman poikkeavaksi, mutta positiivinen korrelaatio sen ja epäsanantoiston välillä vahvistui, kun poikkeava havainto poistettiin analyysistä. Yhteyteen ei myöskään näyttänyt vaikuttavan se, arvioitiinko kielitehtävät rakennetta vai merkitystä painottaen.

Myös viitteitä fonologisen työmuistin kielispesifisyydestä löydettiin. Ainostaan suomenkielinen epäsanantoistotesti ennusti suomen kielen lukutaitoa. Luultavasti englanninkielinen epäsanantoistotesti oli alttiimpi kielitaidon vaikutukselle tai englannin kielen taito vaihteli tutkimushenkilöiden välillä enemmän kuin heidän äidinkielen taitonsa, ja tämä heijastui epäsanantoistotehtävien erilaisina suhteina kielitaitotehtäviin. Saattaa myös olla, että lukutaitotehtävässä ja suomalaisen epäsanantoistotestin suorittamisessa oli yhteisiä piirteitä, joita englanninkielisessä epäsanantoistotehtävässä ei ollut. Epäsanantoistotestin kielellä joka tapauksessa oli merkitystä sille, millaisena fonologisen työmuistin ja kielitaidon välinen yhteys näyttäytyi, mutta etenkin laajempi suomen kielitaidon arviointi olisi ollut tarpeen lisäselvitysten tekemiseksi.

Fonologisen työmuistin kielispesifisyyttä tutkittiin myös ryhmittelyanalyysillä, joka tuotti tarkempaa tietoa siitä, millaisia yhteyksiä epäsanantoistotestin ja kielitaidon välillä oli yksilöiden tasolla. Vaikka korrelaatioanalyysien perusteella yhteys epäsanantoistotestien ja vieraan kielen taidon välillä vaikutti selvältä, saattoi se yksilöiden kohdalla vaihdella suurestikin. Englannin kielen taidossa olikin suuria eroja etenkin niillä, jotka menestyivät hyvin suomenkielisessä epäsanantoistotestissä.

Tämän tutkimuksen tulosten perusteella yhteys fonologisen työmuistin ja vieraan kielen taidon välillä on selkeä ja ainakin melko alkuvaiheen oppijoiden osalta riippumaton kielitehtävien modaliteetista ja siitä, arvioidaanko rakenteiden vai merkityksen prosessointia. Fonologisessa työmuistissa puolestaan vaikuttaisi olevan sekä kielispesifejä että kielestä riippumattomia piirteitä.

Vaikka tämän tutkimuksen perusteella ei voida suoraan sanoa, mikä lähdekieli olisi sopivin, on tulosten perusteella kuitenkin selvää, että epäsanantoistotehtävän kielellä on merkitystä sille, millainen kuva fonologisesta työmuistis-

ta saadaan. Testin käyttötarkoituksesta riippuen olisikin siis tärkeää pohtia, minkä kielinen epäsanantoistotesti on kulloiseenkin tilanteeseen sopivin. Koska yhteys fonologisen työmuistin ja vieraan kielen taidon välillä vaikuttaa vahvalta, olisi jatkossa syytä keskittyä pohtimaan keinoja, miten opetuksessa voitaisiin mahdollisimman hyvin huomioida fonologisen työmuistin ongelmat ja varmistaa, että niistä johtuvat yksilölliset erot saadaan minimoitua. Tärkeintä on taata kaikille tasavertainen mahdollisuus oppia vierasta kieltä mahdollisista fonologisen työmuistin ongelmista riippumatta.

REFERENCES

Adams, A-M. & S. E. Gathercole 2000. Limitations in working memory: implications for language development. *International Journal of Language & Communication Disorders* 35 (1), 95–116.

Adams, A-M. & C. Willis 2001. Language processing and working memory: a developmental perspective. In J. Andrade (ed.), *Working Memory in Perspective*. London: Psychology Press, 79–100.

Alanen, R. 2003. A sociocultural approach to young language learners' beliefs about language learning. In P. Kalaja & A. M. F. Barcelos (eds.), *Beliefs about SLA: New Research Approaches*. Dordrecht: Kluwer, 55–85.

Alanen, R. 2006. "Birds in the sky": tilanteinen kielellinen tietoisuus ja vieraan kielen oppiminen. In R. Alanen, H. Dufva & K. Mäntylä (eds.), *Kielen päällä. Näkökulmia kieleen ja kielen käyttöön*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 185–222.

Aldenderfer, M. S. & R. K. Blashfield 1984. *Cluster Analysis.* Newbury Park, CA: Sage Publications.

Alkula, T., S. Pöntinen & P. Ylöstalo 1994. *Sosiaalitutkimuksen kvantitatiiviset menetelmät*. Juva: WSOY.

Alloway, T. P., S.E. Gathercole, A-M. Adams, C. Willis, R. Eaglen & E. Lamont 2005. Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology* 23, 417–426.

Altarriba, J. & D. M. Basnight-Brown 2009. An information-processing approach to second language acquisition. In W. C. Ritchie & T. K. Bhatia (eds.), *The New Handbook of Second Language Acquisition*. Bingley: Emerald, 115–136.

Anderson, J. R. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Andrade, J. 2001. An introduction to working memory. In J. Andrade (ed.), *Working Memory in Perspective*. London: Psychology Press, 3–30.

Aro, M. 2001. Lasten käsityksiä englannin ja suomen kielestä. Unpublished Pro Gradu Thesis. University of Jyväskylä, Centre for Applied Language Studies.

Aro, M. 2003. *Speakers and Doers: Polyphony and Agency in Children's Beliefs about Language Learning.* Jyväskylä Studies in Humanities 116. Jyväskylä: University of Jyväskylä.

Aro, M. 2006. Anteeksi, kuka puhuu? Lasten kielikäsitysten moniäänisyydestä. In R. Alanen, H. Dufva & K. Mäntylä (eds.), *Kielen päällä. Näkökulmia kieleen ja kielen käyttöön*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 53–76.

Atkins, P. W. B. & A. D. Baddeley 1998. Working memory and distributed vocabulary learning. *Applied Psycholinguistics* 19, 537–552.

Avons, S. E., C. A. Wragg, L. Cupples & W. J. Lovegrove 1998. Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics* 19, 583–601.

Baddeley, A . 1986. *Working Memory*. Oxford: Clarendon Press.

Baddeley, A. 1992. Working memory. *Science* 255 (5044), 556–559.

Baddeley, A. 1993. Short-term phonological memory and long-term learning: a single case study. *European Journal of Cognitive Psychology* 5 (2), 129–148.

Baddeley, A. 2000. The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences* 11, 417–423.

Baddeley, A. 2003. Working memory and language: an overview. *Journal of Communication Disorders* 36, 189–208.

Baddeley, A, S. Gathercole & C. Papagno 1998. The phonological loop as a language learning device. *Psychological Review* 105 (1), 158–173.

Baddeley, A. D. & G. Hitch 1974. Working memory, in G. H. Bower (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory 8*. New York: Academic Press, 47–89.

Berman, R. A. & D. I. Slobin 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillside, NJ: Lawrence Erlbaum.

Bley-Vroman, R. & C. Chaudron 1994. Elicited imitation as a measure of L2 competence. In E. E. Tarone, S. M. Gass & A. D. Cohen (eds.), *Research Methodology in SLA*. Hillsdale, NJ: Lawrence Erlbaum, 245–261.

Bowey, J. A. 1996. On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology* 63, 44–78.

Brady, S., E. Poggie & M. M. Rapala 1989. Speech repetition abilities in children who differ in reading skill. *Language and Speech* 32 (2), 109–122.

Carroll, J. B. 1981. Twenty-five years of research on foreign language aptitude. In K. C. Diller (ed.), *Individual Differences & Universals in Language Learning Aptitude*. Rowley, MA: Newbury House, 83–118.

Chamot, A. U. & J. M. O'Malley 1994. Language learner and learning strategies. In N. C. Ellis (ed.), *Implicit and Explicit Learning of Languages*. London: Academic Press, 371–392.

Cheung, H. 1996. Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology* 32 (5), 867–873.

Conway, A. R. A. & R. W. Engle 1996. Individual differences in working memory capacity: more evidence for a general capacity theory. *Memory* 4 (6), 577–590.

Dollaghan, C. A., M. E. Biber & T. F. Campbell 1995. Lexical influences on nonword repetition. *Applied Psycholinguistics* 16, 211–222.

Dufva, H. 2003. Beliefs in dialogue: Bakhtinian view. In P. Kalaja & A. M. F. Barcelos (eds.), *Beliefs about SLA: New Research Approaches*. Dordrecht: Kluwer, 131–152.

Dufva, H., R. Alanen & M. Aro 2003. Kieli objektina – miten lapset mieltävät kielen? In M. Koskela & N. Pilke (eds.), *Kieli ja asiantuntijuus.* AFinLAn

vuosikirja 61. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 295–315.

Dufva, H., R. Alanen & K. Mäntylä 2001. Ei kissa ole mikään sana: lasten näkökulmia kieleen. In H. Sulkala & L. Nissilä (eds.), *XXVII Kielitieteen päivät Oulussa 19.–20.5.2000.* Acta Universitatis Ouluensis Humaniora B 41. Oulu: University of Oulu, 35–45.

Dufva, M. & M. Vauras 2002. Promoting at-risk pupils' foreign language literacy learning. In L. Verhoeven, C. Elbro & P. Reitsma (eds.), *Precursors of Functional Literacy*. Amsterdam: John Benjamins, 317–337.

Dufva, M. & M. J. M. Voeten 1999. Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics* 20, 329–348.

Dörnyei, Z. 2005. *The Psychology of the Language Learner*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z. 2009. *The Psychology of Second Language Acquisition*. Oxford: Oxford University Press.

Dörnyei, Z. & P. Skehan 2003. Individual differences in second language learning. In C.J. Doughty & M.H. Long (eds.), *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell, 589–630.

Eisenstein, M., N. Bailey & C. Madden 1982. It takes two: contrasting tasks and contrasting structures. *TESOL Quarterly* 16, 381-93.

Ellis, N. C. 2001. Memory for language. In P. Robinson (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, 33–68.

Ellis, N. C. & S. G. Sinclair 1996. Working memory in the acquisition of vocabulary and syntax: putting language in good order. *The Quarterly Journal of Experimental Psychology* 49A (1), 234–250.

Engle, R. W., J. Cantor & J. J. Carullo 1992. Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology* 18 (5), 972–992.

Erlam, R. 2006. Elicited imitation as a measure of L2 implicit knowledge: an empirical validation study. *Applied Linguistics* 27 (3), 464–491.

Everitt, B. S. & G. Dunn 1991. *Applied Multivariate Data Analysis*. London: Edward Arnold.

Farnia, F. & E. Geva 2011. Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics* 32, 711–738.

Feldman, M. J. 1985. Evaluating pre-primer basal readers using story grammar. *American Educational Research Journal* 22 (4), 527–547.

French, L. 2003. Phonological working memory and L2 acquisition: a developmental study of Quebec francophone children learning English. Unpublished doctoral dissertation. Université Laval Québec, Département de langues, linguistique et traduction.

French, L. M. & I. O'Brien 2008. Phonological memory and children's second language grammar learning. *Applied Psycholinguistics* 29, 463–487.

Gass, S. M. & A. Mackey 2007. *Data Elicitation for Second and Foreign Language Research*. New York: Routledge.

Gathercole, S. E. 1995. Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition* 23 (1), 83–94.

Gathercole, S. E. 2006. Nonword repetition and word learning: the nature of the relationship. *Applied Psycholinguistics* 27 (4), 513–543.

Gathercole, S. E. & T. P. Alloway 2008. *Working Memory and Learning: A Practical Guide for Teachers*. London: Sage.

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

Gathercole, S. E. & A. D. Baddeley 1993a. *Working Memory and Language*. Hove: Lawrence Erlbaum.

Gathercole, S. E. & A. Baddeley 1993b. Phonological working memory: a critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education* 8 (3), 259–272.

Gathercole, S. E. & G. J. Hitch 1993. Developmental changes in short-term memory: a revised working memory perspective. In A. F. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris (eds.), *Theories of Memory*. Hove: Lawrence Erlbaum, 189–209.

Gathercole, S. E., C. Willis & A. D. Baddeley 1991. Nonword repetition, phonological memory, and vocabulary: a reply to Snowling, Chiat, and Hulme. *Applied Psycholinguistics* 12, 375– 379.

Gathercole, S .E., G. J. Hitch, E. Service & A. J. Martin 1997. Phonological short-term memory and new word learning in children. *Developmental Psychology* 33 (6), 966–979.

Gathercole, S. E., C. S. Willis, A. D. Baddeley & H. Emslie 1994. The children's test of nonword repetition: a test of phonological working memory. In S. E. Gathercole & R. A. McCarthy (eds.), *Memory Tests and Techniques*. Hove: Lawrence Erlbaum, 103–127.

Gathercole, S. E., C. Willis, H. Emslie & A. D. Baddeley 1991. The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics* 12, 349–367.

Gathercole, S. E., C. Willis, H. Emslie & A. D. Baddeley 1992. Phonological memory and vocabulary development during the early school years: a longitudinal study. *Developmental Psychology* 28 (5), 887–898.

Gathercole, S. E., E. Service, G. J. Hitch, A-M. Adams & A. J. Martin 1999. Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology* 13, 65–77.

Gathercole, S. E., C. Tiffany, J. Briscoe, A. Thorn & The ALSPAC team 2005. Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry* 46 (6), 598–611.

Gathercole, S. E., J. Briscoe, A. Thorn, C. Tiffany & ALSPAC Study team 2008. Deficits in verbal long-term memory and learning in children with poor phonological short-term memory skills. *The Quarterly Journal of Experimental Psychology* 61 (3), 474–490.

Gernsbacher, M. A. 1994. *Handbook of Psycholinguistics*. San Diego, CA: Academic Press.

Griffiths, C. (ed.) 2008. *Lessons from Good Language Learners*. Cambridge: Cambridge University Press.

Grigorenko, E. L., R. J. Sternberg & M. E. Ehrman 2000. A theory-based approach to the measurement of foreign language learning ability: the Canal-F theory and test. *The Modern Language Journal* 84, 390–405.

Harrington M. & M. Sawyer 1992. L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition* 14 (1), 25–38.

Heikkilä, T. 1998. *Tilastollinen tutkimus*. Helsinki: Edita.

Holsgrove, J. V. & A. F. Garton 2006. Phonological and syntactic processing and the role of working memory in reading comprehension among secondary school students. *Australian Journal of Psychology* 58 (2), 111–118.

Hulstijn, J. H. 2007. The shaky ground beneath the CEFR: quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal* 91 (4), 663–667.

Hummel, K. M. & L. M. French 2010. Phonological memory and implications for the second language classroom. *The Canadian Modern Language Review* 66 (3), 371–391.

Just, M. A. & P. A. Carpenter 1992. A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99 (1), 122–149.

Kaushanskaya, M. & J. Yoo 2011. Rehearsal effects in adult word learning. *Language and Cognitive Processes* 26 (1), 121–148.

Kintsch, W. & T. A. van Dijk 1978. Toward a model of text comprehension and production. *Psychological Review* 85 (5), 363–394.

Kintsch, W. & K. A. Rawson 2005. Comprehension. In M. J. Snowling & C. Hulme (eds.), *The Science of Reading: A Handbook*. Malden, MA: Blackwell, 209–226.

Kintsch, W., A. F. Healy, M. Hegarty, B. F. Penington & T. A. Salthouse 1999. Models of working memory: eight questions and some general issues. In A. Miyake & P. Shah (eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. New York: Cambridge University Press, 412–441.

Kormos, J. & A. Sáfár 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11 (2), 261–271.

Krashen, S. D. 1981. Aptitude and attitude in relation to second language acquisition and learning. In K. C. Diller (ed.), *Individual Differences & Universals in Language Learning Aptitude*. Rowley, MA: Newbury House, 155–175.

Kyllonen, P. C. & R. E. Christal 1990. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14, 389–433.

Lehto, J. 1996. *Työmuistin yhteys tekstin tiivistämiseen, ongelmanratkaisuun ja koulumenestykseen*. Helsinki: Helsingin yliopisto.

Lindeman, J. 1998. *Ala-asteen lukutesti: tekniset tiedot*. Åbo: Åbo Akademis Förlag.

Lovatt, P. & S. Avons 2001. Re-evaluating the word-length effect. In J. Andrade (ed.), *Working Memory in Perspective*. London: Psychology Press, 199–218.

Lust, B., S. Flynn & C. Foley 1996. What children know about what they say: elicited imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee & H. Smith Cairns (eds.), *Methods for Assessing Children's Syntax*. Cambridge, MA: MIT Press, 55–76.

Majerus, S., M. Poncelet, C. Greffe & M. Van der Linden 2006. Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology* 93 (2), 95–119.

Masoura, E. V. & S. E. Gathercole 1999. Phonological short-term memory and foreign language learning. *International Journal of Psychology* 34 (5/6), 383–388.

Masoura, E. V. & S. E. Gathercole 2005. Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in foreign language. *Memory* 13 (3/4), 422–429.

McLauglin, B. 1987. *Theories of Second-Language Learning*. London: Edward Arnold.

Metsala, J. L. 1999. Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology* 91 (1), 3–19.

Michas, I. C. & L. A. Henry 1994. The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology* 12, 147–163.

Miettinen, H. 2003. Nonword repetition and L2 knowledge. Unpublished Pro Gradu Thesis. University of Jyväskylä, Department of Languages.

Miyake A. & P. Shah 1999. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* Cambridge: Cambridge University Press.

Munnich, E., S. Flynn & G. Martohardjono 1994. Elicited imitation and grammaticality judgement tasks: what they measure and how they relate to each other. In E. E. Tarone, S. M. Gass & A. D. Cohen (eds.), *Research Methodology in SLA*. Hillsdale, NJ: Lawrence Erlbaum, 227–243.

Muter, V. 1998. Concurrent and longitudinal predictors of reading: the role of metalinguistic and short-term memory skills. *Reading Research Quarterly* 33 (3), 320–337.

Mäntylä, K. 2004. "Ilo on aika yleistä: lapset kielitieteen termejä soveltamassa". In P. Muikku-Werner & H. Stotesbury, *Minä ja kielitiede – soveltajan arki.* AFinLAn vuosikirja 62. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 43–54.

Mäntylä, K. 2008. Vocabulary in an EFL classroom. In M. Edwardes (ed.), *Proceedings of the BAAL Annual conference 2007*. Scitsiugnil Press/BAAL. CD- publication.

Nation, K. 2005. Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (eds.), *The Science of Reading: A Handbook*. Malden, MA: Blackwell, 248–265.

Numminen, H., E. Service & I. Ruoppila 2002. Working memory, intelligence and knowledge base in adult persons with intellectual disability. *Research in Developmental Disabilities* 23, 105–118.

O'Brien, I., N. Segalowitz, J. Collentine & B. Freed 2006. Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics* 27, 377–402.

O'Brien, I., N. Segalowitz, B. Freed & J. Collentine 2007. Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition* 29, 557–582.

O'Malley, J. M. & A. U. Chamot 1990. *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

Oxford, R. L. 1990. *Language Learning Strategies. What Every Teacher Should Know*. Boston: Heinle & Heinle.

Page, M. & R. Henson 2001. Computational models of short-term memory: modeling serial recall of verbal material. In J. Andrade (ed.), *Working Memory in Perspective*. London: Psychology Press, 177–198.

Papagno, C. & G. Vallar 1995. Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology* 48A (1), 98–107.

Papagno, C., T. Valentine & A. Baddeley 1991. Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language* 30, 331–347.

Potter, M. C. & L. Lombardi 1990. Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29, 633–654.

Ranta, L. 2008. Aptitude and good language learners. In C. Griffiths (ed.), *Lessons from Good Language Learners*. Cambridge: Cambridge University Press, 142–155.

Robinson, P. 1995. Attention, memory and the "noticing" hypothesis. *Language Learning* 45 (2), 283–331.

Robinson, P. 2001. Preface. In P. Robinson (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, viii–x.

Robinson, P. 2002a. Learning conditions, aptitude complexes, and SLA: a framework for research and pedagogy. In P. Robinson (ed.), *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins, 113–133.

Robinson, P. 2002b. Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: a replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (ed.),

Individual Differences and Instructed Language Learning. Amsterdam: John Benjamins, 211–266.

Robinson, P. 2005. Aptitude and second language acquisition. *Annual Review of Applied Linguistics* 25, 46–73.

Rubin, J. 1975. What the "good language learner" can teach us. *TESOL Quarterly* 9 (1), 41–51.

Sachs, J. S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* 2 (9), 437–442.

Sáfár, A. & J. Kormos 2008. Revisiting problems with foreign language aptitude. *IRAL* 46, 113–136.

Sawyer, M. & L. Ranta 2001. Aptitude, individual differences, and instructional design. In P. Robinson (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, 319–353.

Scheibner-Herzig, G., H. Sauerbrey & S. Kokoschka 1991. Repetition – a means to predict foreign language oral proficiency. *IRAL* 29, 229–240.

Segalowitz, N. & J. Hulstijn 2005. Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press, 371–388.

Seigneuric, A., M-F. Ehrlich, J. V. Oakhill & N. M. Yuill 2000. Working memory resources and children's reading comprehension. *Reading and Writing: An Interdisciplinary Journal* 13, 81–103.

Service, E. 1989. *Phonological Coding in Working Memory and Foreign-Language Learning*. Helsinki: Helsingin yliopisto.

Service, E. & F. I. M. Craik 1993. Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language* 32, 608–623.

Service, E. & V. Kohonen 1995. Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* 16, 155–172.

Sim, J. & C. C. Wright 2005. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy* 85 (3), 257–268.

Skehan, P. 1989. *Individual Differences in Second-Language Learning*. London: Edward Arnold.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. 2002. Theorising and updating aptitude. In P. Robinson (ed.), *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins, 69–93.

Smith, B. 2006. Precautions regarding nonword repetition tasks. *Applied Psycholinguistics* 27 (4), 584–587.

Snowling, M., S. Chiat & C. Hulme 1991. Words, nonwords, and phonological processes: some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics* 12, 369–373.

Speciale, G., N. C. Ellis & T. Bywater 2004. Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25 (2), 293–312.

Sternberg, R. J. 1988. *The Triarchic Mind: A New Theory of Human Intelligence*. New York: Viking.

Swanson, H. L. & M. Howell 2001. Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology* 93 (4), 720–734.

Swanson, H. L., L. Sáez, M. Gerber & J. Leafstedt 2004. Literacy and cognitive functioning in bilingual and nonbilingual children at or not at risk for reading disabilities. *Journal of Educational Psychology* 96 (1), 3–18.

Takala, S. 1984. *Evaluation of Students' Knowledge of English Vocabulary in the Finnish Comprehensive School*. Kasvatustieteiden tutkimuslaitoksen julkaisuja 350. Jyväskylä: Kasvatustieteiden tutkimuslaitos.

Thorn, A. S. C. & S. E. Gathercole 1999. Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology* 52A (2), 303–324.

van Bon, W. H. J. & J. M. L. van der Pijl 1997. Effects of word length and wordlikeness on pseudoword repetition by poor and normal readers. *Applied Psycholinguistics* 18, 101–114.

van der Lely, H. K. J. & N. Gallon 2006. Exploring the impact of higher level linguistic representations on nonword repetition performance. *Applied Psycholinguistics* 27 (4), 591–594.

Vinther, T. 2002. Elicited imitation: a brief overview. *International Journal of Applied Linguistics* 12, 54–73.

Vitevitch, M. S. 2006. Manipulating the characteristics of words and nonwords to better understand word learning. *Applied Psycholinguistics* 27 (4), 594–598.

Wajnryb, R. 1990. *Grammar Dictation*. Oxford: Oxford University Press.

Wesche, M. B. 1981. Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (ed.), *Individual Differences & Universals in Language Learning Aptitude*. Rowley, MA: Newbury House, 119–154.

Westlake, P., M-L. Lintunen, E-L. Pitkänen & S. Satamo 1997. *Yes Adventures*. Helsinki: WSOY.

APPENDIX 1: The Finnish nonwords

ruuppa
tusti
lanni

uoraste
tuilasmo
uroula

piuttuikere
lärskätelo
kitsamellu

surstisynäyppö
tiurunauppama
aalsannokuurste

kuossakeetitarsko
uispauttuhaahuumu
vokkohotoiheltsu

naihtotunkkoesniikki
mauppohuottiallamo
vuoksattiisisuilasmo

kiippakonnomeittomella
ottiisiettaiskelame
lumponeetrikysättäämö

APPENDIX 2: The English nonwords

sep
clird
tull

diller
hampent
bannow

commerine
brasterer
skiticult

pennerriful
blonterstaping
empliforvent

versatrationist
sepretennial
altupatory

prindlefenneriser
bannifertrumpetine
smipdefermication

APPENDIX 3: The Written vocabulary task

From Finnish to English	From English to Finnish
Practice item: mennä naimisiin (get married)	Practice item: a horse
Actual list:	Actual list:
anteeksi (sorry, pardon)	an ankle
aurinkovarjo (a sun umbrella)	a berry
diakuva (a slide)	a breakfast
herra (Mr)	bring
hylje (a seal)	a cup
japanilainen, japanin kieli (Japanese)	half-price
	interview
keppi (a stick)	a legend
kun (when, as)	marmalade
kylpyhuone (bathroom)	May
Lepakkomies (Batman)	a moment
muuttaa (move, change)	obey
myöskään (either)	onto
outo (strange, weird)	polite
pikkutakki (a jacket)	some
porkkana (a carrot)	steel
Ranska (France)	a tooth
rotta (a rat)	a tortoise
seisoa (stand)	a village
sen sijaan (instead)	a kingdom
tehdä hulluksi (drive mad)	

APPENDIX 4: The Spoken vocabulary task

From Finnish to English	**From English to Finnish**
Practice item: koskettaa (touch)	Practice item: a book
Actual list:	Actual list:
askel (a step)	among
bussi (a bus)	blue
Espanja (Spain)	a brick
joka tapauksessa (anyway)	company
kuluttaa (spend)	dark
kunnes (until)	a frog
kynä (a pen)	a fruit
Lumikki (Snow White)	a helicopter
läikyttää (spill)	march
mekko (a dress)	a mat
pikkuinen (small, tiny)	a match
piparkakku (gingerbread)	Mexican
planeetta (a planet)	a nose
pyyhekumi (a rubber)	only
saada (get)	Pippy Long Stocking
sattua (hurt, happen)	ring
takki (a coat)	sore
toukokuu (May)	still
ulkopuolella (outside)	a tomato
vakava (serious)	a vegetable

APPENDIX 5: The sentences in the written sentence task

Tom likes ice-cream a lot, but he doesn't eat it at school.

They are washing dishes but they aren't cleaning the whole house.

The boys in this class are going to the cinema tomorrow.

He does his job very well, and he doesn't waste any time.

I have got two younger sisters, so I am the oldest.

The girls in my group walked to school yesterday.

I am reading a book and my friend is watching TV.

She has got three older brothers, so she is the youngest.

They do their homework, but they don't enjoy it very much.

I love movies but I didn't see the movie about the dog.

APPENDIX 6: The sentences in the spoken sentence task

Emma can speak French but she doesn't do it very often.

She is waiting for her friends, but they aren't coming tonight.

My family is going to Africa next summer.

She has to walk to school because she doesn't have a bike.

I have got a new computer, but Tom's computer is better.

His cousins travelled to Germany last winter.

We are playing football today with Mike and some other friends.

My sister has got a cat, but I think that dogs are nicer.

We don't have much time so we really have to hurry.

Tom didn't do his homework, so he can't go home yet.

APPENDIX 7: The language-memory clusters described with participants' (composite) scores on task

	Participant	English nonword repetition	Finnish nonword repetition	English language tasks	Finnish language tasks
Cluster A	*1*	14	17	87.94	44
	2	14	15	87.04	42
	3	13	17	94.74	43
Cluster B	*4*	8	9	42.56	22
	5	9	10	53.82	25
	6	7	12	62.41	26
	7	13	10	49.37	31
Cluster C	*8*	11	18	63.15	36
	9	10	15	66.50	39
	10	8	15	48.90	39
Cluster D	*11*	12	15	76.85	26
	12	12	12	78.54	29
	13	14	12	83.82	32
	14	13	14	80.22	32
	15	14	18	88.80	25

JYVÄSKYLÄ STUDIES IN HUMANITIES

1 KOSTIAINEN, EMMA, Viestintä ammattiosaamisen ulottuvuutena. - Communication as a dimension of vocational competence. 305 p. Summary 4 p. 2003.

2 SEPPÄLÄ, ANTTI, Todellisuutta kuvaamassa – todellisuutta tuottamassa. Työ ja koti television ja vähän radionkin uutisissa. - Describing reality – producing reality. Discourses of work and home in television and on a small scale in radio news. 211 p. Summary 3 p. 2003.

3 GERLANDER, MAIJA, Jännitteet lääkärin ja potilaan välisessä viestintäsuhteessa. - Tensions in the doctor-patient communication and relationship. 228 p. Summary 6 p. 2003.

4 LEHIKOINEN, TAISTO, Religious media theory - Understanding mediated faith and christian applications of modern media. - Uskonnollinen mediateoria: Modernin median kristilliset sovellukset. 341 p. Summary 5 p. 2003.

5 JARVA, VESA, Venäläisperäisyys ja ekspressiivisyys suomen murteiden sanastossa. - Russian influence and expressivity in the lexicon of Finnish dialects. 215 p. 6 p. 2003.

6 USKALI, TURO, "Älä kirjoita itseäsi ulos" Suomalaisen Moskovan-kirjeenvaihtajuuden alkutaival 1957–1975. - "Do not write yourself out" The beginning of the Finnish Moscow-correspondency in 1957–1975. 484 p. Summary 4 p. 2003.

7 VALKONEN, TARJA, Puheviestintätaitojen arviointi. Näkökulmia lukioikäisten esiintymis- ja ryhmätaitoihin. - Assessing speech communication skills. Perspectives on presentation and group communication skills among upper secondary school students. 310 p. Summary 7 p. 2003.

8 TAMPERE, KAJA, Public relations in a transition society 1989-2002. Using a stakeholder approach in organisational communications and relation analyses. 137 p. 2003.

9 EEROLA, TUOMAS, The dynamics of musical expectancy. Cross-cultural and statistical approaches to melodic expectations. - Musiikillisten odotusten tarkastelu kulttuurienvälisten vertailujen ja tilastollisten mallien avulla. 84 p. (277 p.) Yhteenveto 2 p. 2003.

10 PAANANEN, PIRKKO, Monta polkua musiikkiin. Tonaalisen musiikin perusrakenteiden kehittyminen musiikin tuottamis- ja improvisaatiotehtävissä ikävuosina 6–11. - Many paths to music. The development of basic structures of tonal music in music production and improvisation at the age of 6–11 years. 235 p. Summary 4 p. 2003.

11 LAAKSAMO, JOUKO, Musiikillisten karakterien metamorfoosi. Transformaatio- ja metamorfoosiprosessit Usko Meriläisen tuotannossa vuosina 1963–86. - "Metamorphosis of musical characters". Transformation and metamorphosis processes in the works of Usko Meriläinen during 1963–86. 307 p. Summary 3 p. 2004.

12 RAUTIO, RIITTA, *Fortspinnungstypus* Revisited. Schemata and prototypical features in J. S. Bach's Minor-Key Cantata Aria Introductions. - Uusi katsaus kehitysmuotoon. Skeemat ja prototyyppiset piirteet J. S. Bachin kantaattien molliaarioiden alkusoitoissa. 238 p. Yhteenveto 3 p. 2004.

13 MÄNTYLÄ, KATJA, Idioms and language users: the effect of the characteristics of idioms on their recognition and interpretation by native and non-native speakers of English. - Idiomien ominaisuuksien vaikutus englannin idiomien ymmärtämiseen ja tulkintaan syntyperäisten ja suomea äidinkielenään puhuvien näkökulmasta. 239 p. Yhteenveto 3 p. 2004.

14 MIKKONEN, YRJÖ, On conceptualization of music. Applying systemic approach to musicological concepts, with practical examples of music theory and analysis. - Musiikin käsitteellistämisestä. Systeemisen tarkastelutavan soveltaminen musikologisiin käsitteisiin sekä käytännön esimerkkejä musiikin teoriasta ja analyysistä. 294 p. Yhteenveto 10 p. 2004.

15 HOLM, JAN-MARKUS, Virtual violin in the digital domain. Physical modeling and model-based sound synthesis of violin and its interactive application in virtual environment. - Virtuaalinen viulu digitaalisella alueella. Viulun fysikaalinen mallintaminen ja mallipohjainen äänisynteesi sekä sen vuorovaikutteinen soveltaminen virtuaalitodellisuus ympäristössä. 74 p. (123 p.) Yhteenveto 1 p. 2004.

16 KEMP, CHRIS, Towards the holistic interpretation of musical genre classification. - Kohti musiikin genreluokituksen kokonaisvaltaista tulkintaa. 302 p. Yhteenveto 1 p. 2004.

17 LEINONEN, KARI, Finlandssvenskt sje-, tje- och s-ljud i kontrastiv belysning. 274 p. Yhteenveto 4 p. 2004.

18 MÄKINEN, EEVA, Pianisti cembalistina. Cembalotekniikka cembalonsoittoa aloittavan pianistin ongelmana. - The Pianist as cembalist. Adapting to harpsichord technique as a problem for pianists beginning on the harpsichord. 189 p. Summary 4 p. 2004.

19 KINNUNEN, MAURI, Herätysliike kahden kulttuurin rajalla. Lestadiolaisuus Karjalassa 1870-1939. - The Conviction on the boundary of two cultures. Laestadianism in Karelia in 1870-1939. 591 p. Summary 9 p. 2004.

20 Лилия Сиберг, "БЕЛЫЕ ЛИЛИИ". ГЕНЕЗИС ФИНСКОГО МИФА В БОЛГАРИИ. РОЛЬ РУССКОГО ФЕННОИЛЬСТВА. ФИНСКО-БОЛГАРСКИЕ КОНТАКТЫ И ПОСРЕДНИКИ С КОНЦА XIX ДО КОНЦА XX ВЕКА. 284 с. - "Belye lilii". Genezis finskogo mifa v Bolgarii. Rol' russkogo fennoil'stva. Finsko-bolgarskie kontakty i posredniki s konca XIX do konca XX veka. 284 p. Yhteenveto 2 p. 2004.

21 FUCHS, BERTOLD, Phonetische Aspekte einer Didaktik der Finnischen Gebärdensprache als Fremdsprache. - Suomalainen viittomakieli vieraana kielenä. Didaktinen fonetiikka. 476 p. Yhteenveto 14 p. 2004.

22 JÄÄSKELÄINEN, PETRI, Instrumentatiivisuus ja nykysuomen verbinjohto. Semanttinen tutkimus. - Instrumentality and verb derivation in Finnish. A semantic study. 504 p. Summary 5 p. 2004.

23 MERTANEN TOMI, Kahdentoista markan kapina? Vuoden 1956 yleislakko Suomessa. - A Rebellion for Twelve Marks? – The General Strike of 1956 in Finland. 399 p. Summary 10 p. 2004.

24 MALHERBE, JEAN-YVES, L'œuvre de fiction en prose de Marcel Thiry : une lecture d'inaboutissements. 353 p. Yhteenveto 1 p. 2004.

25 KUHNA, MATTI, Kahden maailman välissä. Marko Tapion *Arktinen hysteria* Väinö Linnan haastajana. - Between two worlds. Marko Tapio's Arktinen hysteria as a challenger to Väinö Linna. 307p. Summary 2 p. 2004.

26 VALTONEN, HELI, Minäkuvat, arvot ja mentaliteetit. Tutkimus 1900-luvun alussa syntyneiden toimihenkilönaisten omaelämäkerroista. - Self-images, values and mentalities. An autobiographical study of white collar women in twentieth century Finland. 272 p. Summary 6 p. 2004.

27 PUSZTAI, BERTALAN, Religious tourists. Constructing authentic experiences in late modern hungarian catholicism. - Uskontoturistit. Autenttisen elämyksen rakentaminen myöhäismodernissa unkarilaisessa katolisuudessa. 256 p. Yhteenveto 9 p. Summary in Hungarian 9 p. 2004.

28 PÄÄJOKI, TARJA, Taide kulttuurisena kohtaamispaikkana taidekavatuksessa. - The arts as a place of cultural encounters in arts education. 125 p. Summary 3 p. 2004.

29 JUPPI, PIRITA, "Keitä me olemme? Mitä me haluamme?" Eläinoikeusliike määrittelykamppailun, marginalisoinnin ja moraalisen paniikin kohteena suomalaisessa sanomalehdistössä. - "Who are we? What do we want?" The animal rights movement as an object of discursive struggle, marginalization and moral panic in Finnish newspapers. 315 p. Summary 6 p. 2004.

30 HOLMBERG, JUKKA, Etusivun politiikkaa. Yhteiskunnallisten toimijoiden representointi suomalaisissa sanomalehtiuutisissa 1987–2003. - Front page politics. Representation of societal actors in Finnish newspapers' news articles in 1987-2003. 291 p. Summary 2 p. 2004.

31 LAGERBLOM, KIMMO, Kaukana Kainuussa, valtaväylän varrella. Etnologinen tutkimus Kontiomäen rautatieläisyhteisön elinkaaresta 1950 – 1972. - Far, far away, nearby a main passage. An ethnological study of the life spans of Kontiomäki railtown 1950 – 1972. 407 p. Summary 2 p. 2004.

32 HAKAMÄKI, LEENA, Scaffolded assistance provided by an EFL teacher during whole-class interaction. - Vieraan kielen opettajan antama oikea-aikainen tuki luokkahuoneessa. 331 p. Yhteenveto 7 p. 2005.

33 VIERGUTZ, GUDRUN, Beiträge zur Geschichte des Musikunterrichts an den Gelehrtenschulen der östlichen Ostseeregion im 16. und 17. Jahrhundert. - Latinankoulujen musiikinopetuksen historiasta itäisen Itämeren rannikkokaupungeissa 1500- ja 1600-luvuilla. 211 p. Yhteenveto 9 p. 2005.

34 NIKULA, KAISU, Zur Umsetzung deutscher Lyrik in finnische Musik am Beispiel Rainer Maria Rilke und Einojuhani Rautavaara. - Saksalainen runous suomalaisessa musiikissa, esimerkkinä Rainer Maria Rilke ja Einojuhani Rautavaara. 304 p. Yhteenveto 6 p. 2005.

35 SYVÄNEN, KARI, Vastatunteiden dynamiikka musiikkiterapiassa. - Counter emotions dynamics in music therapy. 186 p. Summary 4 p. 2005.

36 ELORANTA, JARI & OJALA, JARI (eds), East-West trade and the cold war. 235 p. 2005.

37 HILTUNEN, KAISA, Images of time, thought and emotions: Narration and the spectator's experience in Krzysztof Kieslowski 's late fiction films. - Ajan, ajattelun ja tunteiden kuvia. Kerronta ja katsojan kokemus Krzysztof Kieslowskin myöhäisfiktiossa. 203 p. Yhteenveto 5 p. 2005.

38 AHONEN, KALEVI, From sugar triangle to cotton triangle. Trade and shipping between America and Baltic Russia, 1783-1860. 572 p. Yhteenveto 9 p. 2005.

39 UTRIAINEN, JAANA, A gestalt music analysis. Philosophical theory, method, and analysis of Iegor Reznikoff's compositions. - Hahmoperustainen musiikkianalyysi. Hahmofilosofinen teoria, metodi ja musiikkianalyysi Iégor Reznikoffin sävellyksistä. 222 p. Yhteenveto 3 p. 2005.

40 MURTORINNE, ANNAMARI, *Tuskan hauskaa*! Tavoitteena tiedostava kirjoittaminen. Kirjoittamisprosessi peruskoulun yhdeksännellä luokalla. - Painfully fun! Towards reflective writing process. 338 p. 2005.

41 TUNTURI, ANNA-RIITTA, Der Pikareske Roman als Katalysator in Geschichtlichen Abläufen. Erzählerische Kommunikationsmodelle in *Das Leben des Lazarillo von Tormes*, bei Thomas Mann und in EinigenFinnischen Romanen. 183 p. 2005.

42 LUOMA-AHO, VILMA, Faith-holders as Social Capital of Finnish Public Organisations. - Luottojoukot – Suomalaisten julkisten organisaatioiden sosiaalista pääomaa. 368 p. Yhteenveto 8 p. 2005.

63 RUUSKANEN, ESA, Viholliskuviin ja viranomaisiin vetoamalla vaiennetut työväentalot. Kuinka Pohjois-Savon Lapuan liike sai nimismiehet ja maaherran sulkemaan 59 kommunistista työväentaloa Pohjois-Savossa vuosina 1930–1932. - The workers' halls closed by scare-mongering and the use of special powers by the authorities. 248 p. Summary 5 p. 2006.

64 VARDJA, MERIKE, Tegelaskategooriad ja tegelase kujutamise vahendid Väinö Linna romaanis "Tundmatu sõdur". - Character categories and the means of character representation in Väinö Linna's Novel *The Unknown Soldier*. 208 p. Summary 3 p. 2006.

65 TAKÁTS, JÓZSEF, Módszertani berek. Írások az irodalomtörténet-írásról. - The Grove of Methodology. Writings on Literary Historiography. 164 p. Summary 3 p. 2006.

66 MIKKOLA, LEENA, Tuen merkitykset potilaan ja hoitajan vuorovaikutuksessa. - Meanings of social support in patient-nurse interaction. 260 p. Summary 3 p. 2006.

67 SAARIKALLIO, SUVI, Music as mood regulation in adolescence. - Musiikki nuorten tunteiden säätelynä. 46 p. (119 p.) Yhteenveto 2 p. 2007.

68 HUJANEN, ERKKI, Lukijakunnan rajamailla. Sanomalehden muuttuvat merkitykset arjessa. - On the fringes of readership. The changing meanings of newspaper in everyday life. 296 p. Summary 4 p. 2007.

69 TUOKKO, EEVA, Mille tasolle perusopetuksen englannin opiskelussa päästään? Perusopetuksen päättövaiheen kansallisen arvioinnin 1999 eurooppalaisen viitekehyksen taitotasoihin linkitetyt tulokset. - What level do pupils reach in English at the end of the comprehensive school? National assessment results linked to the common European framework. 338 p. Summary 7 p. Sammanfattning 1 p. Tiivistelmä 1 p. 2007.

70 TUIKKA, TIMO, "Kekkosen konstit". Urho Kekkosen historia- ja politiikkakäsitykset teoriasta käytäntöön 1933–1981. - "Kekkonen´s way". Urho Kekkonen's conceptions of history and politics from theory to practice, 1933–1981 413 p. Summary 3 p. 2007.

71 Humanistista kirjoa. 145 s. 2007.

72 NIEMINEN, LEA, A complex case: a morphosyntactic approach to complexity in early child language. 296 p. Tiivistelmä 7 p. 2007.

73 TORVELAINEN, PÄIVI, Kaksivuotiaiden lasten fonologisen kehityksen variaatio. Puheen ymmärrettävyyden sekä sananmuotojen tavoittelun ja tuottamisen tarkastelu. - Variation in phonological development of two-year-old Finnish children. A study of speech intelligibility and attempting and production of words. 220 p. Summary 10 p. 2007.

74 SIITONEN, MARKO, Social interaction in online multiplayer communities. - Vuorovaikutus verkkopeliyhteisöissä. 235 p. Yhteenveto 5 p. 2007.

75 STJERNVALL-JÄRVI, BIRGITTA, Kartanoarkkitehtuuri osana Tandefelt-suvun elämäntapaa. - Manor house architecture as part of the Tandefelt family´s lifestyle. 231 p. 2007.

76 SULKUNEN, SARI, Text authenticity in international reading literacy assessment. Focusing on PISA 2000. - Tekstien autenttisuus kansainvälisissä lukutaidon arviointitutkimuksissa: PISA 2000. 227 p. Tiivistelmä 6 p. 2007.

77 KŐSZEGHY, PÉTER, Magyar Alkibiadés. Balassi Bálint élete. - The Hungarian Alcibiades. The life of Bálint Balass. 270 p. Summary 6 p. 2007.

78 MIKKONEN, SIMO, State composers and the red courtiers - Music, ideology, and politics in the Soviet 1930s - Valtion säveltäjiä ja punaisia hoviherroja. Musiikki, ideologia ja politiikka 1930-luvun Neuvostoliitossa. 336 p. Yhteenveto 4 p. 2007.

79 SIVUNEN, ANU, Vuorovaikutus, viestintäteknologia ja identifioituminen hajautetuissa tiimeissä. - Social interaction, communication technology and identification in virtual teams. 251 p. Summary 6 p. 2007.

80 LAPPI, TIINA-RIITTA, Neuvottelu tilan tulkinnoista. Etnologinen tutkimus sosiaalisen ja materiaalisen ympäristön vuorovaikutuksesta jyväskyläläisissä kaupunkipuhunnoissa. - Negotiating urban spatiality. An ethnological study on the interplay of social and material environment in urban narrations on Jyväskylä. 231 p. Summary 4 p. 2007.

81 HUHTAMÄKI, ULLA, "Heittäydy vapauteen". Avantgarde ja Kauko Lehtisen taiteen murros 1961–1965. - "Fling yourself into freedom!" The Avant-Garde and the artistic transition of Kauko Lehtinen over the period 1961–1965. 287 p. Summary 4 p. 2007.

82 KELA, MARIA, *Jumalan kasvot* suomeksi. Metaforisaatio ja erään uskonnollisen ilmauksen synty. - God's face in Finnish. Metaphorisation and the emergence of a religious expression. 275 p. Summary 5 p. 2007.

83 SAARINEN, TAINA, Quality on the move. Discursive construction of higher education policy from the perspective of quality. - Laatu liikkeessä. Korkeakoulupolitiikan diskursiivinen rakentuminen laadun näkökulmasta. 90 p. (176 p.) Yhteenveto 4 p. 2007.

84 MÄKILÄ, KIMMO, Tuhoa, tehoa ja tuhlausta. Helsingin Sanomien ja New York Timesin ydinaseuutisoinnin tarkastelua diskurssianalyyttisesta näkökulmasta 1945–1998.

- "Powerful, Useful and Wasteful". Discourses of Nuclear Weapons in the New York Times and Helsingin Sanomat 1945–1998. 337 p. Summary 7 p. 2007.

85 KANTANEN, HELENA, Stakeholder dialogue and regional engagement in the context of higher education. - Yliopistojen sidosryhmävuoropuhelu ja alueellinen sitoutuminen. 209 p. Yhteenveto 8 p. 2007.

86 ALMONKARI, MERJA, Jännittäminen opiskelun puheviestintätilanteissa. - Social anxiety in study-related communication situations. 204 p. Summary 4 p. 2007.

87 VALENTINI, CHIARA, Promoting the European Union. Comparative analysis of EU communication strategies in Finland and in Italy. 159 p. (282 p.) 2008.

88 PULKKINEN, HANNU, Uutisten arkkitehtuuri - Sanomalehden ulkoasun rakenteiden järjestys ja jousto. - The Architecture of news. Order and flexibility of newspaper design structures. 280 p. Yhteenveto 5 p. 2008.

89 MERILÄINEN, MERJA, Monenlaiset oppijat englanninkielisessä kielikylpyopetuksessa - rakennusaineita opetusjärjestelyjen tueksi. - Diverse Children in English Immersion: Tools for Supporting Teaching Arrangements. 197 p. 2008.

90 VARES, MARI, The question of Western Hungary/Burgenland, 1918-1923. A territorial question in the context of national and international policy. - Länsi-Unkarin/Burgenlandin kysymys 1918–1923. Aluekysymys kansallisen ja kansainvälisen politiikan kontekstissa. 328 p. Yhteenveto 8 p. 2008.

91 ALA-RUONA, ESA, Alkuarviointi kliinisenä käytäntönä psyykkisesti oireilevien asiakkaiden musiikkiterapiassa – strategioita, menetelmiä ja apukeinoja. – Initial assessment as a clinical procedure in music therapy of clients with mental health problems – strategies, methods and tools. 155 p. 2008.

92 ORAVALA, JUHA, Kohti elokuvallista ajattelua. Virtuaalisen todellisen ontologia Gilles Deleuzen ja Jean-Luc Godardin elokuvakäsityksissä. - Towards cinematic thinking. The ontology of the virtually real in Gilles Deleuze's and Jean-Luc Godard's conceptions of cinema. 184 p. Summary 6 p. 2008.

93 KECSKEMÉTI, ISTVÁN, Papyruksesta megabitteihin. Arkisto- ja valokuvakokoelmien konservoinnin prosessin hallinta. - From papyrus to megabytes: Conservation management of archival and photographic collections. 277 p. 2008.

94 SUNI, MINNA, Toista kieltä vuorovaikutuksessa. Kielellisten resurssien jakaminen toisen kielen omaksumisen alkuvaiheessa. - Second language in interaction: sharing linguistic resources in the early stage of second language acquisition. 251 p. Summary 9 p. 2008.

95 N. PÁL, JÓZSEF, Modernség, progresszió, Ady Endre és az Ady–Rákosi vita. Egy konfliktusos eszmetörténeti pozíció természete és következményei. 203 p. Summary 3 p. 2008.

96 BARTIS, IMRE, „Az igazság ismérve az, hogy igaz". Etika és nemzeti identitás Sütő András Anyám könnyű álmot ígér című művében és annak recepciójában. 173 p. Summary 4 p. 2008.

97 RANTA-MEYER, TUIRE, Nulla dies sine linea. Avauksia Erkki Melartinin vaikutteisiin, verkostoihin ja vastaanottoon henkilö- ja reseptiohistoriallisena tutkimuksena. - *Nulla dies sine linea*: A biographical and reception-historical approach to Finnish composer Erkki Melartin. 68 p. Summary 6 p. 2008.

98 KOIVISTO, KEIJO, Itsenäisen Suomen kanta-aliupseeriston synty, koulutus, rekrytointitausta ja palvelusehdot. - The rise, education, the background of recruitment and conditions of service of the non-commissioned officers in independent Finland. 300 p. Summary 7 p. 2008.

99 KISS, MIKLÓS, Between narrative and cognitive approaches. Film theory of non-linearity applied to Hungarian movies. 198 p. 2008.

100 RUUSUNEN, AIMO, Todeksi uskottua. Kansandemokraattinen Neuvostoliitto-journalismi rajapinnan tulkkina vuosina1964–1973. - Believed to be true. Reporting on the USSR as interpretation of a boundary surface in pro-communist partisan journalism 1964–1973. 311 p. Summary 4 p. 2008.

101 HÄRMÄLÄ, MARITA, Riittääkö *Ett ögonblick* näytöksi merkonomilta edellytetystä kielitaidosta? Kielitaidon arviointi aikuisten näyttötutkinnoissa. – Is *Ett ögonblick* a sufficient demonstration of the language skills required in the qualification of business and administration? Language assessment in competence-based qualifications for adults. 318 p. Summary 4 p. 2008.

102 COELHO, JACQUES, The vision of the cyclops. From painting to video ways of seeing in the 20th century and through the eyes of Man Ray. 538 p. 2008.

103 BREWIS, KIELO, Stress in the multi-ethnic customer contacts of the Finnish civil servants: Developing critical pragmatic intercultural professionals. – Stressin kokemus suomalaisten viranomaisten monietnisissä asiakaskontakteissa: kriittis-pragmaattisen kulttuurienvälisen ammattitaidon kehittäminen. 299 p. Yhteenveto 4 p. 2008.

104 BELIK, ZHANNA, The Peshekhonovs' Workshop: The Heritage in Icon Painting. 239 p. [Russian]. Summary 7 p. 2008.

105 MOILANEN, LAURA-KRISTIINA, Talonpoikaisuus, säädyllisyys ja suomalaisuus 1800- ja 1900-lukujen vaihteen suomenkielisen proosan kertomana. – Peasant values, estate society and the Finnish in late nineteenth- and early

and early twentieth-century narrative literature. 208 p. Summary 3 p. 2008.

106 PÄÄRNILÄ, OSSI, Hengen hehkusta tietostrategioihin. Jyväskylän yliopiston humanistisen tiedekunnan viisi vuosikymmentä. 110 p. 2008.

107 KANGASNIEMI, JUKKA, Yksinäisyyden kokemisen avainkomponentit Yleisradion tekstitelevision Nuorten palstan kirjoituksissa. - The key components of the experience of loneliness on the Finnish Broadcasting Company's (YLE) teletext forum for adolescents. 388 p. 2008.

108 GAJDÓ, TAMÁS, Színháztörténeti metszetek a 19. század végétől a 20. század közepéig. - Segments of theatre history from the end of the 19th century to the middle of the 20th century. 246 p. Summary 2 p. 2008.

109 CATANI, JOHANNA, Yritystapahtuma kontekstina ja kulttuurisena kokemuksena. - Corporate event as context and cultural experience. 140 p. Summary 3 p. 2008.

110 MAHLAMÄKI-KAISTINEN, RIIKKA, Mätänevän velhon taidejulistus. Intertekstuaalisen ja -figuraalisen aineiston asema Apollinairen L'Enchanteur pourrissant teoksen tematiikassa ja symboliikassa. - Pamphlet of the rotten sorcerer. The themes and symbols that intertextuality and interfigurality raise in Apollinaire's prose work L'Enchanteur pourrissant. 235 p. Résumé 4 p. 2008.

111 PIETILÄ, JYRKI, Kirjoitus, juttu, tekstielementti. Suomalainen sanomalehtijournalismi juttutyyppien kehityksen valossa printtimedian vuosina 1771-2000. - Written Item, Story, Text Element. Finnish print journalism in the light of the development of journalistic genres during the period 1771-2000. 779 p. Summary 2 p. 2008.

112 SAUKKO, PÄIVI, Musiikkiterapian tavoitteet lapsen kuntoutusprosessissa. - The goals of music therapy in the child's rehabilitation process. 215 p. Summary 2 p. 2008.

113 LASSILA-MERISALO, MARIA, Faktan ja fiktion rajamailla. Kaunokirjallisen journalismin poetiikka suomalaisissa aikakauslehdissä. - On the borderline of fact and fiction. The poetics of literary journalism in Finnish magazines. 238 p. Summary 3 p. 2009.

114 KNUUTINEN, ULLA, Kulttuurihistoriallisten materiaalien menneisyys ja tulevaisuus. Konservoinnin materiaalitutkimuksen heritologiset funktiot. - The heritological functions of materials research of conservation. 157 p. (208 p.) 2009.

115 NIIRANEN, SUSANNA, «Miroir de mérite». Valeurs sociales, rôles et image de la femme dans les textes médiévaux des *trobairitz*. - "Arvokkuuden peili". Sosiaaliset arvot, roolit ja naiskuva keskiaikaisissa *trobairitz*-teksteissä. 267 p. Yhteenveto 4 p. 2009.

116 ARO, MARI, Speakers and doers. Polyphony and agency in children's beliefs about language learning. - Puhujat ja tekijät. Polyfonia ja agentiivisuus lasten kielenoppimiskäsityksissä. 184 p. Yhteenveto 5 p. 2009.

117 JANTUNEN, TOMMI, Tavu ja lause. Tutkimuksia kahden sekventiaalisen perusyksikön olemuksesta suomalaisessa viittomakielessä. - Syllable and sentence. Studies on the nature of two sequential basic units in Finnish Sign Language. 64 p. 2009.

118 SÄRKKÄ, TIMO, Hobson's Imperialism. A Study in Late-Victorian political thought. - J. A. Hobsonin imperialismi. 211 p. Yhteenveto 11 p. 2009.

119 LAIHONEN, PETTERI, Language ideologies in the Romanian Banat. Analysis of interviews and academic writings among the Hungarians and Germans. 51 p. (180 p) Yhteenveto 3 p. 2009.

120 MÁTYÁS, EMESE, Sprachlernspiele im DaF-Unterricht. Einblick in die Spielpraxis des finnischen und ungarischen Deutsch-als-Fremdsprache-Unterrichts in der gymnasialen Oberstufe sowie in die subjektiven Theorien der Lehrenden über den Einsatz von Sprachlernspielen. 399 p. 2009.

121 PARACZKY, ÁGNES, Näkeekö taitava muusikko sen minkä kuulee? Melodiadiktaatin ongelmat suomalaisessa ja unkarilaisessa taidemusiikin ammattikoulutuksessa. - Do accomplished musicians see what they hear? 164 p. Magyar nyelvü összefoglaló 15 p. Summary 4 p. 2009.

122 ELOMAA, EEVA, Oppikirja eläköön! Teoreettisia ja käytännön näkökohtia kielten oppimateriaalien uudistamiseen. - Cheers to the textbook! Theoretical and practical considerations on enchancing foreign language textbook design. 307 p. Zusammanfassung 1 p. 2009.

123 HELLE, ANNA, Jäljet sanoissa. Jälkistrukturalistisen kirjallisuuskäsityksen tulo 1980-luvun Suomeen. - Traces in the words. The advent of the poststructuralist conception of literature to Finland in the 1980s. 272 p. Summary 2 p. 2009.

124 PIMIÄ, TENHO ILARI, Tähtäin idässä. Suomalainen sukukansojen tutkimus toisessa maailmansodassa. - Setting sights on East Karelia: Finnish ethnology during the Second World War. 275 p. Summary 2 p. 2009.

125 VUORIO, KAIJA, Sanoma, lähettäjä, kulttuuri. Lehdistöhistorian tutkimustraditiot Suomessa ja median rakennemuutos. - Message, sender, culture. Traditions of research into the history of the press in Finland and structural change in the media. 107 p. 2009.

126 BENE, ADRIÁN Egyén és közösség. Jean-Paul Sartre *Critique de la raison dialectique* című műve a magyar recepció tükrében. - Individual and community. Jean-Paul Sartre's