Marta Marecka

The relationship between phonological processing and lexical acquisition in a foreign language. A study on Polish primary school students learning English.

Związki pomiędzy przetwarzaniem fonologicznym, a akwizycją leksykalną w języku obcym. Badanie na polskich uczniach szkoły podstawowej uczących się języka angielskiego.

Praca doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza w Poznaniu
pod kierunkiem prof. zw. Katarzyny Dziubalskiej-Kołaczyk i dr Karoliny Rataj

Poznań, 2015

ACKNOWLEDGEMENTS

I want to thank my supervisors - Katarzyna Dziubalska-Kołaczyk and Karolina Rataj - for giving me intellectual freedom and supporting me in my pursuits. I am grateful for their patience, for their guidance, and for the inspiration they offered me.

I owe an enormous deal to Kuba Szewczyk, with whom I had numerous intellectually stimulating conversations. Much of the theoretical content in this thesis has been inspired by those conversations. Moreover, I am grateful to Kuba for his help with the statistical analysis of the study and with the design of the stimuli.

I am grateful to Donata Janiszewska and Ania Jelec for helping me conduct the study and for being a professional, reliable and awesome research team.

I want to thank Michał Jankowski for poring over corpora and helping me design the non-words for the study.

I am grateful to my friends who helped me record the stimuli - Konrad Juszczyk, Gosia Lauks, Gosia Góraj, Halszka Bąk, Agnieszka Czoska, Ron Kim, and Laura Hawbaker.

I want to thank Beata Maj for creating lovely pictures for the vocabulary test, Ania Basińska for helping me organise the pilot study for this research, Agnieszka Otwinowska-Kasztelanic for her useful comments, Tim Fosker for his kindness and encouragement.

I am grateful to my participants and the schools that allowed me to conduct this study.

I want to thank the organisations that supported this research: National Science Centre (grant "The interactions between phonological processing and lexical acquisition in a foreign language. A longitudinal study on Polish primary school students learning English", 2012/07/N/HS6/00415), European Social Fund, which sponsored my PhD programme (UDA- POKL.04.01.01-00-019/10-00) and the PRO-ED publishing company, which has provided me with a free kit of the Comprehensive Test of Phonological Processing 2 for the research purposes.

Finally, I want to thank Szymon Butkiewicz, my friends from the LST&C PhD Programme (Halszka, Marta, Michał, Paula, Rafał) and my family for being there for me.

Poznań, dnia ....

OŚWIADCZENIE

Ja, niżej podpisany/a .... student/ka Wydziału Anglistyki Uniwersytetu im. Adama Mickiewicza w Poznaniu oświadczam, że przedkładaną pracę dyplomową pt: .... napisałem/napisałam samodzielnie. Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałem/am z pomocy innych osób, a w szczególności nie zlecałem/am opracowania rozprawy lub jej części innym osobom, ani nie odpisywałem/am tej rozprawy lub jej części od innych osób.Oświadczam również, że egzemplarz pracy dyplomowej w formie wydruku komputerowego jest zgodny z egzemplarzem pracy dyplomowej w formie elektronicznej.

Jednocześnie przyjmuję do wiadomości, że przypisanie sobie, w pracy dyplomowej, autorstwa istotnego fragmentu lub innych elementów cudzego utworu lub ustalenia naukowego stanowi podstawę stwierdzenia nieważności postępowania w sprawie nadania tytułu zawodowego.

[ ]* - wyrażam zgodę na udostępnianie mojej pracy w czytelni Archiwum UAM

[ ]* - wyrażam zgodę na udostępnianie mojej pracy w zakresie koniecznym do ochrony mojego prawa do autorstwa lub praw osób trzecich

*Należy wpisać TAK w przypadku wyrażenia zgody na udostępnianie pracy w czytelni Archiwum UAM, NIE w przypadku braku zgody. Niewypełnienie pola oznacza brak zgody na udostępnianie pracy.

....

Table of contents

List of tables

List of figures

Introduction

It is often the case that different facets of human cognition and language are treated as separate entities. There are researchers specialising in studying phonological systems, experts on mental lexicon and scientists whose research is devoted solely to memory systems. This kind of specialisation is necessary for managing the information about the vast and complicated systems of human cognition. Thanks to narrowing down their subject of investigation, these researchers managed to amass a great body of knowledge. However, the danger of these approaches is that they disregard the context in which human cognition and language operate. Language is a product of the brain - a system based on connections and associations. It is intimately connected to cognitive systems. Researching the interconnections between different aspects of language (such as phonology, syntax or lexicon) and applying the knowledge amassed by psychologists and neuroscientists in this research can significantly further the understanding of human linguistic processing. The present dissertation is devoted to the interconnections between phonology, memory and word learning. In particular, it focuses on the interconnection between sound processing and vocabulary acquisition. As will be shown, there is ample evidence for a close relationship between the acquisition of words and sounds in human minds. Apart from exploring these connections, this dissertation will also investigate the possible interplay between the phonological processing and phonological short-term memory in learning new words.

The thesis begins with an introductory chapter that presents the mainstream theories related to the interconnections between memory, sounds and words. Within these theories, memory is the key concept. Those theories assume the existence of a specialised memory component, which is crucial for learning novel words. This component is called phonological short-term memory (henceforth phonological STM), since it used precisely for encoding phonological forms of the words. However, as will be shown at the end of the chapter, the theory of phonological short-term memory as a vocabulary learning device is not without problems. Indeed, some researchers suggest that some tests of phonological short-term memory might in fact tap into another skill - the phonological processing. Within these theories it is phonological processing that is central to learning new words. The rest of the thesis explores and significantly extends this theory.

First of all, it is observed that the concept of phonological processing is in itself a subject to considerable controversy. Throughout the years different and often contradictory ideas about phonological processing have been offered and some theories even deny the existence of the mechanism. Therefore, the second chapter will constitute an attempt to analyse different theories of phonological processing and to distil from them a coherent description of the concept. The chapter thus ends with a theoretical framework within which the notion of phonological processing will be considered.

Equipped with the working definition and description of the key notions, the reader will be directed to the third chapter, in which the relationship between phonological processing and word learning in first language (L1) acquisition will be explored. The chapter will begin with a presentation of studies related to the acquisition of phonology in the L1 - a description of how phonological processing develops in young children. Since different theoretical approaches offer different visions as to the shape of phonological development, this chapter will attempt to provide a unified account of the process. In particular, it will be suggested that phonological development in L1 consists of several different processes - notably initial prosodic processing, abstraction, specification and chunking. The chapter will end by showing how each of these processes is related to the lexical development in L1 acquisition.

The fourth chapter will continue to tackle the topic of phonological development and word learning, but it will describe them in the context of second language (L2) acquisition. This chapter will begin with the description of phonological development in L2, using the previously introduced notions of initial prosodic processing, specification, abstraction and chunking. Then it will show how these processes could be related to word learning in L2. It will be noted, however, that there is a severe lack of research that would support these hypotheses.

All the four chapters will prepare the reader for the experimental part of the thesis, which ties the three areas (phonology, lexicon and memory) into one study, described in the fifth chapter. The study investigates a group of Polish nine-year-olds learning a second language (English) at school. It examines how fast the children learn new L1, L2 and completely foreign word forms in an experimental tasks. It also tracks the participants' progress in their English classes in terms of vocabulary. These measures of word learning efficiency are then correlated with participants' phonological processing skills and phonological short-term memory.

The results of the study, which are discussed in the sixth chapter, suggest that the some aspects of phonological processing (initial prosodic processing) play an important role in learning words of a foreign language at the initial stages of acquisition. At the same time, phonological STM seems to play a role in learning new words of the native language. These results clearly show the interdependence of different cognitive and linguistic systems in the human mind.

Chapter 1: Memory for words, memory for sounds. The relationship between phonological STM and word learning

1.1. Introduction

This chapter will introduce the topic of memory, phonology and word learning. In particular it will focus on the relationship between the memory module called phonological STM and vocabulary acquisition in L1 and L2. In order to allow for a better understanding of the topic, the chapter will begin with a short introduction into the theories of memory systems that will provide a context for the phonological STM theory. In the second section, the concept of phonological STM will be further developed and the relationship between this memory component and word learning in L1 will be explored. The third section will focus on the relationship between phonological STM and word learning in L2. The chapter will end with a critique of the studies investigating the phonological STM-vocabulary relationship. It will be hypothesised that some research results taken as the evidence for this relationship can be attributed to the effects of phonological processing on lexical development. With that, the reader will be smoothly directed to the next chapter, which will introduce the topic of phonological processing.

1.2. Memory systems

1.2.1. Multicomponent models: Atkinson's and Shiffrin's memory model, Baddeley's Multicomponent Working Memory Model, Long-term memory models

The notion of phonological STM cannot be understood without reference to the classic model of memory introduced by Atkinson and Shiffrin's model (Atkinson and Shiffrin 1968). According to this model, human memory is divided into three different components: sensory memory, short-term memory and long-term memory. Each of these components is connected with a different stage of memory processing. The sensory memory is the first contact point between the stimuli and human memory system. This memory module takes in a large number of raw sensory information and stores it for a very short amount of time (no more than few seconds) for further analysis. Most information stored in the sensory memory is subjected to fast decay. However, the most relevant information is then carried into the short-term memory, where it is held for further analysis for up to 30 seconds. Short-term memory is limited not only in time, but also in capacity - it cannot store more than 5-9 items at the same time (Miller 1956). A significant portion of information in the short-term memory decays, but some of it can be carried into the long-term memory, which according to Atkinson and Shiffrin, has unlimited capacity and provides relatively permanent storage for information.

The model by Atkinson and Shiffrin still constitutes the popular point of reference for many memory studies, but it has also been criticised by many researchers, and the criticisms have often been the basis of improvements and developments. One of the criticism has been put forward by Baddeley and Hitch, who argued with the notion of unitary short-term memory and proposed that it should be replaced with the concept of multi-component working memory (Baddeley and Hitch 1975). The working memory model by Baddeley and Hitch was initially composed of three separate components: phonological loop, visuospatial sketchpad and the central executive. Phonological loop was conceptualised as a short-term memory store for verbal information, such as lists of words, sentences or letters. Visuospatial sketchpad was considered to be a store for visual and spatial information, for instance shapes, colours and location of objects. The central executive was the element responsible for controlling working memory, assigning attentional resources to particular stimuli and for dividing attention. Twenty five years after the multi-component model of working memory was introduced, it was further enhanced by adding another component, the episodic buffer. The role of this additional component is to integrate information from the verbal store with the information from the visuospatial sketchpad (Baddeley 2000). Moreover, since Baddeley in his theoretic approach still assumed the strict separation between the working memory and the long-term memory, the episodic buffer was conceptualised as the point of connection between these two components. It was a centre in which the information from the working memory and long-term memory could be integrated. To recap, the development of the classic memory model by Baddeley and his co-workers consisted in replacing the notion of unitary short-term memory with the concept of multicomponent memory system, containing a verbal store, a visual store, a central executive and an episodic buffer.

The perception of unitary long-term memory was also soon challenged. First of all, it has been discovered that there are different long-term memory systems for conscious knowledge of facts (the declarative memory) and the unconscious learning of skills and procedures (Cohen and Squire 1980). Patients with amnesia who have a deficit in declarative memory (for instance, a lesion to the medial-temporal part of their brain) cannot consciously learn facts, but they can learn skills, such as reading mirror images of texts. The declarative memory has been then broken down into further components (Tulving 1985). One of these is semantic memory, which is responsible for the acquisition, storage and retrieval of facts (such as facts learned at school). The other is episodic memory, which is the memory for events in one's life. Emotionally loaded facts can be also considered a separate type of memory, because an additional part of the brain (the amygdala) takes part in encoding of these memories and gives them an additional boost over the non-emotional stimuli (Gabrieli 1998; Milner et al. 1998).

The unconscious memory has also been divided into different types. In his review, Gabrieli distinguishes between procedural memory, repetition priming and conditioning (Gabrieli 1998). Procedural learning involves learning new skills of different kinds. Thanks to this type of memory, an amnesiac patient with his declarative memory impaired can learn how to ride a bike or how to draw with altered visual feedback (sensorimotor skills learning). Patients with deficits in declarative memory can also learn perceptual skills such as reading mirror-reversed texts, and even cognitive skills, such as the ability to perform a complex cognitive task involving problem-solving and planning (although to a limited degree). On the other hand there can be patients, whose declarative memory is intact, but who have specific deficits in this kind of learning. This includes individuals with neurological diseases such as Parkinson's disease, Huntington's disease or Tourette syndrome, which affect the part of the brain called basal ganglia.

Repetition priming (Milner et al. 1968), the second type of unconscious memory, is defined as the effect of certain stimuli on the following stimuli. For example, when participants hear a word that has negative connotations, they might perceive (that is - be primed to perceive) the next word they hear more negatively. Priming can be supraliminal (when the learner can consciously perceive the words or pictures that are the primes) or subliminal (when the prime has been presented for such a short time that the learner was unable to consciously register it). It can be observed for different modalities (visual, auditory, tactile priming) and it can be based on different types of associations. For instance, there can be phonological and orthographic priming (the word "tribe" inducing a quicker reaction to a word "bribe"), affective priming (negative word triggering negative reaction to the subsequent stimuli), semantic priming (the word "cat" priming the word "dog") etc. Priming seems to be dissociated from declarative memory, because amnesiac patients can be primed. However, there are certain kinds of priming that cannot be performed by individuals with amnesia. For instance, if the patient is shown a pair of two unrelated words (cat-spoon) and then shown one of the words from the pair (cat), this word does not prime the second word from the pair (spoon) in the amnesiac person as it would in a normal person. Overall, it might be concluded that priming is a separate type of memory process, but it can be associated with other types in certain tasks.

The last type of unconscious memory mentioned by Gabrieli (1998) is conditioning. In this type of memory processing, a person is exposed to a stimulus, called conditioned stimulus, which is followed by another stimulus, called unconditioned stimulus. An unconditioned stimulus is one that produces an automatic response (for instance a finger pointed at the eye causes the eye to close) and as a result of conditioning this automatic response becomes associated with the conditioned stimulus. An example of classical conditioning in an experiment in which the learner hears a sound (conditioned stimulus), which is followed by a puff of air to the eye (unconditioned stimulus), triggering an automatic response (blinking). With repeated exposure to this series of stimuli, the participant will start blinking upon hearing the sound, even when it is not followed by the puff of air. This kind of motor learning is not associated with other types of memory like procedural and declarative. Patients with Huntington's disease, who display procedural memory deficits, are prone to this type of conditioning, although for unknown reason this type of learning can be impaired in Alzheimer's disease. More complex types of conditioning seem to be related to the declarative memory. For instance, trace conditioning (Bangasser et al. 2006), in which there is a short pause between the conditioned stimulus (like a sound) and the unconditioned stimulus (like the puff of air), produces a learning effect in people without deficits, but is impaired in amnesiac individuals with deficits to the medial temporal parts of the brain. Yet another type of conditioning, fear conditioning, is associated with the activity of the amygdala brain region, which is also involved in learning emotional stimuli within the declarative memory. Fear conditioning involves pairing a neutral stimulus (a word and a picture) with an unpleasant stimulus like pain or loud bursts of boat-horn. This kind of treatment usually quickly produces an automatic adverse reaction to the neutral stimulus in participants. However, in individuals with amygdala lesions fear conditioning does not work. All in all, conditioning seems to be a complex process that cannot be entirely subsumed under any of the previous types of learning, but can be associated with them.

1.2.2. Unitary memory models

As can be seen from the above review, it seems that memory is a complex system composed of multiple interacting subsystems. The notion that there is one short-term memory and one long-term memory does not seem to be accurate. Some researchers, however, have gone as far as suggesting that even the division into short-term and long-term memory is wrong. Instead they propose what is called unitary models of memory (Cowan 1988; Jonides et al. 2008).

A very special case of the unitary memory model is the so-called Levels-of-processing Theory. This model assumes that instead of short-term and long-term stores there are different levels of information storage in memory, which are closely related to how deeply a given stimulus has been processed (Craik and Lockhart 1972). The theory is based on the assumption that analysing information and learning this information is essentially the same process. Learners memorise a particulate item (for example, a sentence) when they process this item at a number of levels. Some of these levels of processing are rather superficial. For example, during the analysis of a sentence, such a superficial level of processing would involve assessing the loudness or pitch of the stimuli and other physical characteristics. On further levels, learners might start to analyse the patterns in the stimulus. For example, if they analyse a sentence, they might process its phonological or grammatical structure. On the deepest levels the stimulus is analysed conceptually. This means that the learners would analyse the meaning of the sentence. According to the model, the deeper the level of analysis performed on the stimulus, the better and more durable the memory of it will be. A sentence processed on a very deep conceptual level will be better remembered than a sentence that has been analysed only superficially. The theory assumes that the number of repetitions of the stimuli is not as relevant to remembering it as the depth of processing.

There are several studies showing that information that is processed deeper is indeed more robust in the mind. However, Craik and Lockhart's theory is not without problems. It is difficult, for instance, to assess what the depth of processing means and to offer an objective measure of the concept. In fact, so far, there has been no way of measuring the depth of processing apart from asking a group of judges to assess what kind of processing should be considered deep and what kind of processing should be categorised as shallow (Craik 2002). Nevertheless, the level-of-processing models has been one of the most famous and interesting alternatives to the traditional models involving the division into the three separate storage types. While the model assumes different levels of processing, it is unitary in the sense that it does not see short-term memory as a different store of information from the long-term memory. Rather, it assumes that what is called short-term memory is the activation of the information in the long-term memory connected with the perceptual and conceptual aspects of the stimuli. In other words, short-term memory is nothing more than memories that are attended to. This idea is present in one form or another in most unitary models (Jonides et al. 2008). In general, unitary models see short-term memory more in terms of process rather than in terms of store. Baddeley's model conceptualises this memory component as a set of stores from which information can be moved into long-term memory. Unitary model proponents, such as Cowan (Cowan 1988), perceive short-term memory as a state - the activation of particular information in the brain as a result of attention directing. The idea behind unitary models is that there is only one kind of storage but that it can be activated to different degrees. In their review of unitary memory models Jonides and colleagues (2008) conceptualise the long-term memory as dormant information and the short-term memory as active information. The authors support their claim by showing that the same brain regions that are active for the long-term memory storage of information (such as medial-temporal lobe) also take part in the encoding of information in short-term memory tasks. Moreover, the authors quote evidence indicating that the regions that are responsible for the initial perception and encoding of the stimuli are also the regions, in which later the short- and long-term representations of the stimuli are stored (Damasio 1989). In their paper, short-term memory is conceptualised as the firing of neurons in response to the stimuli and the short-term changes in brain structure as a result of this neuronal activation. Long-term memory, on the other hand, are the more permanent changes to the brain structure following the activation of neurons.

The unitary models are more and more popular in the current psychological literature. In their 2009 review, Suprenant and Neath argue that most of the phenomena which have been treated as evidence for the existence of short-term memory can also be interpreted in terms of unitary models (Suprenant and Neath 2009). Nevertheless, at the moment both the unitary and the multicomponent models are considered possible representations of the human memory. In his recent paper, Cowan indicates that both kinds of models can be used to account for the data related, for instance, to the capacity of working memory (Cowan et al. 2012). In other words, the judgement is still out on the best theory describing human memory. Nevertheless, when it comes to research on memory and language learning, it is the non-unitary Baddeley's model that has been used most often as the reference point (Baddeley et al. 1998; Baddeley 2003; Gathercole 2006). Therefore this model will be used as the basis of further considerations in this chapter.

To sum up, the current theories of memory are largely inspired by Atkinson and Shiffrin's model that distinguishes between sensory memory, short-term memory and long-term memory. Further research largely expanded this model. In particular, Alan Baddeley proposed the existence of multicomponent working memory system in place of the unitary short-term memory (Baddeley 2000, 2007; Baddeley and Hitch 1975; Repovs and Baddeley 2006) and several researchers indicated the existence of different subsystems within the long-term memory (Milner et al. 1968; Cohen and Squire 1980;

Tulving 1985; Gabrieli 1998; Milner et al. 1998). A number of psychologists have also challenged the idea of separate short-term (working) and long-term memory components and instead proposed that there is one type of memory. Within these models working/short-term memory is simply the process of activating representations in this unitary memory store (Craik and Lockhart 1972; Craik 2002; Cowan 1988; Jonides et al. 2008; Suprenant and Neath 2009). Nevertheless, since much research on the relationship between memory and language has been done within the separate systems approach, notably, Baddeley's multicomponent model of working memory, this approach will be followed in further sections of this thesis.

1.3. Memory for sounds and words: phonological STM and the acquisition of L1 vocabulary

The previous section introduced the most important models in the current memory research and indicated that some of these models have been used heavily in studies on the relationship between memory and language learning. In particular, Baddeley's multicomponent working memory model has been very often used as a theoretical framework for studies exploring the memory effects on language learning. Most of these studies focus on the working memory component called phonological loop, which, according to Baddeley, is a dedicated "language learning device" (Baddeley et al. 1998; Baddeley 2003). In the subsequent literature this component has been named differently: phonological store (Gathercole 2006), verbal short-term memory (Gupta and MacWhinney 1997), phonological memory (Bowey 1996), etc. To avoid confusion, for the purpose of the current thesis, the name phonological short-term memory (or phonological STM) will be consistently adopted to refer to this memory system.

In Baddeley's working memory model, phonological STM is a specialised component responsible for short-term storage of verbal information (Repovs and Baddeley 2006; Baddeley 2007). The tasks that are used to tap this component usually involve remembering words or non-words (invented words such as *kipser*) for short periods of time and then reporting them back to the researchers. In general, the two tasks that are most often used to tap into this memory system are Immediate Serial Recall (ISR) tasks and non-word repetition tasks. ISR involves remembering and repeating lists of words or non-words. The most classic example of this task is digit span, in which the participant is given longer and shorter lists of digits and is asked to repeat them in the order in which they were presented. There are also variations of this task involving repetition of the words in backward order (from the last items presented to the first item presented). A similar type of task is free recall, in which participants are required to recall the presented items in any order they wish. The non-word repetition tasks do not involve remembering multiple items, but instead require repetition of shorter and longer non-words. These tasks are often used in studies with children.

On the basis of various studies with these tasks, two interesting characteristics of phonological STM have been identified. First of all, it appears that the coding of information in this memory system is based on phonological coding - it is primary the auditory form and not the meaning of the words that is stored. This is evidenced by the so-called phonological similarity effect, word-length effect and articulatory suppression effect. Phonological similarity effect (Conrad and Hull 1964; Baddeley 1966) can be demonstrated very easily with the ISR tasks. Tasks with lists of words or letters that are phonologically similar are much more difficult for the participants than tasks with lists of words that are phonologically dissimilar. This effect is not observed for semantic similarity - i.e. using lists with words that are semantically similar does not affect performance on the ISR tasks (Baddeley 1966). This indicates that phonological STM is devoted to storing form rather than meaning.

The word-length effect can be demonstrated with digit span tasks in bilingual speakers (Ellis and Hennelly 1980; Shebani et al. 2005). In those studies bilingual participants are asked to perform two digit span tasks - one in each of their languages. The participants are selected in such a way that in one of their languages the digit names are much longer than in the other language. The studies have shown that participants recall more digits when they perform the task in the language with shorter digit names. This means that what they store in the phonological STM is not concepts of the digits, but rather their auditory form.

The final piece of evidence for the phonological coding in the phonological STM comes from the so-called articulatory rehearsal studies (Murray 1967; Caplan and Waters 1995; Larsen and Baddeley 2003; Eiter and Inhoff 2010). Those studies indicate that whenever learners store some verbal information in their phonological STM, they use an articulatory rehearsal mechanism to remember this information better. What this means is that, essentially, learners mentally repeat the words to be memorised in their minds so as not to forget them. This mechanism has been demonstrated by a series of experiments using ISR tasks with articulatory suppression, that is tasks, in which participants have to remember lists of words, while at the same time repeating an unrelated sound or word. During ISR tasks with articulatory suppression people remember significantly less words than during normal ISR tasks without articulatory suppression (Murray 1967; Larsen and Baddeley 2003; Eiter and Inhoff 2010). Apart from these studies, compelling evidence for the existence of articulatory rehearsal comes from the case study by Caplan and Waters (2005) investigating the phonological STM in R.W., an aphasic patient with deficits in articulatory planning. In this patient, the articulatory rehearsal was likely suppressed due to his inability to plan articulatory movements. Consequently, his performance on the ISR tasks was severely impaired. Taken together, these studies provide strong evidence for articulatory rehearsal mechanism within the phonological STM and consequently support the claim that phonological STM stores the auditory form of words.

Another interesting characteristic of phonological STM is that is seems crucially involved in vocabulary acquisition. It appears necessary for learning novel word forms. According to Baddeley (Baddeley et al. 1998; Baddeley 2003, 2007), facilitation of novel word form learning is the main function of phonological STM and is the precise reason for calling the memory component, the "language learning device" (Baddeley et al. 1998). One of the first studies supporting this theory has been published in 1989. In this study (Gathercole and Baddeley 1989), 104 children aged 4 and 5 have been tested on their non-verbal intelligence, the size of receptive vocabulary (with British Picture Vocabulary Scale) and phonological STM (with a non-word repetition task). The children were retested on the receptive vocabulary one year later. It turned out that the phonological STM scores were significantly correlated with the vocabulary scores in children both at the time of testing and one year later.

The results of this study have been replicated later on. The replications involved tracing the vocabulary development of 118 children over the period of 4 years (Gathercole et al. 1992). During the initial testing, the participants were 4 to 5 years old. Each year the participants were tested on non-word repetition, expressive and receptive vocabulary size and non-verbal intelligence. At each point in time the non-word repetition task scores correlated with the vocabulary scores. Moreover, the non-word repetition scores collected in the first year of the study were also significant predictors of children's vocabulary size a year later. This, according to the authors, indicates that there is causal relationship between phonological STM performance and word learning. Similar relationships have not been observed in the subsequent years of observation, which lead the researchers to the conclusion that phonological STM influences word learning only at the initial stages of language acquisition. However, the result indicating the causal relationship between phonological STM and vocabulary learning in children aged 4 to 5 should be treated with caution (Melby-Lervåg et al. 2012). This is because in their statistical analyses the researchers did not really take into consideration the increase in vocabulary over the year, but rather the raw vocabulary scores at the second time point. To be more specific, they measured the correlation between non-word repetition at first time point and the vocabulary size at the second time point without controlling for the vocabulary size at the first time point. The reanalysis of the data by Melby-Larvåg and colleagues (2012) shows that when the previous vocabulary scores are controlled for, there is no effect of non-word repetition at the age of four on vocabulary size at the age of five. In other words, on the basis of Gathercole and colleagues' data it is not possible to establish the direction of relationship between phonological STM scores and vocabulary learning.

Nevertheless, there are other, experimental studies that have shown this direction of influence and have in fact indicated that phonological STM plays a role in word learning. Those studies involve experimental word learning paradigms and thus measure the speed of word learning in controlled, laboratory conditions. In one of the first such studies, 118 children aged 5-6 were divided into two groups on the basis of their non-word repetition scores (low repetition group with lower scores on the task and high repetition group with higher scores on the task) (Gathercole and Baddeley 1990). The children were then given word learning tasks, in which they had to learn names of four plastic toys. In one task, the names of the toys were regular English names (Simon, Michael, Peter, Thomas), while in the other tasks the names were non-words created by scrambling the letters of the regular names (Sommel, Meton, Pimas, Tike). The task in both conditions looked the same. The experimenter showed the children the toys one by one and named them. Then the children were shown each toy and were asked to name them themselves. Those presentation and testing procedures were repeated 15 times or until the child named all the toys correctly in two successive trials. The number or trials needed to learn the toys names was then calculated for each child. While children in both groups did similarly well on the task involving learning the familiar English names, the low repetition group was significantly worse than the high repetition group on the task involving learning novel names. The study involved also a delayed memory test, in which the participants were asked to name the toys 24 hours after the initial exposure. Again, the high repetition group outperformed the low repetition group on this task. According to the researchers, this result indicates that phonological STM is involved in learning novel words.

The study was conceptually replicated 7 years later with different sets of learning tasks (Gathercole et al. 1997). In this study, 65 participants aged 5 to 6 were tested on digit span, non-word repetition, non-verbal intelligence, expressive and receptive vocabulary. Then they were also asked to perform two types of non-word learning tasks. In one type of tasks, paired associates, the participants were asked to learn pairs of known words (for instance *donkey-flower*) or pairs consisting of a word and a non-word (for example, *chicken-kipser*). As in the Gathercole and Baddeley (1990) study, the task consisted of interchanging learning phases, in which all pairs were presented to children, and the testing phases, in which the participants were given the cue word from the pair and asked to provide either the word or the non-word that went with the cue. In the second task, called story learning, the children were provided with two novel names that were paired with short definition (for example *foltano - noisy dancing fish*). Following the presentation of the novel names, the children were asked to give the definition for the non-words. Finally, they were given the definition and were asked to provide the non-word that went with it. The results indicate that the recall of the novel names in both the paired associates task and the story learning tasks was correlated with the non-word repetition and the digit span scores. Of the two, the digit span was a stronger predictor. It was associated with novel word learning even when the vocabulary scores or non-word repetition scores were controlled for. The findings of the study, combined with the original study by Gathercole and Baddeley (1990) provide the strongest support for the relationship between phonological STM and new word learning to date.

The final piece of evidence for the involvement of phonological STM in word learning process comes from the brain lesion research. In one study (Gupta et al. 2003), 11 children with normal intelligence and vocabulary size, but with brain lesions affecting the performance on non-word repetition and digit span, were given a word learning task. In the task, the participants were presented with pictures of 9 novel objects and had to learn the names of these objects. The objects were grouped in blocks of three. In each block, there were five cycles of learning and testing phases. During the learning phase the children were presented with the pictures and given the names of the pictures twice. During the testing phase, they were given the pictures and had to provide the names. The results indicate that the brain lesion severely impaired the performance on the word learning task as compared against the control group of 70 children without any brain lesions. This study again provides some evidence for the involvement of phonological STM in the learning process, although it is important to note that the study participants have also been impaired on general linguistic processing, as measured with the Clinical Evaluation of Language Fundamentals. Therefore, it is possible that the word learning deficit cannot be attributed solely to the impairment of phonological STM.

To conclude, the data presented in this section seem to suggest that the phonological STM, i.e. short-term store for verbal information characterised by phonological coding, facilitates novel vocabulary learning in children acquiring their first language. The studies described indicate that phonological coding, memory and vocabulary are intimately linked with each other. The data provided by Gathercole and colleagues (1992) suggest, however, that this relationship might not be as strong in older learners. Even though this data has to be treated with caution, it is worth investigating whether the relationship phonological STM and word learning obtains only for the early stages of human development or whether it can also be observed in adults. In particular, it is interesting to investigate whether phonological STM plays a role in L2 acquisition in older learners. Studies exploring this topic will be described in the next section.

1.4. Memory for different words: phonological STM and L2 vocabulary learning

The previous section reported on several studies on the relationship between phonological STM and word learning in L1. However, there is also substantial evidence that this relationship obtains also for L2 acquisition. One of the first studies to support this is the case study presented by Baddeley, Papagno and Vallar as early as in 1988 (Baddeley et al. 1988). The participant of the study, P.V., was a native speaker of Italian with a very pure phonological STM deficit. P.V. was given a series of word learning tasks. In one, she had to learn eight pairs of words in her own language, in another pairs composed of an Italian and a Russian word. The participant performed within the norm on the task involving learning Italian word pairs, but has shown a clear deficit on the task that involved learning the words in a foreign language. This indicates that phonological STM is of importance for learning words of a foreign language. This would explain why students displaying marked foreign language learning difficulty have a significant deficit in phonological STM (Palladino and Ferrari 2008), while polyglots display superior performance in tasks such as digit span or non-word repetition (Papagno and Vallar 1995). Further support for the idea comes from the data showing that phonological STM (or, to be more specific, English non-word repetition scores) is a good predictor of learning English as a second language (Service 1992). Overall, it seems possible that phonological STM facilitates foreign language learning by helping with the acquisition of novel word forms.

To explore the link between the phonological STM and word learning in a foreign language, a number of studies have been conducted. One of these is the polyglot study by Papagno and Valar (1995), in which a group of 10 polyglots (speakers of at least three languages) and 10 non-polyglots were tested on the measures of phonological STM (non-word repetition and digit span), non-verbal intelligence, native vocabulary size, visuospatial span and visuospatial learning. The participants were also given two experimental word learning tasks in the paired associates paradigm. In one of the tasks the participants had to learn eight pairs of native words and in the other they had to learn eight pairs, each consisting of a native word cue and a Russian non-word target. The polyglots differed from the non-polyglots in terms of phonological STM task scores and the performance on the paired associates task with non-words. Specifically, they learned the foreign words significantly faster than their non-polyglot peers. This study provides some evidence for the relationship between phonological STM and word learning in a foreign language, although the results should be not treated as conclusive, since the two groups also differed probably on the metalinguistic knowledge and linguistic experience and this factor might have been underlying the polyglots' superior performance on both the learning and the phonological STM tasks.

More convincing evidence for the involvement of phonological STM in foreign word learning comes from correlation studies on large populations of L2 learners. One of such studies involved testing 45 of Greek primary school students learning English with the same teacher (Masoura and Gathercole 1999). The children have learned English for an average of 3 years (range 1 - 5 years). In the study, the participants were tested with non-word repetition tasks in English (L2) and in Greek (L1), non-verbal intelligence measures, as well as native and English vocabulary tests. In the English vocabulary test, the students had to translate 60 English words into Greek and 60 Greek words into English. The words were taken from the English handbook that the children used in the classroom. The results of the study indicate that the performance on the foreign vocabulary test was significantly related to the scores on the non-word repetition tests. However, since the study was simple correlation research, it is impossible to establish the direction of influence between those two variables. A similar problem could be observed in another study investigating 41 Finnish children learning English at school (Service and Kohonen 1995). Also here, the correlation between English non-word repetition scales and English vocabulary has been found, but it is impossible to establish the direction of the influence. This problem has been somewhat rectified in further studies which employed the word learning tasks.

In one of those studies (Cheung 1996), 84 12-year-old students from Hong-Kong (native speakers of Cantonese) were tested on the measures of non-verbal intelligence, an ISR task with simple English words and an ISR task with English non-words. The participants also took an English vocabulary test, English reading comprehension task and a paired associates word learning task, in which they had to learn three new English words (*jocular*, *succulent* and *egregious*) along with their Cantonese equivalents. The results point to the ISR with English non-words as a unique predictor of the speed of learning novel English words. This finding supports the hypothesis that phonological STM facilitates L2 word learning. The study did not show, however, the relationship between the phonological STM scores and the vocabulary size in English. The author of the study interprets this result as an indication that the phonological STM stops predicting the vocabulary increase in a given language at a certain stage of acquisition. This explanation is probable and it has been proposed also by other researchers (Gathercole et al. 1992; Baddeley et al. 1998). However it is also possible that the lack of effect stems from problems with Cheung's vocabulary data. First of all, the data might have been influenced by extraneous variables (such as students' motivation to learn, the quality of English teaching). Second of all, it is possible that the vocabulary task (the Crichton Vocabulary Scale) was not the best test to measure the acquisition of L2 vocabulary in foreign language classroom. Perhaps the results would be different if Cheung used a vocabulary test based on the handbooks used in participants' English classes. This seems probable, since in studies were the vocabulary tests have been based on the materials used for teaching the students, there is usually a strong correlation between phonological STM and the L2 vocabulary size. This is the pattern of results that has been obtained, for example, in a study of 80 Greek children (aged 8-13) learning English at school (Gathercole and Masoura 2003). The participants of the study were asked to perform a non-word repetition task in English and in Greek, a non-verbal intelligence test and a test of L2 vocabulary. The vocabulary task was based on English handbooks used in Greece and contained 80 Greek items that had to be translated into English as well as 80 English items that to be translated into Greek. There was a strong correlation ($r = 0.48$, $p < 0.01$) between the performance on the tests and the English non-word repetition scores.

The problem with Masoura and Gathercole's study is of different kind, however. In contrast to the Cheung (1996) data, this study did not confirm the straightforward relationship between phonological STM and novel word learning in an experimental task. Following the initial testing procedures, the children examined by Masoura and Gathercole have been divided into four groups. There were two vocabulary groups: one with high vocabulary knowledge and one with low vocabulary knowledge, but both matched on non-word repetition and non-verbal intelligence (although the high vocabulary group had on average learned English longer). There were also two non-word repetition groups: one with high non-word repetition scores and one with low non-word repetition scores. Those two groups did not differ in terms of English vocabulary size. The four groups were presented with eight English words and eight pictures illustrating the words, taken from British Picture Vocabulary Scale. All children were asked to repeat the words and to learn the associations between the words and the pictures. Following the presentation, there was a testing phase, in which the children were presented with the pictures and asked to produce the English name of the object or animal depicted. The children had to learn the associations in ten trials consisting of interchanging presentation and testing phases. The results show a significant difference in learning speed between the high vocabulary group and the low vocabulary group. However, there was no significant difference between the low non-word repetition group and the high non-word repetition group. Therefore, the results obtained by Gathercole and Masoura (2003) are the mirror image of the results by Cheung (1996) - the two researchers found a relationship between non-word repetition and L2 vocabulary size, but no relationship between non-word repetition and L2 word learning. Again, different explanations for these findings can be offered, however, it is possible that the result of the learning study might be a simple methodological issue. According to the authors, the words taught to the children during the word learning task "were not likely to have been encountered by the children in their studies so far" (Gathercole and Masoura, 2003: 425). However, they also admit that the children might have been exposed to English outside school. Therefore it is possible that some of the words taught in the experiment were known to the participants. It is also possible that the mere difference in exposure to English among the participants could have influenced the results. As indicated previously, the children in the high vocabulary group have been, on average, learning English longer than the children in the low vocabulary group, so their linguistic experience and exposure to English might have been the third factor underlying both the performance on the vocabulary size tasks and the experimental word learning tasks.

To conclude, barring certain methodological issues in some of the experiments, data seem to indicate that that there is a relationship between phonological STM and L2 word learning. In particular, non-word repetition in L2 appears to be a significant predictor of vocabulary acquisition - both in experimental settings and in classroom environment. This relationship is also indirectly confirmed by data from language deficits and from extraordinarily gifted individuals. It seems that impairments affecting phonological STM affect L2 learning in general and L2 word learning specifically. There is also a relationship between superior phonological STM skills and polyglottism.

On the face of it, it seems that the issue of relationship between phonology, memory and foreign language learning is settled. The data suggest that at least at the initial stages of L2 acquisition, when individuals rely on phonological learning more than semantic learning, phonological STM facilitates vocabulary acquisition. However, the issue of the interrelations between phonology and lexicon is more complex than that. In the last 20 years strong criticism has been levelled at the phonological STM theory. In particular, it has been claimed that the non-word repetition task, which is used in most of the studies quoted, is not a pure measure of phonological STM. Several researchers put forward a hypothesis that non-word repetition might tap into another variable - phonological processing - and it is this variable that facilitates novel word learning. The next section will explore this hypothesis, as well as the criticism levelled at phonological STM and language learning research in more detail.

1.5. Beyond memory: the critique of the phonological STM hypothesis

While research on the relationship between phonological STM and word learning is well established in the field, it is still subject to problems and controversies. Some of the problems are related to the contradictory research results. While studies with experimental non-word learning tasks almost unequivocally show a relationship between non-word repetition scores and the efficiency of novel word learning, studies conducted in real life classroom do not show consistent results. Gathercole et al. (1992) claims that there is a relationship between vocabulary increase and non-word repetition, but only in children aged 4 to 5. Even this claim has been questioned by Melby-Lervåg (2012) and colleagues, who reanalysed Gathercole's data and found no effect of phonological STM on vocabulary size increase. Moreover, a longitudinal study performed by Gathercole and colleagues in 2005 indicates that there is no effect of persistent phonological STM deficit on language development in children. (Gathercole et al. 2005). The participants of the study, who were recruited from a large-scale longitudinal project, were tested on phonological STM and non-verbal intelligence at the age of 4 and 5. 39 children who were diagnosed with phonological STM deficits on the basis of those screenings and 15 children who also took part in the initial testing but did not have any disorders were given additional evaluation at the age of eight. This second evaluation included phonological STM tests (digit span and non-word repetition), a series of language tests (Wechsler Objective Language Dimensions) and vocabulary measures (British Picture Vocabulary Scale). Of the 39 children initially diagnosed with deficits, 24 displayed a persistent phonological STM disorder also at the age of eight, while 15 did not show deficit in phonological STM at the time of the second testing. Thus the researchers divided the eight-year-olds into three groups - a group with early phonological STM deficit, a group with persistent STM deficit and a control group. What is surprising, is that on the second testing the persistent phonological STM disorder group turned out to be indistinguishable from the control group on vocabulary tests and other language development measures. However, the children with early phonological STM deficits, displayed significant language deficits at the age of eight. In particular, their oral production scores and vocabulary scores were lower than that of controls. Interestingly, this group was also characterised by lower verbal IQ than the other two groups. These results suggest, that it is not the deficit in phonological STM per se that impairs vocabulary learning. Instead it is a general verbal processing deficit that initially impairs the phonological STM performance and leads to a general language delay, including a delay in vocabulary acquisition.

The above data question the involvement of phonological STM in vocabulary development. Yet, on the other hand, most research findings indicate a correlation between the vocabulary scores and phonological STM scores in children acquiring their first language and in older learners acquiring the second language, i.e. in learners at the early stages of language acquisition. The problem, however, is that there is little research that would track vocabulary progress over time in the natural settings of the learner and then use the increase in the vocabulary size as a dependent variable in the studies. Especially in the field of second language acquisition there is a severe lack of research that would investigate the relationship between phonological STM scores and the increase in L2 vocabulary over a particular period of time. For this reason, the real impact of phonological STM on word learning in a natural learning environment is still not known.

This is not the only problem with the hypothesis that phonological STM facilitates word learning. One of the most serious criticisms levelled at the theory is that the tests used to tap into phonological STM might not, in fact, be good measures of the concept. This claim has been made especially about the non-word repetition tests, which are often used as the main measure of phonological STM in studies on language development. The first problem with non-word repetition is that it seems to tap into different cognitive modules depending on the type of non-words used in the task. For instance, tasks with non-words resembling real words of a given language (i.e. word-like non-words) produce different correlation patterns with word learning than tasks with non-words not resembling real words (i.e. non-word-like non-words). This has been shown in the research by Gathercole herself (Gathercole 1995). In this study, the experimenter asked a group of 20 adults to assess the word-likeness of the non-words used in the Children's Test of Non-word Repetition (Gathercole et al. 1994), one of the most popular and widely used non-word repetition tasks. Then she analysed the data from 111 four- and five-year-olds who have performed the Children's Test of Non-word Repetition, alongside a receptive vocabulary test and a digit span task. She calculated non-word repetition scores for the word-like and non-word-like items separately and then for each set of non-words she performed separate statistical analyses on the relationship between the repetition scores, digit span and vocabulary size. It turned out that the word-likeness of the items was a very significant factor in the analysis. First of all, children aged four and five were better at repeating word-like non-words than non-word-like non-words. Second of all, the results of non-word-like non-word repetition were correlated to a greater extent with the digit span scores in the participants. Third of all, repetition of non-word-like non-words was a better predictor of vocabulary size in children than the repetition of word-like non-words. All this suggests that the Children's Test of Non-word Repetition could tap into several measures. The repetition of non-word-like items is, according to the author, a purer measure of phonological STM. The word-like items, on the other hand, tap into long-term lexical knowledge - including knowledge about the most common speech sound combinations (studies show that non-words containing frequent sound combinations are assessed as more word-like than non-word containing infrequent sound combinations – cf. Munson et al. 2005). This indicates that treating non-word repetition tests (such as Children's Test of Non-word Repetition) as a pure measure of phonological STM might be problematic.

And yet those tests are still often used as indicators of phonological STM. Many claims about the memory effects on language learning have been made on the basis of the observed relationship between non-word repetition scores and vocabulary scores. This led a number of researchers to criticise the theory of phonological STM as a language learning device. Many of those critics suggest that the factor underlying relationship between non-word repetition and word learning is not phonological STM, but a completely different faculty - the faculty of phonological processing. As pointed out by Snowling and colleagues (1991), non-word repetition requires a different set of skills than digit span which is considered a classical test of phonological STM. Repetition of non-word requires phonological analysis of completely novel words and creating temporary representations of these words. As such it engages not only phonological STM, but also the ability to phonologically process verbal information. It is quite possible that phonological processing also facilitates the acquisition of novel vocabulary, since learning new words also requires creating new phonological representations. Thus it is plausible that phonological processing is the factor mediating the relationship between non-word repetition and novel word learning.

A similar argument has been made by Bowey (1996, 2001) and Metsala (1999). However, those authors criticise in particular the conclusion made by Gathercole and Baddeley (1989) that the correlation between non-word repetition and vocabulary size in children reflects the causal influence of phonological STM on vocabulary acquisition. Bowey and Metsala argue for a reverse direction of influence and claim that it is vocabulary development that influences the performance on non-word repetition. According to Bowey (1996), as children develop greater vocabulary, they become better at phonological processing and have greater sensitivity to phonological structures. As a result they also perform better on non-word repetition tasks, which, as mentioned earlier, require phonological processing. This hypothesis is supported by her own studies (Bowey 1996, 2001) and data collected by Metsala (1999). In one of these studies (Bowey 1996), 205 five-year-olds have been tested on measures of receptive vocabulary, non-word repetition, digit span, non-verbal intelligence and tasks tapping into phonological processing. In one of the phonological tasks, the children were presented with three words and had to point out which of these words ended with a phoneme different from the two other items. In the second phonological task, the children had to identify which word in a set of three began with a sound different from the two other words. The study has shown correlation between the phonological measures and expressive vocabulary scores that remained significant even when the non-word repetition scores were controlled for. On the other hand, the effect of non-word repetition scores on the vocabulary measure disappeared when the phonological scores were controlled for. It has to be mentioned however, that digit span was also a contributor to the vocabulary scores, even when phonological task scores have been controlled for. In general therefore, the study does not disprove the involvement of phonological STM in word learning, however, it suggests that the factor underlying the relationship between non-word repetition and word learning is phonological processing.

Similar results have been obtained by Metsala (1999) in a study performed on 36 children aged 3-4. Here also receptive vocabulary, digit span, non-word repetition and phonological tasks have been administered to the participants. Phonological tasks turned out to be significant predictors of vocabulary size in the participants, even when the non-word repetition scores have been controlled for. The non-word repetition, on the other hand, was not a predictor of vocabulary size when phonological tasks have been controlled for. Metsala has also noted that the phonological task scores and the non-word repetition scores were significantly correlated. This study thus again indicates that the relationship between non-word repetition and word learning is mediated by phonological processing.

Finally, one of the few studies that registered vocabulary progress in children over time also suggests that phonological processing might underlie the facilitative effect of non-word repetition on word learning (Bowey 2001). In this study, 71 children were tested in two separate sessions about one year apart. The average age of children was 4;10 at the time of the first session. The participants were given tests of non-word repetition, phonological processing, receptive vocabulary, non-verbal IQ and grammar during the first session. During the second session, they were tested on non-word repetition and receptive vocabulary. Once the non-verbal IQ of the participants and the vocabulary size at session one were controlled for, non-word repetition predicted the vocabulary size at session two. However, so did phonological tasks. In fact, non-word repetition did not explain any unique variability in the vocabulary scores at second session, once the phonological tasks have been controlled for. There was also no unique contribution of the phonological task once that non-word repetition scores have been controlled for, which suggests that these two tasks might tap into the same ability - the phonological processing ability.

1.6. Conclusion

All in all, the data presented in this chapter suggest that there is still much to be discovered about the relationship between phonological STM, phonological processing and word learning. Even though many mainstream researchers (Ellis 2001; Baddeley 2003; Gathercole 2006) support the theory that phonological STM facilitates word learning, there is convincing evidence suggesting that phonological processing also plays a role in vocabulary acquisition. Moreover, it has been suggested that some phonological STM tasks might actually tap into phonological processing skills and that these skills might in fact underlie the apparent relationship between phonological STM and vocabulary acquisition in many studies (Snowling et al. 1991; Bowey 1996, 2001 Metsala 1999).

There is a need for more research on phonological processing and word learning, especially in the second language acquisition, where the data is really scarce. Therefore the study presented in this thesis aims to fill in the gap by exploring the relationships between phonological STM, phonological processing and word learning in both the native and a foreign language. However, before such a study can be carried out, there is also a need to clarify how should phonological processing be understood. This, as will turn out, is a truly gargantuan task.

The research that supports the involvement of phonological processing in word learning is usually written from the perspective of developmental psychology, clinical linguistics or psycholinguistics. Often this research is very applied in nature and does not go into much detail regarding what "phonological processing" means. "Phonological sensitivity" and "phonological awareness" - both used to describe the set of phonological skills used in the acquisition of vocabulary are defined in the literature as "the ability to perceive, discriminate and manipulate syllables, rhymes and phonemes" (Beattie and Manis 2014: 120). "Phonological processing" is described even more vaguely as "the use of phonological information (i.e. the sounds of a given language) in processing written and oral language" (Wagner and Torgesen 1987: 192). The problem is that the nature of phonological information and its use is a subject of huge controversies. It is not known how people perceive, discriminate and manipulate syllables, rhymes and phonemes and whether anything like phonemes or syllables even exists. The core of the problem stems partly from the fact that the topic of sound perception and phonological processing lies at the intersection of different research disciplines - linguistics, psycholinguistics, acoustics, computational linguistics and psychology. All of those disciplines have vastly different ideas about what phonological processing is and whether there is at all a phonological level in speech perception at all. Therefore, it seems prudent to evaluate these theories and arrive at one coherent theoretical framework concerning the nature of phonological processing before carrying out any further investigations. The attempt to create such a theoretical approach will be the subject of the next chapter. The chapter will attempt to disentangle different notion of phonological processing. From there, the reader will be directed to the chapters describing how phonological factors interact with vocabulary acquisition in different languages at different stages of acquisition.

Chapter 2: Sounds turned into system: theories of phonological processing

2.1. Introduction

The following chapter will be devoted to different theories of phonological processing, which is tentatively defined as a process of translating the speech input into linguistic code. As such, phonological processing is closely connected with speech perception and thus the theories of speech perception will feature significantly in the sections to follow. The chapter will begin with classical theories of phonology, including structuralist and generative theories. Then, psycholinguistic, neurobiological and computational models will be discussed in more detail. These include TRACE, Cohort models, Shortlist models and neurobiological models proposed by Poeppel and colleagues (2008) and Hawkins (Hawkins 2010b). These are only examples of speech perception models, but they should give the reader a very general overview of the current strands of thought in the area. As will be easy to observe, the models presented are often contradictory. Some researchers assume that phonological processing involves a segmentation mechanism that cuts speech input into smaller units that are further analysed and categorised. Others believe that there is no segmentation process - speakers just recognise speech units in input. Another area of controversy is the product of phonological processing. When listeners process speech, do they store actual speech samples in their minds or do they translate the speech input into a set of abstract units and only this abstract information is stored in memory? If there are indeed some abstract units, then what are they? Do listeners have phonological representations (i.e. representations of phonemes and syllables) or do they store whole word-forms in their minds? Finally, also the actual mechanism of processing speech is subject to many arguments. Some models see this process as an orderly mechanism, composed of three separate, consecutive processes. First the continuous stream of speech is translated into a set of phonemes (phonological processing), then the combinations of phonemes are identified as words (lexical selection) and finally combination of words are combined into sentences (integration) (Dahan and Magnuson 2006). Other models see phonological processing, lexical selection and integration as mechanisms that interact with each other and occur at the same time. Resolving those contradictions is of consequence for the investigation of the relationships between phonological processing and word learning. For one, if there is no level of phonological representation (and this has been claimed), then it is hard to argue that there is anything like phonological processing at all, much less to argue that it influences any kind of learning. Therefore, at the end of the chapter the author will attempt to propose solutions to some of these issues and present a unified theoretical framework of speech perception and phonological processing.

2.2. Traditional phonology: linguistic perspective

Phonological processing as described in classical linguistic theories can be understood as analysing the stream of speech and turning it into an abstract linguistic code. This idea can be traced back to the father of modern linguistics, Ferdinand de Saussure. More specifically, it can be found in the three ideas attributed to him and described in the *Course in linguistics* collated on the basis of de Saussure's lectures. The first idea is that there exists abstract linguistic processing and that language is, in fact, not just a stream of sounds produced by human beings, but an underlying system. The second idea is that listeners do not take in the speech sounds as they appear, but that they perform a certain kind of processing or categorisation on the input they receive. The third idea is that this kind of processing is based on segmentation of speech and on selective attunement to certain acoustic features (de Saussure [1916] 2013; Anderson 1985: 34-43).

De Saussure was one of the first to believe that language is a system of signs, that is pairings of forms and meanings, and that people interpret the speech signal they receive and translate it into the code understood by this system. According to him, listeners have sound images in their minds that they use as standards for producing and perceiving speech. These sound images correspond to the letters of the alphabet and are not fully specified in terms of reflecting each minute acoustic cue that appears in the speech signal. In other words, de Saussure assumed that people perceive speech as a sequence of separate abstracted segments. Listeners actively segment speech signal as they hear it and perform certain categorisations of speech sounds (Anderson 1985: 37-40). Even though de Saussure's works do not describe the exact mechanisms of these processes, they laid the foundation for the notion of phonological representation and phonological processing that other linguists elaborated on.

The idea of phonological representation was investigated in more detail by Mikołaj Kruszewski and Jan Niecisław Baudouin de Courtenay from the Kazan School. Like de Saussure, these two linguists believed that speech is segmented by the listeners and that speech segments are categorised into general, abstract units, which they called "phonemes". Baudouin, however, was more explicit in underlining that phoneme is a psychological entity - an ideal speech unit that can be variously realised (de Courtenay 1894). He clearly distinguished between the speech signal and phonological representations, which he defined as a set of idealised speech units. Thus, he implicitly suggested the existence of phonological processing - the processing that turns signal into abstract linguistic representations. The imperfect speech signal must be interpreted by the listeners to match the ideal phonemes in their minds. The Kazan school linguists did not elaborate on the mechanisms of this interpreting, yet important ideas about how this interpretation takes place can be can be found in later theories connected with Prague school of phonology and developed by Nikolai Trubetzkoy and Roman Jakobson.

Trubetzkoy and Jakobson introduced into the phonological theory the notion of distinctiveness and of distinctive features, used later on in other theories of speech perception (Poeppel et al. 2008). It must be underlined that both of these linguists were more interested in researching the language system, rather than exploring the psychological reality of speech perception. While Baudouin wrote explicitly about psychology and about phonemes as linguistic units in the speaker's mind, Prague school linguists defined phonemes as minimal units in a linguistic system (Trubetzkoy [1939] 1970). Nevertheless, their ideas have been influential in theories regarding the translation of the stream of speech into linguistic structure. The key idea in their theory was that phonemes are sets of distinctive features, i.e. features that make them different from other phonemes. These features have physical correlates. For instance the phoneme /t/ has a feature of voicelessness that makes it different from a voiced phoneme /d/. This is the feature that distinguishes the two phonemes from one another and this feature has a clear physical correlate - the lack of vocal folds vibrations.

While in Trubetzkoy's theory distinctive features were merely the characteristics of phonemes, for Jakobson, these were the basic units of speech (Jakobson [1939] 1980; Jakobson and Halle [1956] 1975). Jakobson devoted significant portion of his writing to describing and defining these features. In his understanding, distinctive features were binary in nature, general and identifiable at the level of acoustics, articulation and perception. The features had a binary character, according to Jakobson, because he assumed the whole process of phonological analysis is based on detecting oppositions (the idea of oppositions occurred earlier in the works of Trubetzkoy (1970), but Jakobson organised his theory around this concept). The speech segment can either have a particular feature or not. When small children start acquiring language, they do it by finding oppositions in speech. First they learn to distinguish between vowels and consonants, then between the oral and nasal sounds and so forth, adding one opposition at a time until the whole phonological system of distinctive features is developed (Jakobson 1980). The features devised by Jakobson were very general, since he wanted them to be applicable to the phonological systems of the world languages (Jakobson and Halle 1975). Most importantly though, Jakobson insisted that the distinctive features needed to be grounded in phonetic reality and have both articulatory and acoustic correlates. In this way, he made an important contribution to the theories of speech perception and phonological processing. His hypothesis suggested that translating speech output into linguistic structure can be done by detecting particular acoustic cues in the signal that will be characteristic for particular speech sounds. While contemporary studies indicate that the mechanism of speech perception is more complicated than that, the idea of distinctive features detectable in speech continues to be widely influential in the field.

Later phonological theories developed the ideas of the first linguists, adding new insight to the same basic notions of phonological representation. Bloomfield, for instance, assumed the existence of alphabetic phoneme, which he defined as "a minimal unit of distinctive sound feature" (Bloomfield 1962). He believed that phonemes were characterised by distinctive features that had clear physical correlates, although he did not claim these features to be binary in nature like Jakobson had done. One innovation added by Bloomfield was the notion that phonemes were characterised by their structural properties in addition to distinctive features. In other words, he proposed that phonotactic information was built into the phonemes. During phonological processing, listeners identify particular phonemes not only on the basis of acoustic cues but also on the basis of their position in the syllable and word (they take into consideration whether the phoneme is initial or final, whether it serves the role of sonority peak and so on). The idea that structural information and phonotactics can play a role in the recognition of speech sound and phonological processing is interesting, because it has been borne out by modern research (see McQueen 1998; Saffran et al. 1996a).

The American structuralists that grew out of Bloomfieldian linguistics in large part accepted the system of alphabetic phonemes. They did not add much insight into the mechanism of phonological processing. They believed speech perception to be like typewriting. Listeners simply hear speech and convert it into the string of phonemes on the fly. The string of phonemes is then subject to higher-level lexical and grammatical analyses (Anderson 1985: 280-286). Implicit in this view is that both speech and speech perception is sequential. Structuralists believed that all the information needed for the translation of speech input into string of phonemes must be contained in the phonetic representations. Listeners cannot use any higher level information (for instance grammatical information or frequency of particular sequences) to identify phonemes in speech. Structural linguists believed also in a very clear-cut correspondence between phonetic realisations and phonemes. They assumed biuniqueness, the notion that from concrete phonetic realisations one can always arrive at a particular phoneme and that each particular phoneme has a clearly defined range of possible phonetic realisations ("once a phoneme always a phoneme"). All in all, their vision of phonological processing was very simple. It assumed that speech itself was easily segmented and rendered itself easily to phonemic analysis. Unfortunately, as will be pointed out in the sections to follow, this idea does not reflect reality very well. Speech is a mass of overlapping cues that does not render itself very easily to the segmental analysis, as suggested by repeated failures to create good speech recognition systems in the past decades (Kluender and Kiefte 2006). Also the biuniqueness assumption is false, since different speech segments can have virtually identical realisations. Nevertheless, the idea of unproblematic phonological processing and clear-cut speech-to-phonemes correspondence proposed by the structuralists has been a very popular one and it still permeates thinking of phonology by many linguists today. Even though structuralism has been mostly replaced by other linguistic theories, courses of linguistics are often organised with the implicit assumptions that phonological processing is simple, that it does not mix with other levels of analysis (hence separate courses devoted to phonology and to other aspects of language) and that one can easily define and describe phonemes. This is despite the fact that these ideas have been largely questioned, even within the field of linguistics itself.

The biuniqueness account was already criticised by the representatives of generative phonology - the next major theory in phonology following structuralism. However, in generativism this criticism was based on abstract arguments. Like most linguistic theories, generativism distinguished between the level of phonological representations (ideal phonemes and distinctive features) and the level of phonetic realisation (the speech sounds that are actually uttered). It was believed, however, that the transition between phonological representation and the actual stream of speech is governed by a set of rules (Chomsky and Halle 1968). Generativists argued against biuniqueness on the grounds that through the operation of rules, the phoneme could be altered in such a way that its realisation was similar to another phoneme (Chomsky and Halle 1968; Halle 1959). One could determine the underlying segment mostly by looking at the morphological structure of the word and also by using the knowledge of the grammatical rules governing the changes in phonemes. Thus the insight provided by generativists into the notion of phonological processing was that morphophonemic representation and grammatical rules could play a role in speech recognition.

Unlike structuralists, generativists placed great emphasis in their theory on the importance of binary distinctive features. They saw phonological segments as matrices with binary features and assumed that the phonological rules operated by changing a certain feature in the phoneme. In the early generativist accounts, one can find the idea that distinctive features are organised within the phoneme in binary branching diagrams. According to this view, listeners assess and categorise phonemes by making a series of choices about the existence of particular distinctive features (Is a given phoneme an obstruent or sonorant? Is it voiced or voiceless? Etc.). It was also assumed that phonological segments contained only as many features as was necessary to distinguish this segment from other ones. This idea, however, was abandoned in later versions of generative phonology (Anderson 1985: 122-127).

In the early accounts of generative phonology (for instance, Chomsky and Halle 1968) it was also hypothesised that there was only one level (or tier) of phonological structure. For generativists, just as for structuralists, the utterance was a simple, flat string of phonemes and all the processes operated on the level of phonemes (or, strictly speaking, on distinctive features within phonemes). Later on, however, Goldsmith (Goldsmith 1976) proposed Autosegmental Theory, which assumed that phonological representations did not consist of one phonemic tier, but rather of several different (although linked) tiers. Apart from the phonemic (segmental) tier, there was also a tier where processes relating to suprasegmental features (for instance tone) were assumed to take place. There have also been proposals to introduce lower levels of structure related to parts of complex segments, such as diphthongs (Van der Hulst and Smith 1982). Also in another late generative framework, Metrical Theory (Liberman 1975; Hayes 1981), it was assumed the phonological structure of utterances is composed of many levels that are ordered hierarchically. There was the level of segments (phonemes), the level of syllables, the level of feet, of phonological words etc. Each of these levels had its specific phonological rules and operations. Metrical and autosegmental theories are important in that they underline that there are different levels of phonological analysis, an idea that will be important later on, for instance, for Poeppel et al. (2008).

Since the 70s, a number of phonological theories have grown out of or in dialogue with generative phonology. There is Natural Generative Phonology (Hooper 1976), which comes back to the idea that there should be a clear relationship between phonemes and phonetic realisations. There is also Optimality Theory (Prince and Smolensky [1993] 2008) and Natural Phonology (Donegan and Stampe 1979, 2009; Dziubalska-Kołaczyk 2002a, 2002b, 2006), which currently seem to be the most popular alternatives to the classic generative Government Phonology.

The Optimality Theory continues the tradition of formal linguistics set by researchers such as Chomsky, but it rejects the notion of rules governing the realisation of phonemes and replaces them with the notion of constraints. This theory assumes that the phonological system is composed of underlying representations (input) and a set of universal and violable constraints on articulation. These constraints state what is less acceptable in a given language and should be avoided in speech production. In this way they govern what realisations of the phonological representation are possible or not. According to Prince and Smolensky (2008), there are underlying speech sound representations in the speaker's minds. On the basis of those representations speakers generate a large number of possible outputs (realisations) for those representations. Then these outputs are judged against constraints, which help the speaker evaluate which of these representations are the most well-formed and at the same time faithful to the underlying representations. The constraints form hierarchies with regard to their importance, which means that in each language there are some constraints which are less violable than other. When the speaker evaluates the possible outputs, she usually chooses the most optimal one, so the one that violates the smallest number of constraints or one that does not violate the most important constraints. As can be seen from the above review, Optimality Theory in its classical form focuses mostly on speech production – the translation of linguistic representation into speech output. However, the theory might also be reversed and it might be argued that the knowledge of the constraint hierarchies in a given language is also a tool for decoding language input.

The Natural Phonology (Donegan and Stampe 1979, 2009; Dziubalska-Kołaczyk 2002a, 2002b, 2006) is similar to Optimality Theory in that is also rejects the assumptions of formal rules altering the realisation of phonemes in production, but it is a functional rather than formal approach. This means that it sees linguistic form as governed by function. In particular, Natural Phonology believes that the linguistic form is a compromise between human drive for communication which requires clarity of perception, and the ease of production. Natural Phonology follows linguists such as Baudouin de Courtenay in the belief that phoneme as a psychological entity – the intention of the speaker and a "fully specified, pronounceable percept" (Dziubalska-Kołaczyk 2002b: 11). What is crucial to the theory is the idea that a phoneme can be altered in production by the speakers to make it easier to articulate or clearer to perceive for the listeners. Those alterations are called phonological processes. The phonological processes that serve to ease the articulation of speech sounds are called lenitions. Consonant deletion or devoicing are examples of such processes. The phonological processes that serve to increase the clarity of perception are called fortitions and here examples include vowel lengthening. Each language chooses its own set of phonological processes that can be applied in speech, but not all of these processes are compulsory. Therefore, Natural Phonology is a theory of preferences rather than of rules. Language users might display a preference for reducing a given sound, but this does not mean that the reduction will apply across the board. Natural Phonology is thus different from the previously presented generative frameworks in that it replaces the notion of rules with the notion of preferences in processes. Consequently, the process of speech perception – translation of the altered speech input into linguistic code – is slightly different than in the account of generativists. Listeners know the possible phonological processes that can be applied in production by the speakers, so they can use this knowledge to decode the speech signal. However, since those processes do not apply universally, this decoding speech will go beyond the simple "reverse-engineering" the input into the phonological representations with the use of phonological rules.

While Natural Phonologists do not go into significant detail concerning how exactly the translation of speech into phonological representations takes place, they suggest three characteristics of the mechanism. The first is the already mentioned use of phonological processes to decode speech signal. The second is the claim of biuniqueness (Dressler 1984, 1996), which suggests the rather that each speech sound can be easily translated into an appropriate phoneme. This claim is not borne out by the data and simplifies a rather complex process of speech perception. However, the third characteristic of phonological processing in Natural Phonology and one which has been heavily underlined by Dressler (1984, 1996) is the idea that speech perception is based on contrast. This idea is not new in linguistics, it could be found even in Jakobsonian theories, however, it is heavily emphasised in Dressler's description of Natural Phonology. Dressler claims that speech perception is influenced by the figure-and-ground principle. This principle states that there are more perceptually salient elements in the speech (figures) and less salient elements in speech (ground). The figures become more salient due to the presence of the ground. Speech perception is largely driven by the contrast between the figures and ground. This idea is somewhat related to the ideas of perceptual contrast in the theory of Kluender and Kiefte (2006) and the theory of perceptual anchors by Sarah Hawkins (2010).

In general, when it comes to the issue of speech perception and phonological processing, most traditional approaches to linguistics offer a similar insight. Starting with de Saussure, the majority of linguists see speech perception as a process, in which the stream of speech is segmented and then the resulting segments are placed into general categories called phonemes. In many theories it is assumed that speakers categorise segments by identifying distinctive features. The more formal approaches suggest the existence of rules allowing for recognising a given segment as a particular phoneme.

Some frameworks suggest that several levels of segmentation are possible (syllable levels, phonological feet levels), others argue for speech being represented as a simple string of phonemes in the mind. A group of linguists claim that the knowledge of neighbouring structure or morphological boundaries or grammatical rules are involved in speech recognition. Natural Phonology suggests the importance of contrast in the process of turning speech input into linguistic code. Overall, the picture of speech perception and phonological processing in traditional linguistics is focused on turning the speech signal into sequences of phonemes that is the basis for all the further analyses.

Most linguistic theories do not elaborate to a great extent on the exact mechanisms of phonological processing, understood as the translation of the speech input into linguistic code. This is probably due to the fact that linguists usually focus very strongly on the structure of the language itself and try to explain how this structure relates to speech output. As a result, there are numerous accounts in linguistics describing how the underlying linguistic code might be turned into the actual speech with the use of phonological rules (Government Phonology), constraints (Optimality Theory) or preferences (Natural Phonology). Yet there are not that many descriptions of the other direction of processing – from the actual speech to the code. However, as will be shown, some ideas about this mechanism can be gleaned from the work of psycholinguists, computational scientists, neuroscientists and interdisciplinary linguists of all kinds. As will be seen, in their studies, speech perception turns out to be a very complex tasks, and the idea that speech is segmented into a string of phonemes becomes much more controversial.

2.3. Auditory phonetics perspective

Phonological processing, is closely connected with speech perception and auditory phonetics. Therefore, looking into empirical data in this area of research can offer valuable insights regarding the mechanisms of the process. In particular it can help evaluate the belief present in many linguistic theories that phonological processing is a simple process involving segmentation of speech and that is quite separated from other modes of perceptual processing. Auditory phonetics provides data that allow to question some of these assumption. First of all, research in this area shows that the idea of unproblematic segmentation into phonemes is misguided. Second of all, while it provides some evidence for the fact that speech processing as a unique type of processing it also shows many similarities between speech perception and other types of perception (for instance, vision).

2.3.1. Phonological processing is a complex issue

The problems with the idea of easy segmentation into phonemes has been noticed by many phoneticians across the years. Those issues have been aptly summarised by Port (2007b, 2008). In his papers, he notices that linguists see speech as easily divisible into phonemes - abstract, invariant units that are discrete from each other, and have easily identifiable boundaries. However, as he points out, empirical data from auditory phonetics do not confirm any of these assumptions.

The first problem is the invariance assumption, which states the speech units - phonemes - are the same in all contexts - /d/ is the same sound in the word "do" and in the word "day". And since they are the same, they can be readily identifiable from different context by some specific phonetic cues. However, speech sounds are altered by different contexts. The articulation of a given speech sound is influenced by the articulation of the sound neighbouring it. This phenomenon is called coarticulation (Liberman et al. 1967). An obstruent followed by a high vowel will be different from the same obstruent followed by a low vowel, because during the production of the obstruent the articulators will already prepare for the next sound. In fact, as pointed by Kluender and Kiefte, automatic speech recognition systems based on the assumption that there is one particular acoustic template of each phoneme simply fail (Kluender and Kiefte 2006). Identifying reliable acoustic cues - distinctive features that will be always associated with a particular phoneme - seems to be impossible.

A related problem is the issue of overlapping acoustic cues. The phoneme boundaries in speech are blurry. For instance, in a sequence of a plosive and a vowel, the information about the place of articulation for the plosive is retained on the formants of the vowel. In fact, the plosive is harder to identify if it is not followed by a vowel (Redford and Diehl 1999). If this is the case, it becomes very difficult to establish where the plosive ends. The problem with segmental boundaries is aptly illustrated by a large-scale experiment, in which the Dutch speakers were asked to identify phonemes in 1179 sequences of diphones (Smits et al. 2003). The diphone sequences were sliced into six time frames called gates and the speakers were asked to identify both phonemes at each gate (so 1/3 into the first phoneme, 2/3 into the first phonemes, 3/3 into the first phonemes, 1/3 into the second phoneme, 2/3 into the second phoneme and at the end of the sequence). The results show how fluid the boundaries between the phonemes are. The recognition rates for the first phoneme peak around the fourth gate (1/3 into the second phoneme), while the recognition rate of the second phoneme go up already at the third gate (and for some specific cases - as early as the second gate - 2/3 into the first phoneme). It is clear that in speech two sounds in the sequence permeate each other.

The presented problems are connected mostly with the assumption that phonemes are the basic units of speech. If one assumes that the basic processing is a larger sequence - such as a syllable or a word, the problems stemming from blurred boundaries or co-articulation are less severe. This might be the reason why children have problems finding phoneme boundaries before they receive literacy training (Ziegler and Goswami 2005), but are much better at segmenting speech into syllables. Nevertheless, even for larger speech units, phonological processing is incredibly complex. Speech sequences can be altered by the ambient sounds. They are produced differently by various people and will sound differently with various intonation patterns. They can be slurred, mispronounced and mangled. There is nothing easy or straightforward about processing them.

2.3.2. Phonological processing is a process different from ordinary sound processing, but it uses certain universal perceptual mechanisms

Despite all those problems, humans are incredibly apt at processing speech. This can be attributed to two seemingly contradictory features of speech processing. On the one hand, it seems that speech processing enjoys a special status in humans and that it is qualitatively different from other forms of auditory processing. On the other hand, it seems to depend on the process that humans are universally very skilled at – the perception of change.

The unique status of speech processing among other kinds of auditory processing is described in a review article by Remez (2008). First of all, the author provides evidence that processing of speech sounds is much faster than processing of other sounds and that it cannot be fully explained by generic theories of auditory perception such as auditory scene analysis (Bregman 1990). Second of all, he proposed that processing speech phonologically is an inborn quality in humans and requires very little learning. Finally, he postulates that this kind of processing is unique in that it requires very little input - speech can be significantly distorted or stripped of multiple acoustic cues and will still be understood.

The first claim comes from the studies that investigate the perception of speech vs. other sounds. In non-speech sounds people categorise tones as belonging to one stream (in folk understanding "one sound") depending on its acoustic categories. Tones that start at exactly the same moment or tones that are acoustically similar are grouped together into one stream. If one of the tones starts with a slight offset, it is grouped with another stream. This mechanism, however, does not explain speech perception well, since speech is highly unstable and varied, with frequencies changing rapidly over time. In speech, the similarity and timing principles are simply not enough to classify all the tones correctly to particular streams. For instance, in vowels different formants often change over time to a varying degree. If the principles of general auditory perception applied to them, people would perceive a vowel not as one stream but as several streams (several speech sounds).

Speech perception utilises different mechanisms than ordinary sound perception. For instance it relies heavily on fundamental frequency. Two formants that are harmonics of the same fundamental frequency will be perceived as belonging to one sound. Two sounds excited on two different fundamental frequencies will be assigned by the listener to two speakers. Speech perception is also different from ordinary auditory perception in that it is categorical, i.e. a speech sound is always placed by the listeners into one of the known speech segment categories. This is not the case in ordinary auditory perception, which, with certain exceptions, is governed by the Weber's law (i.e. the smaller the amplitude of two sounds, the greater the perceptual difference between them). In a perceptual experiment performed by Mattingly et al. (1971), it turned out that a given acoustic stimulus will be perceived differently depending on whether it is a part of speech stream or not. In this study the participants were given formant transitions beginning at different frequencies. Some of these formant transitions were followed by steady-state vowel formants, so they occurred in a context of an artificial syllable. Others were isolated, i.e. not followed by steady-state formants. The participants were given a categorization task with the artificial syllables – they had to judge whether the formant transition indicated the sound /b/, /g/ or /d/. Then the participants were given a discrimination task. In each trial, the participants were given three formant transitions – two which were the same and one that was different – and they had to point to the odd sound. Some of the trials contained only isolated formant transitions, some only artificial syllables. When participants of this study heard a second formant frequency transitions in an artificial syllables, they perceived the differences most clearly along phonetic boundaries established with the perceptual categorization task. For the isolated formant transitions there were no peaks of discrimination.

Studies like these show that speech perception is indeed somewhat different from ordinary sound perception, but interestingly, this kind of perception is natural to human beings and virtually unlearned. Research shows that infants at the age of 3 and 4 months integrate acoustic cues into syllables just like adults do (Eimas and Miller 1992). They can distinguish between two artificial syllables, but not between non-speechlike chirps. This means that the mechanisms typical for phonological processing are present in very young children and do not depend on complex, conscious learning.

The third claim regarding speech perception put forward by Remez is that it requires very little input. Even strong distortions do not hamper phonological processing significantly. People understand speech on the phone or speech produced by imperfect speech synthesisers. They also understand speech that is stripped from most of its features, as in the sine wave replicas studies (Remez et al. 1994). In these studies, researchers synthesise speech stimuli from 3-4 tones that correspond to the central frequency of nasal, oral or fricative resonance. These stimuli are so basic on the acoustic level that on the surface they do not even sound like speech. In fact listeners at first perceive them as abstract whistles. Yet, when participants of these studies are informed that the whistles constitute distorted speech, they are able to recover speech signal from the few presented tones with the accuracy ranging between 50 and 85%. Thus it seems that perceiving the change in frequency of three or four tones in speech is enough for phonological processing. Of course, it should be noted that the more acoustic cues, the better the recognition. Usually, in speech there are multiple acoustic cues that help listeners recognise the sounds (Diehl 2011). The mechanism of speech perception is very flexible and goes by whatever is available. A study by Kiefte and Kluender (2005) shows that listeners can identify vowels either by formant frequencies or by spectral tilt - depending on which information is better preserved in speech. Such evidence clearly indicates that speech perception is a very robust mechanism that can be performed on limited and distorted input and utilises whatever is most effective in this input.

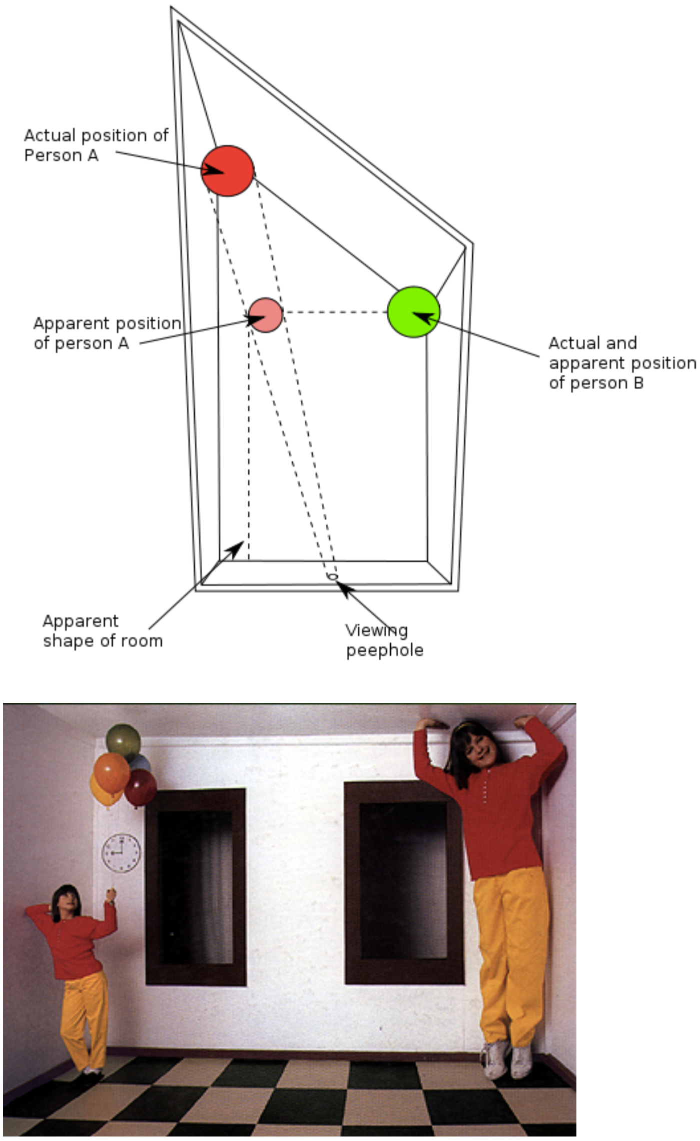

Figure 1: Ames room illusion illustrates the inversion problem in vision. The viewer is fooled as to the size of the people in the room due to a perspective trick (sources: http://en.wikipedia.org/wiki/Ames_room#mediaviewer/File:Ames_room.svg; http://psylux.psych.tu-dresden.de/)

Overall, the arguments quoted by Remez indicate that speech perception (and phonological processing that accompany it) is a specified, efficient process, which is natural for human beings. However, even though this process is so unique, it draws upon one of the most universal human abilities - change detection. In that it is similar to other forms of perception. This argument has been put forward in a paper by Kluender and Kiefte (2006), which explores the parallels between visual and speech perception. As indicated by the authors, there are several similarities between these two types of perception and the three described comprehensively in the paper is the inversion problem, the categorical perception and the sensitivity to change.

The inversion problem states that the input to human senses contains essentially too limited information to interpret the rich environment correctly. All input humans receive is to a certain extent ambiguous and open to various interpretations. For instance, in visual perception, the image of some object on a retina can correspond to smaller object close to the observer or a bigger one further from the observer. Without further cues regarding the proximity of the shape the observer has seemingly no way of knowing how big the object is. And indeed, sometimes the observer is fooled with regard to the size of the object as in the Ames room optical illusion (Figure 1). Similarly, in speech perception, a given realisation of speech sound can correspond to different phonemes. In general it is the case that many sources of sound can produce similar waveforms. And yet people rarely have problems interpreting speech and as has been already mentioned - they can decipher correctly even unusually distorted signal.

This is largely due to categorical perception that, again, applies both to speech and to visual stimuli. The categorical perception, as already indicated, is a mechanism of assigning a given stimulus to a meaningful category. Due to this mechanism, people usually do not perceive reality as fuzzy, even though given input can be interpreted in multiple ways. Instead they choose one (the most probable) interpretation of a given input and stand by it, or, if the signal is particularly ambiguous, they switch between interpretations. In visual perception, this can be illustrated with the duck-rabbit illusion (Figure 2), where the same line drawing can be interpreted by the brain as either the picture of a duck or a picture of a rabbit, but never is it seen as "something in-between" a duck and a rabbit. This is also the case with speech perception. Listeners always categorise the speech sound they hear (Miller and Eimas 1995). If, for instance, they are presented with a sound in between /ta/ and /da/, they will always make a decision whether this sound is /ta/ or /da/. Sometimes they will switch between the categories as in the duck-rabbit picture. Interestingly, people can notice also within-category acoustic details (Miller and Eimas 1995; Hawkins 2010a) - after all, they can notice slight articulatory idiosyncrasies such as lisping. Nevertheless the categorisation tendency is very strong and guides the perception of speech. This categorisation mechanism and perceiving boundaries between sounds as clearly delineated, when in fact they are fuzzy, makes speech perception possible. It allows to interpret all the sounds intended by the listeners as speech, even when the sounds are distorted. Even though sometimes listeners choose wrong speech category (and thus mishear a word), most of the time choosing the most probable interpretation leads to successful deciphering of the speech signal.

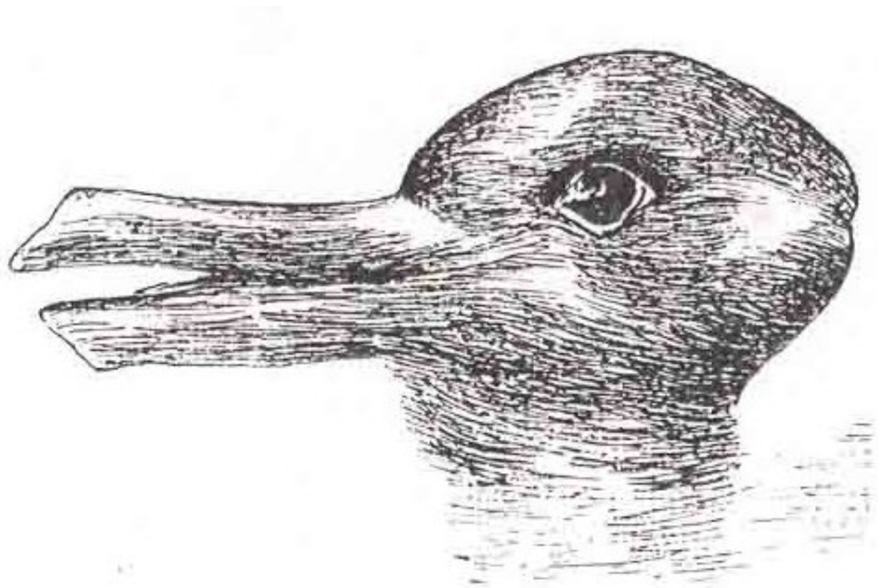

Figure 2: The Duck-Rabbit illusion illustrates categorical perception in vision - the viewer interprets the picture as either a duck or a rabbit, but never as something in-between (source: http://en.wikipedia.org/wiki/Optical_illusion)

This deciphering, according to the Kluender and Kiefte, takes place through detection of change. Again, they draw analogy with visual processing to indicate that receptivity to change is one of the most prominent features of human perceptual system. The most dramatic demonstration of that notion is human perception of colour and lightness. None of these features are assessed by human perceiver in absolute terms - instead they are perceived in relation to the environment. If people are shown different shades of grey in a darkened room, they will essentially judge the lightest grey colour as white no matter what its absolute luminance really is (Palmer 1999). Also colours are categorised in comparison with their environment, as clearly demonstrated by the Munker illusion presented in Figure 3. This is a very practical solution, because while the ambient lighting around objects constantly changes, the relations between objects stay the same. Thanks to the relative perception of colour and lightness, people will always see text on a page in a book as black on white, regardless whether they sit outside in the sun or read by a lamp. If they simply perceived the absolute luminance of objects, they would have the feeling that it is the objects, and not the ambient lighting, that are constantly changing. A similar mechanism applies to speech. It has already been mentioned that speech sounds are changing depending on the context. Also the ambient acoustic setting can change the spectral qualities of speech sounds. The acoustic signal that reaches the listener in a quiet room will be different from the signal received in a noisy bar. Yet, listeners have the feeling of sound constancy - they perceive the context and the ambience as changing, but the speech sounds as relatively stable. As a result, they recognise vastly different realisation of words and can judge /d/ in /da/ and /di/ as essentially the same, despite the actual acoustic difference. These processes are possible, as Kluender and Kiefte posit, because features of speech sounds are not assessed in absolute terms. What is perceived is the change in the sound quality - the contrast. Detecting sound modulations within speech sequence is the basis of speech perception and phonological processing.

Kluender and Kiefte (see also Kluender and Lotto 1999) claim that the categorical perception and the reliance on contrast are essentially linked in speech processing. Categorical perception works by strengthening the contrasts between particular speech sounds. If there are similar correlations between the acoustic cues in two sounds, i.e. similar acoustic features are correlated to a similar degree in both sounds, the contrast between these two sounds is minimised. However, if the correlation patterns in the two sounds are different, the perceptual contrast between these sounds is strengthened. This theory explains why humans can perceive two sounds as belonging to the same category even though these sounds have different absolute values of particular acoustic parameters. The framework proposed by Kluender and Kiefte sees phonemes as patterns of acoustic feature correlations that are different from other patterns of acoustic feature correlations. This definition defies the classical description of phonemes as collection of characteristic, distinctive features that can be used to identify a particular speech segment in the absolute sense.

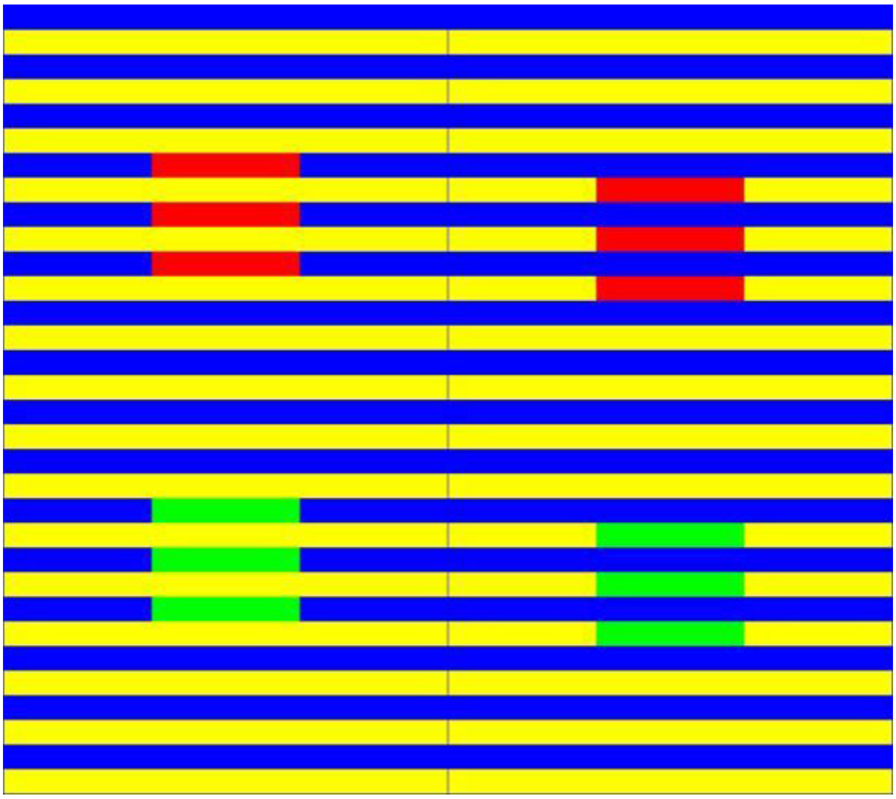

Figure 3: Contrast as the basis of perception in vision illustrated by the Munker illusion - red stripes appear to be orange when interchanged with yellow stripes and pink when interchanged with blue stripes (source: http://deesaturate.blogspot.com/2012/05/munker-illusion.html)

In principle, the most important concept proposed in the chapter by Kluender and Kiefte is a theory of phonological processing based on change detection, in place of the model focusing on distinctive features identification and division of speech into phonemes. Their theory suggests that speech perception might be based on the analysis of pairs of speech sounds (whether they are alphabetic phonemes or not) and detection of changes between these two sounds. It might be the case that the process of phonological processing is not a simple act of cutting speech signal into phonemes, but rather the detection of sound changes in diphones and matching appropriate patterns available in the mind of the speaker to match these.

The papers discussed in this section focused on research in acoustic and auditory phonetics. The first insight offered by studies in this area is that phonological processing is a highly complex process and that the segmentation into phonemes is greatly complicated by issues such as variance, coarticulation and fuzzy boundaries. The second insight of the phonetic research is that speech processing is a highly specified mechanism that differs from other types of auditory processing, operates very fast, goes by very little input and is highly flexible. The ability to process speech occurs at very early stages of human life. The third insight is that despite this highly specialised nature of speech processing, it utilises the universal human capability to detect change that is also used in other perceptual systems, such as the visual system. The identification of speech sounds probably occurs through detection of contrast between the neighbouring sounds. While these insights are very valuable and provide useful pointers for assessing the theories of phonological processing they do not reveal much about the subject of phonological representations in mind and of mental processing that turns sounds into language. This is because theories of mental processing that operate on linguistic stimuli belong the domain of psycholinguistics rather than auditory phonetics. Therefore, in the next section, the reader will be guided into psycholinguistic theories that will focus on the mechanism that turns sound signal into linguistic code.

2.4. Psycholinguistic perspective

When one compares psycholinguistic theories of speech processing with linguistic or phonetic frameworks, a very noticeable difference is that psycholinguists focus strongly on recognition of words. It is words, and not phonemes, that are considered the basic units of speech in psycholinguistics. This shift in perspective can be very illuminating, because it turns the whole problem of speech processing around. When researchers stop being fixated on sublexical segments, a whole range of ideas come to light and new questions about speech perception are being asked. These questions can be grouped into three major issues. One is the problem of segmentation - how is speech segmented and categorised. The second one is the problem of representation - how are speech units represented in the mind. The third problem is the integration problem - how are other aspects of linguistic processing (for example lexical) connected with phonological processing. These three questions are largely interrelated and finding answers to them is essential in defining phonological processing and uncovering the mechanisms that govern it.

2.4.1. Segmentation of speech - how does it happen

The issue of segmentation is basically the question of whether listeners first divide the stream of speech into smaller pieces and then analyse those smaller pieces or whether they simply recognise phonological and lexical units in speech and segmenting speech is a side effect of this process. Those two ideas correspond to the two types of segmentation theories in psycholinguistic literature - pre-lexical theories and post-lexical theories (Miller and Eimas 1995). Pre-lexical segmentation theories state the listener first divides the stream of speech into phonological units and then these units are classified into particular categories. This is the view presented by Anne Cutler and her associates (Norris and Cutler 1985; Cutler 1990) and it is utilised in speech perception models such as Shortlist B (Norris and McQueen 2008). This theory states that listeners have to detect the boundaries of linguistic units in the stream of speech before any further speech processing and speech recognition takes place. This means that these unit boundaries in speech are marked by certain concrete features and are easily recognisable. Cutler (1990) argues that each language has its own specific unit boundary marker and boundary detection strategy. In English, for instance, segmentation is based on metrical criteria. By detecting strong syllables (i.e. syllables containing a full vowel in contrast to schwa), listeners can identify word onsets and thus detect junctions between words and units.

Post-lexical theories present a reverse idea. There is no segmentation prior to identification of sounds and words and thus no need for a unique junction marker. In these models, division of speech into units happens as a result of recognition. This is the assumption of interactive activation models such as contemporary versions of cohort models (Marslen-Wilson 1987; Gaskell and Marslen-Wilson 1997), Neighbourhood Activation Model (NAM – Luce and Pisoni 1998) and TRACE (Elman and McClelland 1984). Cohort models and NAM assume that listeners segment speech into words by recognising those words in speech. In cohort models (Marslen-Wilson 1987; Gaskell and Marslen-Wilson 1997), it is believed that upon hearing the beginning of the word, listener activates a set (cohort) of words that are consistent with this beginning. As the speech flow progresses, the activation is inhibited for the words that do not match the input and in the end the word that fits the input best is chosen. NAM (Luce and Pisoni 1998) is very similar, but it assumes that words are activated on the basis of their global similarity to the speech input. Therefore, a word such as "cat" would activate a number of phonologically similar words ("bat", "kit", "cab", "at", "scat" etc), while in cohort models the same word "cat" would activate only words with similar onset ("camera", "cattle", "castle"). Overall, however, both model are similar in their assumption that language users detect word boundaries in the process of speech recognition

TRACE (Elman and McClelland 1984) is a model that is more complex than cohort and NAM models. It assumes that recognition (and thus segmentation) happens at three levels simultaneously: at the level of phonological features, at the level of phonemes and at the level of whole words. As listeners receive speech input, they create hypotheses about this input on these three levels. If the hypotheses are mutually consistent, they are strengthened. For instance, if the listener has a hypothesis that there is the feature "vocalic" and "open" in the middle of the word, that the word contains the phoneme /ɑː/ and that the word is "card", these hypotheses are mutually consistent, so they will be activated more strongly in the mind of the listener. Hypotheses that are mutually exclusive (for instance that there is a /iː/ vowel inside the word and that the word is "card") inhibit each other. TRACE continues the Jakobsonian tradition of distinctive features and phonemes. However it completely excludes the possibility of pre-lexical segmentation. One of the assumptions made during the creation of the model was that overlapping of cues in speech effectively prevents establishing clear-cut boundaries between sounds. Therefore the model does not segment speech into phonemes. It merely recognises the peaks of phonemic features as they appear in certain moment in time and makes hypotheses about phonemes and words that fit the patterns observed.

As pointed out by Dahan and Magnuson (2006), the post-lexical theories are gaining more and more support in the field of psycholinguistics. This is not surprising, since those theories seem to fit better into the phonetic research described in the previous section. Phoneticians have problems identifying clear-cut acoustic cues that would signify unit boundaries. Their research shows that that speech segmentation is a notoriously difficult task. Therefore, it is not very probable that the success of speech recognition hinges on the isolated segmentation process that occurs before any other phonological and lexical processing comes into play. Nevertheless, there are certain phonological cues that can mark the junctures between units in the stream of speech and these cues are used by listeners as additional source of information about the word boundaries. Research shows that phonotactics (McQueen 1998), phonetic detail (Quené 1992) and prosody (Salverda et al. 2003) all help listeners segment speech. Another very important type of information that is used to facilitate segmentation is transitional probabilities (Saffran et al. 1996a, 1996b; Saffran 2001). Transitional probabilities refer to the information about which sequences of sounds frequency occur within words, and which can occur mainly across word boundaries. To be more specific, transitional probability is the frequency of a pair of syllables (for instance, /beɪ/+/bi/), relative to the frequency of the first syllable (/beɪ/). Listeners store this kind of information and use it as an additional clue to pinpoint where a word begins and ends by recognising the sequences that are typically occurring at word boundaries. On the whole, it seems that segmentation seems to utilise a roaster of different processes. It takes place via recognition process, but detecting junctures in speech with the help of phonetic and prosodic cues, phonotactics and transitional probabilities is also used. As will be suggested in the chapters 3 and 4, the particular mechanisms of segmentation depend also on the stage of acquisition of a particular language. As will be shown, babies acquiring their first language and foreign language learners tend to rely more on pre-lexical segmentation strategies, while mature language learners use most probably primarily post-lexical segmentation.

2.4.2. Phonological representations - what are they

The problem of segmentation is closely related to the issue of phonological representations in the listener's mind. Do such representations exist? What are their shapes and how do they take place in the phonological processing? In the literature on speech representations, two issues receive particular attention. The first issue is the abstractness of representation, the second one is the granularity of representation.

This question of abstractness of representation can be answered in psycholinguistics from two different perspectives: abstract and episodic (Connine and Pinnow 2006). In abstract theories of speech representation, it is believed that listeners create general abstract categories for the sounds and for words they encounter. This is the model suggested by many traditional linguists. In linguistics, phonemes are usually perceived as abstract categories. This is especially the case for Prague school phonology and generative frameworks, where phonemes were defined as sets of abstracted distinctive features. However, abstract models are also utilised in influential psycholinguistic models of speech perception such as cohort (Marslen-Wilson and Welsh 1978; Marslen-Wilson 1987) or TRACE (Elman and McClelland 1984). Episodic theories (also known as exemplar theories) of speech representations are based on the assumption that listeners store actual tokens of speech sequences that have been heard (Goldinger 1998; Pierrehumbert 2001; Port 2007a). All exemplars of a given word or phrase form a set that can be conceptualised as a cloud in which more similar phrases are more closely connected with one another than less similar ones. The set of exemplars form a category. It is notable that in most explications of the theory, exemplars are assumed to have the form of speech sequences, rather than of particular segments. For instance, Port (2007a) explicitly claims there is no system of systematic symbolic orthographic-like representations. Human minds are filled with various speech chunks. Phonological processing operates by matching the new speech input to the exemplars of speech sequences in the mind.

There is certain evidence supporting the exemplar theory. First of all, as pointed by Port (2007a), listeners seem to store a significant amount of seemingly irrelevant information about given speech stream. They store information about particular idiosyncrasies of the speaker's accent, the tone of voice, the acoustic quality of voice. If representations of speech were abstract, he argues, this would not be the case, because listeners would simply extract from the speech stream the few cues that are essential for grasping linguistic meaning, translate these into linguistic codes and disregard all the additional data. This, however, is clearly not the case, and the richness of representations constitutes evidence for exemplar models. Another argument, put forward by Bybee (2002) is that certain phonological processes that apply only to particular words and phrases. For instance, vowel reduction occurs in frequent words like "memory", but not in rare ones like "mammary". This suggests that listeners store many instances of the word "memory" with reduced vowels, but not many exemplars of the word "mammary" with such reduction. If speech were interpreted by the brain as a string of abstract units, processes such as vowel reduction would always apply in particular contexts and would not be limited to a few chosen phrases. The final piece of evidence comes from computational models. As indicated by Pierrhumbert (2001), computer simulations based on exemplar models are very good at emulating both the perception and production of speech.

Nevertheless, as pointed by Bybee (2002) most exemplar-based theories propose some degree of abstraction within the theory. A number of researchers propose a dual model of phonological representation. This model assumes the existence of the exemplar level, where rich representations of actual utterances are stored, and of the abstract level, containing speech categories extracted (and abstracted) from the exemplars (Pierrehumbert 2003; Beckman et al. 2007; Munson et al. 2011). The abstractions are probably created by extracting patterns observed in the exemplars and grouping them together. Some scientists believe that speakers extract the abstract representations by analysing statistical distributions among sound patterns and establishing prototypical speech sounds. This is the basis of Patricia Kuhl's perceptual magnet theory (Kuhl 1991, Kuhl et al. 2008). According to this theory, within a given set of exemplars, the most frequent tokens are taken to be prototypical members of the category. These prototypical sounds help establish boundaries for the category - everything that is close to the prototype will be automatically assumed to be the member of the category. A theory based on the notion of prototypes and exemplars explains very well certain empirical data, such as the studies on perceptual boundary shifts. Boundary shifts occur when listeners are exposed to different pronunciations of particular speech sounds. As a result of this exposure, listeners change their assumptions about which sound can be placed in a given category and which cannot (Norris et al. 2003). For instance, when listeners are exposed to a lisped production of the /s/ consonant, the boundary of the /s/ category will change for them to include also the distorted, lisped productions. This phenomenon can be easily explained by prototype theory. The prototype is usually established on the basis of token frequency. When frequency of such distorted token increases, the prototype within a given category changes to accommodate this and the change of prototype shifts the boundaries of the category. Overall, it seems that the combination of exemplar and prototype theories constitute a good model of speech representations. Therefore, the dual model of speech representation, composed of both exemplars of speech and abstract categories of speech units created on the basis of those exemplars will be followed in the rest of this thesis.

The question of abstraction is one of the two problems related to the issue of phonological representation. The other one is the problem of granularity, i.e. the shape or size of phonological representation. This is essentially a question of whether listeners segment speech into words, syllables, phonemes or still other units. Many psycholinguistic theories indicate that speech is segmented on the level of words - the whole word is recognised and there is no need to divide it into further units (Dahan and Magnuson 2006). However, there are also reasons to believe in segmentation at sub-lexical level (below the level of words). This is evidenced by the fact that people can segment speech into syllables rather well and early on in their linguistic development. While pre-literate children and illiterate adults have problems with dividing speech into phonemes (Read et al. 1986; Morais et al. 1979) or count the number of phonemes in the word, they find it easy to count syllables in a given word. One of the most convincing support for this idea is a study, in which 947 children aged 24 to 72 months were asked to perform an array of phonological awareness tasks (Anthony et al. 2003). One type of tasks was blending, in which the participants were to put together speech sounds or sequences they heard and produce a word out of them. In one version of the task the children had to put together words (such as cow and boy), in another - syllables (/sɪs/-/tɚ/) in yet another - phonemes (/f/-/ɪ/-/s/-/t/). It transpired that younger children were unable to perform the task at the phonemic level, but were quite good at performing at syllabic levels. Similar results were obtained in another task called elision, in which the children had to say a given word with a particular sound missing. Also in this task there were multiple levels of complexity. In one version of the task the children were asked to delete a syllable (the child was asked to say, for instance, the word pencil without the syllable /sɪl/), in other to delete an onset (farm without /f/) and in yet another to delete a phoneme from an onset cluster (scar without /s/). The results of the study show a clear developmental sequence - recognising and manipulating syllables comes in children before the recognition of onsets, which in turns precedes phonemes identification and manipulation. Recognition of phonemes is largely connected with the literacy training children receive, but division into syllables is rather natural. An article by Ziegler and Goswami (2005), which contains an analysis of several studies investigating the performance of children on syllable and phoneme counting tasks, leads to similar conclusions. The authors show that between 48 and 94 per cent of kindergarten children can without problems detect and count syllables in a word (for phonemes this is between 0 and 67 per cent of children). This indicates that listeners can extract certain sub-lexical units from speech at a very early age.

This ability would probably not exist if it did not have any practical utility. There is also evidence that sub-lexical units are stored by listeners and are used in the process of vocabulary learning. As indicated by Storkel (2001, 2004), children learn new word forms faster if they are composed of common speech sound sequences. This finding indicates that children store common speech sequences and can use them to create representations of new words they learn. Overall, on the basis of existing research it can be concluded that segmentation occurs at both the lexical and the sub-lexical level. It is not clear if the segmentation of speech happens at the level of phonemes, but there is evidence that listeners divide stream of speech into sound sequences or syllables. As will be shown in the section on neurobiological theories of speech perception, there are reasons to believe that there is not one, but at least two levels of sub-lexical representation. Therefore, it can be concluded that listeners store phonological representations - that is the representations of speech units smaller than words - in their minds.

2.4.3. Integration between phonology and lexicon in speech perception

The issue of granularity of segmentation is related to the third problem of speech perception in psycholinguistic theories - the problem of integration of the phonological and lexical levels in linguistic processing. In many linguistic theories, it was assumed that the level of phonology and the level of lexicon were separated. Especially the structuralists indicated that there is no mixing of levels. However, as could be seen previously, interaction models such as TRACE assume that the recognition of phonological units and words happens at the same time. There is no real separation between these two levels. Just as information about phonological units facilitates the activation of particular words, the information from lexicon feeds back into the model and constraints the interpretation of phonological units. This is, however, not the case for all psycholinguistic theories of speech perception. Apart form the interactive models, there are also autonomous (a.k.a. feed forward or modular) models, in which there is no feedback loop from lexical to phonological level.

The most popular autonomous model is Shortlist B, based on Bayesian statistics (Norris and McQueen 2008). In this model, it is assumed that there are two levels of representation: pre-lexical (which is the level of phonemes) and lexical. The information flows from the pre-lexical level to the lexical level, but not the other way round. In shortlist B speech recognition is based on the recognition of phonemes. On the basis of the phoneme confusability data from a gating experiment (Smits et al. 2003), the researchers established a probability for each phoneme being recognised by the listener and fed these data into the model. Speech recognition in the model starts by identifying the first phoneme (on the basis of the phoneme probability) and then goes phoneme by phoneme, producing a list of most probable speech outputs. At the same time, at the level of lexicon a list of most probable candidate words is calculated and presented after the identification of each phoneme. Once all the phonemes are presented, the most probable word candidate is selected. No information from the lexicon flows to the pre-lexical system, i.e. the phoneme probabilities are not determined by information from the lexicon. In this way, the model supports the view that the recognition of phonemes comes before lexical selection and that the levels of analysis do not mix.

Shortlist B is a very good computational model. It is very simple and efficient, it also emulates real-word data related to speech processing very well, but there are two problems with it. The first problem is that it glosses over the issue of phoneme recognition. The authors use real word data about phoneme recognition probabilities and in this way they do not need to explain how exactly phonemes are recognised. The second problem is of an even greater concern. This problem is connected with the model's assumption that information from lexicon does not affect speech sound recognition. A phenomenon called Ganong effect indicates that this assumption might be wrong. Ganong effect (Ganong 1980) concerns the categorisation of ambiguous sounds placed in words and non-words. Listeners tend to perceive ambiguous sounds in speech sequences so that the resulting sequence is a word, rather than non-word. For instance, an ambiguous sound which is on the continuum between /t/ and /d/ is interpreted as /d/ when placed at the beginning of speech sequence /æʃ/, because there is a word *dash* /dæʃ/ in English, but not a word *tash* /tæʃ/. The same ambiguous sound placed at the beginning of the sequence /æsk/ is interpreted as /t/, because there is a word *task* /tæsk/, but not a word *dask* /dæsk/. As indicated by Samuel and Pitt (2003), this effect cannot be attributed solely to transitional probabilities. Ganong effect provides evidence for the feedback loop between the lexicon and the phonology, which is problematic for autonomous theories such as Shortlist B. Thus, overall, it seems that the interactive activation models might reflect the reality of speech processing better. It is most probable that lexical and phonological information are processed at the same time and that there is a feedback loop between these two levels of representation.

Overall, the presented review of psycholinguistic literature gives several insights about phonological processing. First of all, research in this field seems to suggest that listeners do not divide speech into segments prior to categorising phonemes and recognising words. The popular interactive activation models such as TRACE (Elman and McClelland 1984) work under the assumption that the segmentation takes place as a direct result of recognition. As listeners recognise a word of a sequence, they establish its boundary. However, phonetic and phonotactic cues are used by the speakers as additional confirmation of the established boundaries. Also transitional probabilities are used to help ascertain when one word begins and another ends.

Second of all, the studies discussed provided some clues regarding the nature of phonological representations. It seems that the idea that phonological representations are stored in the form of abstract minimally distinct category seems to be misguided. It is more probable, that there are two levels of representations in the minds of the speakers. One consists of exemplars of actual speech sequences, the other contains abstractions from these exemplars (Beckman et al. 2007). This would explain why people are capable of categorical perception, while at the same time they are attuned to fine phonetic details and extralinguistic information such as the tone of voice. The judgement is still out on the shape of linguistic units in speech recognition. All of the theories presented assume the existence of word as a linguistic unit. When it comes to the units below the level of words the issue becomes more complicated. Some theories do not mention any sub-lexical (phonological) units at all. TRACE and Shortlist models assume the existence of phonemes and TRACE even makes assumptions about distinctive features. The notion of phoneme has been heavily criticised in the literature, however, it seems that there is some level of phonological representation. More evidence to support this claim will be presented in the next section (2.5).

Third of all, it seems that the processing of phonological units and words happens at the same time and is interconnected. Just as speech sound recognition aids lexical selection, lexical information aids categorisation of speech segments. So far, the studies from the domain of linguistics, phonetics and psycholinguists have elucidated the nature of phonological processing to a large extent. In the next section the topic will be approached from yet another perspective - a neurobiological one. The presented data from neurobiological studies will provide unique insight especially with regards to the levels of phonological representations. Both accounts quoted in the section to follow suggest the existence of multiple sub-lexical levels of representation.

2.5. Neurobiological perspective

The following section will focus on two neurobiological models of speech processing - Poeppel's model of cortical speech processing (MTRM) and Hawkins's model based on the Adaptive Resonance Theory. Both models provide unique insight into the notion of phonological processing and in particular - of phonological representations. In Poeppel's approach one can find strong support for the notion of phonological representations not only exist in the mind, but also that there are several levels of such representations. Hawkins's theory also supports the notion of different levels of representations and moreover offers a unique insight into the notion of distinctive features.

One of the most interesting phonological processing models elucidating the nature of phonological representations is the multi-time resolution model of cortical speech processing (MTRM) proposed by Poeppel and colleagues in the paper entitled "Speech perception at the interface of neurobiology and linguistics" (Poeppel et al. 2008). As indicated by the title, in this model the authors try to combine traditional linguistic notions (such as the notion of distinctive features) with neurobiological discoveries. There are three assumptions of the model. The first one is that speech perception occurs at several temporal resolutions at the same time. The second is that the speech processing algorithm is analysis by synthesis. The third is that speech recognition is based on distinctive features.

The first assumption is based on the discovery that people process auditory signals of different temporal resolutions in different parts of the brain (Boemio et al. 2005). Fast modulations of the sound within the time scale of 20 to 80 ms (quick changes in pitch, amplitude and other acoustic features) are interpreted primarily by parts of the brain in the left hemisphere, while slower changes occurring at the time scale of 150-300 ms are processed in right hemisphere. From there, Poeppel and colleagues draw a conclusion that speech is processed at the same time on two levels - the levels of smaller units (which they identify as phonemes) and the level of larger units (which they identify as syllables). Thus their model is characterised by what they call multitime resolution processing. The second assumption revolves around the idea that speech perception is an active process. Instead of passively tuning into the input signal, the listener makes hypotheses about what is carried in the input and tests these hypotheses. Upon first hearing the stream of speech, the listeners make rough acoustic measurements of the speech signal and on the bases of these they create hypotheses about the distinctive features in speech and possible syllables conveyed by the input. Then these hypotheses are matched against the lexicon and the most probable word is chosen. The final assumption of the theory is that speech perception operates by identifying distinctive features in the signal. The authors of the paper believe in the existence of invariant phonological features that are connected to acoustic cues on the one hand and to articulatory gestures on the other hand. These features are recognised on the basis of the small window analysis (20-80 ms). Since the model operates on the basis of recognition distinctive features it is abstractionist rather than exemplar. In this way, it returns to the linguistic tradition reviewed at the beginning of this chapter.

The adoption of distinctive features at the basis of the model is controversial, as even the authors themselves admit. Since up to this date no clear correlates of distinctive features have been established, this tenet of the theory might be worth rethinking. However, the greatest value of Poeppel's model is its notion about several levels of sub-lexical representations and in defining the time scale of these representations. The framework is truly revolutionary in that it includes cutting-edge neurobiological research concerning the two time windows of speech processing into a well-rounded model. Poeppel's article provides strong evidence for the existence of sub-lexical processing and moreover shows that this sub-lexical processing probably occurs at more than one level.

The model of speech perception proposed by Sarah Hawkins (2010b) comes, on the other hand, with an ingenious approach to distinctive features. Hawkins's approach is based on the neurobiological Adaptive Resonance Theory (Grossberg 2003), according to which speech perception involves matching patterns in speech input to patterns in long-term memory. In neurobiological terms, this matching is nothing more that "a resonant wave of activation across the [neural] networks" associated with the auditory perception of speech (auditory cortex), memory for phonological segments and attention (medial geniculate nucleus) (Grossberg 2003: 423).

In light of this theory, Hawkins defies the claim that there are any stable, abstract linguistic categories that are real objects in minds. Rather, she believes categories are created ad hoc in the process of speech perception as a result of resonance in neural networks. The theory rests on the concept of auditory objects, which replaces the concept of orderly linguistic units in minds. An auditory object is an entity that comes into existence when the speech input is matched with a memory trace. The resonance between these two results in the sensation of a stable object, a speech category. An auditory object is any piece of data that is recognised. All linguistic units - whether words, syllables or phrases are potential candidates for auditory objects. Hawkins also rejects the notion of the fixed order of recognition (phonemes are the initial objects recognised, then syllables, then words). She claims that people can in fact switch attention between different levels of analysis and a larger portion of speech (like a word of a phrase) might be identified at the same time or before a smaller part, corresponding to a lower level linguistic unit.

Having a wide experience with phonetic analyses, Hawkins holds a very interesting view on distinctive features. She acknowledges that listeners use acoustic cues to identify the speech input. She also recognises that it is difficult to find distinctive features with consistent physical correlates. On the other hand, she identifies a small number of features (such as voicing) that do have quite unambiguous acoustic correlates, especially in certain contexts. Hawkins proposes that these unambiguous features representing a relatively straightforward relationship between the linguistic construct and the physical correlate serve as perceptual anchors. The listeners recognise these anchors in speech input and then use them as a perceptual backdrop to judge other acoustic patterns in speech, in which there is no consistent relationship between the linguistic concept and a physical correlate. In other words, Hawkins combines in her approach the idea that speech perception is based on contrast, as proposed, for instance, by Kluender and Kiefte (2006), with the idea of absolute distinctive features. Her theory assumes the existence of certain absolutes that serve as a background for further context-based analysis for which contrast would be crucial.

Hawkins's theory is noteworthy for its attempts to combine several contradictory ideas - such as the idea of exemplars with the idea of abstractions. As Beckman, Edwards and Pierrehumbert (Beckman et al. 2007; Pierrehumbert 2003) she assumes that listeners create generalisations on the basis of exemplars, although unlike these researchers, she claims that these generalisations are created on the fly. An even more impressive feat of her approach is combining the idea of distinctive features with the idea of context-based recognition into one, compelling theoretical notion of perceptual anchors. This idea, as well as the idea of multiple levels of phonological representations proposed in Poeppel's model provide interesting insights into the discussion of phonological processing.

2.6. Conclusion: arriving at the common framework

The previous sections presented a roaster of theories trying to explain to the mechanisms of phonological processing. The reader was familiarised with models of speech perception grounded in linguistics, auditory phonetics, psycholinguistics and neurobiology. While oftentimes the researchers in those areas provide contradictory accounts of what phonological processing is, each science has provided useful insights into the nature of the mechanism.

Linguists pioneered the field, by showing that there is a level of phonological representation and that there is a need for a system that would translate the speech input into linguistic code. Therefore, in a way, linguists provided the basis for the notion of phonological processing, although their ideas about the mechanism of this processing were slightly simplistic. Research in auditory phonetics has shown that this mechanism is indeed incredibly complex, since it is very difficult to identify invariant units in speech. Auditory phonetic research has also indicated that phonological processing is a unique capability separate from other kinds of auditory processing. Nevertheless some theories point out that speech perception is also based on universal human skills such as change detection. The investigation of psycholinguistic research allowed for a deeper understanding of phonological processing. For instance, it suggested that speech segmentation is based on identification of particular speech sequences in the input, rather than on dividing the speech into smaller parts and then categorising the parts. However, segmentation is also facilitated by additional phonetic and phonotactic cues, as well as by statistical information about frequent speech combination in particular positions. The review of psycholinguistic data also indicated that there are two levels of representations in speech - one is the level of exemplars, that is actual instances of speech sequences, the other is the level of generalisations made on the basis of the exemplars (Port 2007a; Bybee 2002; Beckman et al. 2007). Regarding the granularity of representation, the data did not provide clear-cut answers, but they suggested the existence of at least some kind of phonological (sub-lexical) level of representation below the level of words. Finally, psycholinguistic research suggested that speech is processed at the phonological level and at the lexical level at the same time, that is that listeners recognise both speech segments and words in the stream of speech at the same time. Moreover, there is a feedback loop between the two levels. Finally, the neurobiological approaches provided deepen understanding of the phonological processing by providing interesting evidence pertaining to phonological representations. The model devised by Poeppel and colleagues (2008) indicated that there is not one, but at least two levels of phonological representations and provided convincing neuroscientific evidence for that. The other model, presented by Hawkins (Hawkins 2010a) solves the problem of distinctive features by indicating that there are some reliably recognisable features that serve as perceptual anchors in speech, while the rest of speech recognition is context-based, as proposed by Kluender and Kiefte (2006).

All in all, it seems that despite many controversies and contradictions in the field of speech perception research, a common framework of phonological processing can be drawn from all the presented findings. Within this framework, humans store exemplars of actual speech fragments. From these exemplars they can extract prototypical speech patterns that correspond to words, phrases, but also syllables (or sound combinations) and units below syllables (phonemes or diphone modulations). Whenever listeners encounter new stream of speech, they match sound patterns present in the speech with the sound patterns of snippets stored in their minds. The detection of sound patterns happens on at least two levels. One level is the level of rapid changes (that might correspond to detecting features constituting perceptual anchors or detection modulation in diphone sequences), another level is the level of slower changes that occur over the course of one syllable or a speech sequence of a syllable-length (150-300 ms). This level is possibly responsible for recognition of pitch changes, rhythm detection and other features traditionally referred to as suprasegmental. Though the detection of changes people can identify speech fragments of varying lengths (words, syllables, phrases etc.). Additionally, listeners use phonetic and statistical cues to help establish boundaries in speech segments.

To conclude, phonological processing is the mechanism of matching speech input with the speech representations stored in these minds. Additionally, phonological processing is facilitated by sensitivity to cues establishing unit junctures. As has been also indicated throughout this section, phonological processing is inextricably intertwined with lexical processing. Phonology and lexicon continually cooperate in human mind. This is the reason why phonological processing is involved in word learning. As will be seen in the next sections, the development of phonology and lexicon go hand in hand.

Chapter 3: Sounds turned into words: phonology and word learning in L1

3.1. Introduction

The previous chapter described a unified framework for phonological processing. Within this framework, the interrelations between phonology and lexicon are already quite noticeable. However, nowhere is the relationship between phonological processing and lexical processing as visible as in first language acquisition. The following chapter will describe the process of phonological development in children and explore its relationships with lexical acquisition. Since the focus of the thesis is phonological processing, the sections to follow will concentrate on speech perception process, while issues related to speech production (the developmental sequences in articulation etc.) will be omitted. The chapter will begin with the description of phonological development in infants - the period preceding neural commitment, the period of reliance of prosodic cues, the period of lexical restructuring. Throughout these initial sections it will be noted that various researchers provide contradictory accounts of phonological development in children. Therefore the section 3.2.5 will constitute an attempt at a unified vision of phonological development in L1. It will be proposed that in principle there are several developmental processes that children undergo when they acquire their L1 phonology. These processes include: initial prosodic processing, specification, abstraction and chunking. The final sections of this chapter will explain how each of these processes is related to lexical development in children.

3.2. Phonological development in L1

3.2.1. Perception of universal contrasts in infants

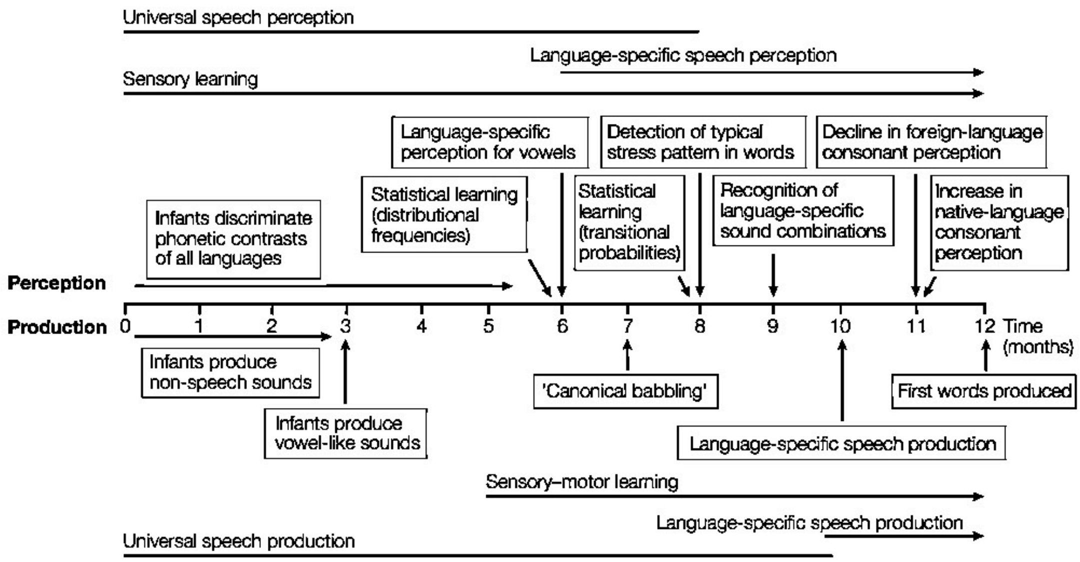

Figure 4: Stages of phonological development (adapted from Kuhl 2004: 832)

Phonological development of infants is usually considered from two points of view (Stoel-Gammon 2011). One is the point of view of production - the development of speech motor control and the ability to pronounce particular speech sounds or speech sound combinations. The other is the point of perception - the cognitive-sensory development allowing children to recognise patterns in speech and create speech sounds categories. Although these two are connected, in this thesis only the perceptual phonological development will be considered. For more information regarding the production patterns of young children see (de Boysson-Bardies 2001). In Figure 4 (Kuhl 2004), which depicts the stages of phonological acquisition in children up to twelve months, the perceptual phonological development is represented in the upper part. As indicated by this illustration, phonological development in infants begins with the development of categorical perception. Up to the 6^{th} month of age, infants can discriminate between all phonetic contrasts. In an experiment conducted by Eimas and colleagues, infants aged 1-4 months have been shown to be sensitive to differences in voice onset time (VOT)[1] as small as 20ms (Eimas et al. 1971). English babies up to the age of 8 months can discriminate between the Hindi retroflex and dental phonemes - a contrast not perceptible to adult speakers of English (Werker and Tees 1984).

3.2.2. Neural commitment - specialisation of perceptual contrasts

While in the first months of life infants are adept at recognising all kinds of phonetic contrasts, with time they become blinded to the non-native phonetic cues and start to be attuned primarily to the ones used in surrounding languages. This phenomenon is called neural commitment (Kuhl 2004). After the age of 8 months, children's sensitivity to all perceptual contrasts starts to fade for consonants - English infants in the study by Werker and Tees (1984) stop being sensitive to contrasts typical for Hindi at this age. This finding has also been confirmed in a recent neuroimaging study, in which a group of infants was exposed to native and non-native consonantal contrasts in MEG (Kuhl et al. 2014). Infants that were 7 months old activated their brain regions responsible for auditory processing equally strongly in response to native and non-native contrasts. Infants aged 11-12 months, however, activated the auditory brain regions more strongly in response to native consonantal contrasts. This is the pattern of activation that is typical also for adults and it seems to indicate, according to the authors, that children start to categorise native consonants like adults do around the age of 8-10 month. Moreover, according to Kuhl, around the age of 6 months, children develop categorical perception

[1] In the production of obstruent sounds, voice onset time indicates when speakers start to voice the sound, i.e. at what time during the production of the consonants their vocal folds start to vibrate. Sound with smaller voice onset time are considered voiced, sounds with greater voice onset time are voiceless. Languages of the world differ in terms of what VOT values will make a sound voiced and which would make a sound voiceless. For instance English voiceless plosives have a very long VOT (they are "very voiceless"), while Polish voiceless plosives have a moderately long VOT (they are "moderately voiceless")

of vowels consistent with the linguistic input they are exposed to (Kuhl et al. 1992; Kuhl 1991). In other words, six-month-olds start to organise speech sound categories around the most frequent realisations of speech sounds in the input they receive. Kuhl has demonstrated this process in an experiment involving the head-turn paradigm (Kuhl et al. 1992). In this study, 6-months-old Swedish and American infants were presented with prototypical English /i/ sound interchanged with Swedish /y/ sound. The children learned to turn their heads upon hearing a change in sound. Then they were presented with variants of the prototypical sounds and the experimenters checked if children perceived these variants as versions of the prototypical sound (i.e. categorised the variants along with the prototype). As it turned out, Swedish children were better at categorising variants of the Swedish prototype and American children at categorising the variants of the English prototype. This means, according to the authors, that half-year-olds start to create phonological categories on the basis of statistical information about the speech sound tokens they encounter. Moreover, Kuhl indicates that this specialisation and development of speech categorisation consistent with surrounding language input is necessary for correct language development. Children, who retain the sensitivity to non-native contrasts at the cost of native categorisation development are more prone to language disabilities later on (Kuhl et al. 2005).

3.2.3. Holistic representations of words in infants - later development

These findings seem to indicate that within the first year of life, children start to develop representations of phonemes (or segments) in their minds. However, many researchers in the field of first language acquisition claim that young children do not have developed phonemic representations before the acquisition of the first words. Instead, it is assumed that words are the earliest units of speech and some researchers indicate even that children store multi-word phrases as whole speech units (Menn and Stoel-Gammon 1995; Metsala and Walley 1998). These claims are grounded in the fact that at some point in development children recognise words on the basis of more holistic cues, seemingly disregarding phonetic detail. Children can recognise words (for instance, their own name) when they are as young as 4.5 months old (Mandel et al. 1995). However, at the initial stages of language acquisition their representations of these words are underspecified. This essentially means that children between the ages of 6 and 11 months have problems distinguishing between minimal pairs. For instance, Hallé and de Boysson-Bardies (1996) have found that 11-months-old children do not see the difference between familiar words produced correctly and the same words with the initial consonant altered (for instance with the voicing parameter changed - as in the non-word *ganard* instead of *canard* - "duck"). Both items are treated as familiar. Similarly, in the experiment carried out by Werker and colleagues (2002), 14-month old children who were taught a new word form *bih* did not notice when it was changed to another word form *dih*.

This is probably because within the first year, children learn to use different types of speech cues which are essential for early word recognition and segmentation - the suprasegmental cues (Curtin and Werker 2009). Between the age of 7 and 11 months, children pay greater attention to suprasegmental features (pitch, rhythm, intonation, affect etc.) than to segmental features (such as voicing or manner of articulation of a particular sound). As indicated by Curtin and colleagues (2005), 7-month-old American children attend to stress patterns in the stream of speech and prefer stressed syllables to unstressed ones. They also use the stress patterns to divide speech into words (taking the stressed syllable as the initial one) and such prosodic cues are of greater importance to them than phonotactic cues in establishing word boundaries (Mattys et al. 1999). In other words, children are fixated on holistic representations and suprasegmental cues for speech perception - so much so that 7.5 month infants have problems recognising a word if it is produced with a voice significantly different from the voice of the original speaker (Houston and Jusczyk 2000) or if it is produced with a different affect (Singh et al. 2004). With time, children become more attuned to the so-called segmental cues (such as voicing or manner of articulation). At the age of 10.5 months they can recognise words regardless of speaker's affect (Singh et al. 2004) and can even recognise a familiar word with an altered stress pattern (Vihman et al. 2004). 18-months-old children are also better at distinguishing between two phonetically similar words (Werker et al. 2002). The initial underspecification of children's representations has led many researchers, such as Marylin Vihman, to the conclusion that children start with phonological inventory containing simply whole words with little phonological structure (Vihman et al. 1994; Vihman 1996). According to these researchers, with time, the system becomes more hierarchical - children begin to store syllable patterns, however still largely underspecified. In the words of Oller and Steffens, "the child's syllables are more like rhythmic chunks with specifiable onsets and offset characteristics than sequences of segmental elements" (Oller and Steffens 1994: 53). In this view, the inventory of phonemes emerges in the child's mind gradually with time.

3.2.4. Lexical restructuring theory and lexical restructuring reversed

A similar view is held by researchers such as Metsala and Walley (1998), see also (Bowey 1996, 2001), who propose that the specification of children's phonological inventory happens as a result of vocabulary acquisition. Like Vihman (Vihman et al. 1994; Vihman 1996), Metsala and Walley assume that when children begin to acquire their first words, they store them as wholes - each word is a separate phonological category, which is not further divided into syllables or phonemes. However, as the child learns more and more lexical items, this system becomes highly inefficient. It is more effective computationally to have a system of phonological units that are universal and limited in number and are able to encode an indefinite number of words. Thus the growing number of words eventually leads to the partitioning of the holistic word representations into unit-based representations. This shift is gradual: the more lexical items children acquire, the more efficient they become at extracting phonological units from words - both due to practice and due to the necessity of handling an ever increasing lexicon. The specification (partitioning) of phonological representations is not global. Children might have specified (partitioned) representations for some words and not others. In principle, words that have more phonological neighbours (i.e. words that are phonologically similar, differing in only one sound) are more phonologically specified than words with fewer neighbours (Storkel 2002). This is because when children have the representations of many phonetically similar words in their minds, there is a greater need to specify these representations to keep the words distinct.

The presented theory, called lexical restructuring model, is supported, for example, by data suggesting children are better at doing phonological tasks on words with many neighbours (Metsala 1999). In her study 36 children aged 3-6 years were given two tasks tapping into awareness of phonological structure of words. In one task, the children heard onset and rhyme of the word (for instance *d* and *ish*), and they had to put the two parts together and point to the picture representing the resulting word. In another task, children had to combine a word from phonemes and also point to the picture representing this word. In both tasks 10 of the stimuli words had many neighbours and 10 had few neighbour. The scores (number of correct answers) were significantly higher for tasks in which words with many neighbours were used. These findings suggest that phonological development is not global, which is predicted by the model.

Partial evidence for the theory is also offered by correlation studies showing the relationship between phonological processing tasks and vocabulary size in kindergarten children (Bowey 1996, 2001; Metsala 1999). These studies have already been reported in chapter 1, but to remind the reader, they employ tasks in which participants are asked to recognise phonemes, word onsets or syllables (for instance by saying whether a group of words start with the same sound) and manipulate these units (for instance by saying a given word without one of the sounds). Children who are better at such tasks also have larger receptive vocabulary. The problem is that these studies cannot show the direction of influence. It is not known whether it is vocabulary that improved phonology or vice versa. However, a study by Girolametto and colleagues indicates that in fact learning vocabulary can improve phonological development (Girolametto et al. 1997). In the study, a group of two-year-old toddlers with speech delays were taught new words. As a result they started to use more consonants and more complex syllable structures compared to the control group who did not receive the vocabulary training. Although the results point to development of phonological inventory in production rather than perception, this study offers convincing evidence for vocabulary influencing phonological development. However, it seems that the relationship between vocabulary and phonology goes also in the other direction.

It appears that vocabulary learning is influenced by phonological development both on the level of production and perception. The influence of productive phonology on word learning is widely acknowledged. Several studies have shown that the amount of speech sounds children know and can produce determines what kind of words the children are going to say (Vihman et al. 1994; McCune and Vihman 2001; Locke 1986). As a result, children with larger productive speech sound inventory have also larger vocabularies (Smith et al. 2006). However, also the phonological knowledge on the level of perception seems to influence word learning. The theories that put forward this idea suggest that either knowledge of phonology or phonological processing in general plays a pivotal role in establishing the representation of the new word forms. For instance, Storkel and Morrisette (2002) claim that establishing common speech sound combinations in a given language and making these combinations more activated in their minds helps learners create the representations of new word forms typical for a given language more efficiently. This hypothesis has been supported by studies showing that children learn words with common speech sound sequences faster (Storkel 2001). In Storkel's study, children were asked to learn four short nonwords with a CVC structure. Half of the non-words were composed of common speech sound combinations and half was composed of rare sound combinations. The non-words were paired with novel objects and presented in a story. Following the exposure, children were asked to match the pictures to the words and to name the novel objects with the non-word name. In all tasks, but especially in the task requiring naming the object, children were much better if the non-word was composed of the common sequences. This finding shows that knowledge of preferable sound combinations in a given language facilitates learning new words of the language.

In Storkel's proposal, the phonological factor facilitating word learning is the knowledge of phonotactics - in other words, having representations of common speech sound combinations. Knowing which sounds are likely to go together and what to expect in a novel native word form helps the learner process and remember the new words. The role of sound combinations in word learning is also underlined by other researcher - Gary Jones (Jones et al. 2007; Jones 2012), although the focus of his theory is slightly different. According to Jones, the phonological factor facilitating word learning is the mechanism called chunking. The researcher assumes that at the beginning of language acquisition children recognise speech and create new representations of words phoneme by phoneme. However, as they encounter more and more words and similar sound combinations, they start to encode larger sequences of phonemes (called chunks) in their minds. This alters their phonological processing mechanism - they start to process speech by identifying larger chunks rather than individual phonemes and they also encode new words with the use of chunks. The chunking ability develops with time as the speakers learn to encode longer and longer phonological sequences. This makes speech perception more efficient, because instead of recognising each phoneme one by one during phonological processing, the listener simply identifies the whole chunk. As a result, children at the age of 4-5 years, whose chunking abilities are more developed, are better at repeating novel word forms than children at the age of 2-3 years. It is this greater efficiency in phonological processing that improves word learning. In Jones's opinion, the sudden increase in vocabulary learning rate during the second year of life, sometimes referred to as vocabulary spurt (Goldfield and Reznick 1990; Ganger and Brent 2004) is the result of chunking strategy being developed in children (Jones 2012: 1). And indeed he has managed to simulate the sudden growth of vocabulary occurring in real-life data with the computational model devised on the basis of his theory.

While Storkel and Jones focus on sound combination as the key to better vocabulary learning, there are studies indicating that the skill to recognise and manipulate phonological units (as tested by phonological tasks mentioned previously) also plays a role in the acquisition of new words. Scores on phonological tasks (phonological awareness tasks) have been shown to correlate with the ability to learn new word forms in laboratory settings. In one study (Windfuhr and Snowling 2001), 75 children aged 6-11 years were given phonological STM tasks, phonological awareness tasks and paired associates task. In the phonological STM tasks, the participants had to repeat lists of words and non-words. In the phonological awareness set there were two tasks. One was phoneme deletion, in which children had to say non-words such as *barp* with one phoneme deleted. The other was rhyme oddity, in which children had to choose from a set of four words the one that did not rhyme with the rest. In the paired associates tasks, children had to learn four non-word names of abstract shapes. Phonological awareness tasks were significantly correlated with the speed of learning the non-word names, while phonological STM tasks (both involving repetition of words and of non-words) did not predict learning efficiency at all. This result indicates that recognition of phonological units plays a role in learning new words.

Similar results were obtained in an experiment using Quick Incidental Learning paradigm, which is a procedure emulating natural word learning (Ramachandra et al. 2011). In the experiment, 40 children were presented with two stories, which had the form of two recorded narratives with an accompanying slideshow. In each story, experimenters introduced three novel names for invented objects. Thanks to the slideshow, the children could both see the objects and hear their names in context. Following each story, the researchers tested if children learned the associations between the objects and the names. Apart from the learning tasks, the participants performed a non-word repetition task and two phonological processing tasks. One was the rhyme oddity task (the child had to say which of the four words did not rhyme with the rest), one was alliteration awareness task (the child had to say which of the four words did not begin with the same sound as the rest). Both phonological tasks (but not the non-word repetition tasks) were significant predictors of the non-word learning performance.

The final piece of evidence for the relationship between the ability to recognise phonological units and learning new words is a training study performed by de Jong et al. (2000). In this study, 28 kindergarten children (5- to 6-year olds) were given either a two-week long phonological training or a conceptual categorisation training. The children in the phonological trainings had 10 15-minute daily sessions during which the teachers had shown them how to break words into phonemes and syllables and taught them 8 grapheme-phoneme correspondences. The children in the control group were given four sessions during which they learned to categorise objects semantically. All participants were tested on their letter knowledge and phonological awareness before and after the 14-day training period. The phonological awareness test involved rhyme oddity task and the alliteration awareness task similar to that in Ramachandra's study. Moreover, following the training period, all children had to perform a paired associates task, in which they had to learn novel names of four cuddly toys. Children who underwent the phonological training improved on the phonological awareness task and they performed better on the paired associates task than the control group. This indicates that a greater awareness of phonological structure and a better ability to identify small units in speech resulted directly in a better word form learning ability.

On the whole, the studies presented in this section strongly suggest that the relationship between lexical and phonologic development is bidirectional. While it seems that vocabulary increase results in more specified and mature phonological representations, it also appears that better phonological skills result in more efficient word learning.

3.2.5. What is phonological development

So far, this chapter introduced certain key studies related to the phonological development in children acquiring their L1. It has also been noted that this phonological development might be intimately related to vocabulary learning in children. There is, however, one problem with the research presented so far - it often promotes contradictory theories of what phonological development is and how it proceeds. These contradictions are also reflected in the studies on the relationships between phonological and lexical development. Since different researchers have different ideas about phonological development and phonological processing, they offer various interpretations as to the nature of the interrelations between these concepts. In the two sections to follow, an attempt to organise these claims will be presented. In this section, the inconsistencies between different theories of phonological acquisition will be pointed out and a common framework will be established. In the next section, the data on the relationship between word learning and phonological processing will be reinterpreted in the light of this framework.

The greatest chasm between different theories of phonological acquisition exists in their approach to the acquisition of phonological categories. The research described in the section on neural commitment (section 3.2.2) and conducted by researchers such as Kuhl or Werker and Tees (Kuhl 1991, 2004; Kuhl et al. 1992, 2005, 2014; Werker and Tees 1984) is taken as the evidence that between the age of 6 months and 12 months children establish the phonological categories of their native language. In other words, it is proposed that infants have representations of phonemes by the end of first year of life and that they essentially start the process of word acquisition with the knowledge of phonemes. This is also the assumption made by Gary Jones (Jones et al. 2007; Jones 2012). In his theory children start by recognising (and learning) speech phoneme by phoneme, but with exposure to linguistic input they start to encode larger sequences of phonemes in their mind. In this view, the phonological development that drives vocabulary is the development from phonemic analysis of speech and into the more holistic processing based on chunks.

However, as has been indicated in the sections on holistic word representations and lexical restructuring (sections 3.2.3 and 3.2.4), other researchers argue precisely the converse idea - that child's phonological development goes from the underspecified chunks to lower level phonological representations (Vihman et al. 1994; Vihman 1996; Bowey 1996; Metsala and Walley 1998; Metsala 1999). Menn and Stoel-Gammon claim even that children at the early stages of acquisition store multi-word phrases rather than words (Menn and Stoel-Gammon 1995). As is easy to observe, these two theoretical approaches stand in complete opposition. One approach sees phonological development as the movement from perception of small phonological units to perception of whole words, the other as the movement from whole word perception to perception of small phonological units. What is problematic is that both approaches are based on empirical data and/or computational modelling. This leads to a conclusion that in some way both of them must be right.

In this situation, it would be desirable to propose a unified framework that could accommodate both of these stances. Such a theory might draw inspiration from the PRIMIR framework proposed by Werker and Curtin (2005) and from the dual representation framework proposed, among others, by Beckman, Edwards and Munson (Beckman et al. 2007; Munson et al. 2012). Both of these theories differentiate between *perceptual attunement* to features, which enables children to make distinctions crucial for a given language and *creating functional abstract speech categories* that are further actively used in the process of speech perception. In PRIMIR, it is assumed that language acquisition occurs on three planes. One of these planes is the so-called general perceptual plane containing phonetic information (including both segmental details and rich phonetic details allowing, for instance, to identify a speaker), one is the phoneme plane containing abstract phonological categories and one is the word plane containing the representation of words (Werker and Curtin 2005). The theory of Beckman and colleagues (2007), which has already been described in the section devoted to the phonological processing in psycholinguistic approaches (section 2.4.2), assumes two planes of representations - the exemplar level (containing rich phonetic detail) and the level of abstracted categories. In both theories, it is assumed that up from the 6th month of life children become attuned to acoustic cues that are meaningful in a given language. At this age children establish perceptual preferences that guide further acquisition process. These preferences, however, are not abstract categories that are actively utilised in the speech processing. These are not the abstract representations in Beckman's theory or the units in the phoneme plane in PRIMIR. They belong to the general perceptual plane.

At the same time, 6-month-old children are very much attuned to the suprasegmental features of their native language such as stress, pitch, rhythm, intonation, indexical features of speech. It is these features that guide the receptive acquisition of first words (Curtin and Werker 2009; Werker and Yeung 2005). In both theories, these words are stored as exemplars, and as indicated by the acquisition literature, these exemplars are largely underspecified. One reason for that is that since the segmentation of speech into words is based on suprasegmental features, segmental information is not well integrated with the representations of words. This integration starts around 10.5 months, when word representations become more specified. With the increase in vocabulary size and the greater specification of representations, abstracted categories (representation on the phoneme plane) start to emerge, just as predicted by lexical restructuring theory (Metsala 1999; Beckman and Edwards 2000; Beckman et al. 2007). According to Beckman, Munson and Edwards (Beckman and Edwards 2000; Beckman et al. 2007; Munson et al. 2012), creating such phonological representations might facilitate further word learning. Abstract sub-lexical representations help children map the speech sequence that constitutes a new word into schematic phonological units in their minds and in this way allow for creating the representation of this word more efficiently. To support this claim the authors quote the study by Merriman and Schuster (1991), in which a group of children were presented with a non-word such as *japple* and were asked to point to a picture that represented the word. One picture was the image of an apple, the other was a picture of an unknown, abstract object. 2-year-old children universally pointed to the picture of an apple, while 4-year-old children pointed to the new object. According to the authors, these findings indicate that older children have more abstract representations that allow them to learn a new word instead of assimilating the new word to the old one. Beckman, Munson and Edwards assume that these sub-lexical representations are phonemes. However, the representations might be phonemes, diphones, pieces of speech sequences of a fixed length, syllables etc. The realisation that the abstract representation might take any form solves simultaneously several problems. First of all, it accommodates the results indicating that children have problems with perceiving phonemes in speech until literacy training. Second of all, it opens a possibility for accommodating Jones's chunking model into the framework. Once children start to extract the basic abstract units (whatever they are), it is possible that they start to combine them into larger, abstracted units - the chunks. These chunks would probably be the common speech sounds combinations that Storkel mentions in her studies (Storkel 2001, 2004).

The model presented above suggests that phonological development can be broken down into four processes. In this thesis, these processes will be called *initial prosodic processing*, *specification*, *abstraction* and *chunking*. The *initial prosodic processing* refers to the process in which children divide speech into smaller units using prosodic cues. Around the 7th month of age children become more focused on suprasegmental cues - they attend to the slow sound modulations connected with pitch, rhythm etc. It seems that this initial speech processing mechanisms allows young language learners to segment speech before they actually have any representations of segments in their heads. This is the kind of pre-lexical processing that triggers the acquisition of new words. As indicated by Curtin and colleagues (2005), American children use stress patterns to establish word boundaries. Since most words in English start with a stressed syllable, children segment speech by attending to stressed syllables and treating them as word beginnings. Apart from the suprasegmental cues, children use also distributional information (Saffran et al. 1996a, 1996b), but the early, primitive phonological skills related to the detection of suprasegmental cues are crucial for initial word learning for children below the age of one year (Mattys et al. 1999). As the initial segmentation mechanism, initial prosodic processing should also play a role in the creation of phonological categories at the later stages of development, because it provides the learners with exemplar "speech-pieces" that can be further fine-tuned and abstracted. This idea has been confirmed in several studies showing in which the detection of pitch changes and the detection of amplitude changes have been named as instrumental in segmenting speech and creating the L1 phonological representations (Richardson et al. 2004; Goswami et al. 2011).

Specification refers to the processes of fine-tuning the exemplar-level representations by adding relevant phonetic details to them, while *abstraction* is the process of developing abstract phonological representations. These are two separate processes that extend from the age of 2.5 years into at least early school years (ages 7-8), and possibly further. The creation of abstract categories begins with the sudden vocabulary growth and is probably solidified with the literacy training (de Jong et al. 2000), although some researchers claim that sharpening of categorical boundaries continues well after the age of 12 years (Hazan and Barrett 2000). At the same time, learners specify their exemplar representations and become more attuned to the minute phonetic details that are relevant for the recognition of words in their language (Edwards et al. 2002). In their 2007 review article Beckman and co-workers quote two studies that tap into the two processes. In one study (Edwards et al. 2002). typically developing children aged 3-4, 5-6 and 7-8 and a group adults were presented with two pictures illustrating two minimal pairs (*cap* vs *cat* and *tap* vs *tack*). Then they heard a word and they needed to identify it by pointing to its referent in the picture. There were three conditions in the task. In the first condition (the whole word condition), participants were presented with whole words, in the second, gated, condition (the 20 ms gate condition) experimenters cut the plosive burst from the final, crucial consonant, making it more difficult to recognise, in the third condition (also gated), the experiments further distorted the final consonant by cutting the vowel immediately following the formant transition (the 40 ms gate condition). Only older children (7-8) and adults could identify words in gated conditions. The recognition of words in the whole word condition improved with age. These results indicate that children become more sensitive to relevant phonetic details with time. In other words, their phonological representations become more specified. In the second study (Edwards et al. 2004), children aged 3-4, 5-6 and 7-8 alongside a group adults were asked to repeat non-words containing frequent and infrequent diphones. As could be predicted, the older the group the better the performance on the task. In all groups, the non-words containing infrequent sequences were more difficult to repeat. However, the difference in performance on frequent and infrequent sequence non-words diminished with age. The most reliable predictor of this diminishing was the vocabulary size of the participants. The authors claim that this result proves that children use their experience with a vocabulary of a given language to create universal sub-lexical representations in this language and that they use these representations to parse novel word-forms. In the 2007 review, Beckman and colleagues indicate that the two processes tapped by the two research paradigms (Edwards et al. 2002 and 2004) can be selectively impaired. In children suffering from phonological disorders, the specification of representation is impaired. These children have problems with the minimal pairs task, however, in the non-word repetition task with frequent and infrequent sound sequences they display developmental pattern similar to those of typically developing children. On the other hand, in children with Specific Language Impairment (SLI), the abnormal behaviour on the non-word repetition task is most pronounced. Children with SLI have problems pronouncing low frequency sequence non-words late into the development, indicating an impairment in the ability to create abstract sub-lexical representations.

The fourth process of phonological development is *chunking*, which might be defined as the process leading to the internalisation of the most frequent sound combinations in a given language. The importance of sound combinations has been evoked by two researchers in the present literature review - Holy Storkel and Gary Jones. Both point to role of the common sound combination knowledge in novel word learning. In Storkel's papers (2001, 2004, 2002), these representations have the form of activated networks. Storkel's theoretical framework, which is based on the theory by Luce et al. (2000), assumes that abstract phonological categories form interactive activation networks. Whenever a person sees a word, the phonological units comprising this word are activated together. Phonological units that are frequently activated together form a strong connection. As a result when one unit is activated, the unit that frequently co-occurs with it is also activated. Therefore, when the listener encounters a new word with common speech sequences, this word is processed faster, because once one of the phonological unit is activated, the rest is also already pre-activated in the mind of the speaker. This is why storing common speech sequences constitutes an advancement in language development - it makes speech processing faster, by pre-activating the units most likely to occur in a particular context.

Jones's explanation (Jones 2011, 2012) for the role of speech sequences representations in language acquisition is slightly different. Like Storkel, he assumes that the representations of common speech sequences happens as a result of linguistic experience, but he does not utilise the interactive activation theories to account for that. He merely states that when human mind encounters a speech sound combination several times, it stores this combination as a separate unit. This is a computational strategy to make word processing more efficient. The efficiency stems from the fact that each word that is processed needs to be held in the phonological STM. As has already been mentioned in chapter 1, phonological STM's capacity is limited, but it can be enlarged by chunking the information. In his 2012 paper, Jones argues after Zhang and Simon (1985) that encoding each speech chunk takes 400 ms and additional 30 ms for each phoneme apart from the first one. In a 4-phoneme word, if each phoneme is one chunk, then the encoding would take 1600 ms (4*400ms). If there are two chunks with two phonemes each - only 860 (2 * (400 ms + 30 ms)). Storing the word as one chunk would take only 490ms. Since STM's capacity is limited to 2000 ms, chunking allows to encode more within this module. Therefore, this model assumes that the increasing chunking ability is the strategy in phonological development that serves to make lexical processing more efficient.

On the whole, it might be proposed that the phonological development of the child consists in establishing well-defined speech representations in the long-term memory and developing strategies of matching speech input with these representations in an efficient and accurate manner. This happens through four different developmental strategies. One of them is initial prosodic processing, which is the segmentation of speech stream on the basis of prosodic cues occurring at the initial stages of L1 acquisition. Another is specification, which involves enriching child's exemplar representations by a number of relevant phonetic details and in this way allows for better identification of the known words. Yet another process is abstraction, which involves creating abstract phonological categories, allowing for efficient processing of novel words. Finally, the process of chunking, or storing speech combinations common in a particular language helps process novel lexical forms more efficiently.

3.3. How is phonological development related to word learning in L1?

The previous section was an attempt to describe the phonological development in a fairly unified fashion by identifying its four constituent processes. What is truly astounding is that each of these processes related in some way to vocabulary development.

The first phonological developmental process is the initial prosodic processing. This mechanism is crucial for word learning, because it allows for the extraction of first words from the stream of speech, before the children acquire any phonological representations (Curtin et al. 2005). It is also possible that it acts as a trigger for further development of phonology and lexicon by providing the learners the first exemplar speech representations. Further down the line, the processes of specification, abstraction and chunking also influence lexical development, each in its specific way. Specification, according to Beckman and colleagues, is connected with the recognition of known words (Beckman et al. 2007). As children's representations of words become richer and filled with relevant phonetic detail, their recognition of words becomes more reliable. As indicated by Edwards et al. (2002) the knowledge of phonetic detail helps distinguish between two similar words. Richer representations allow for differentiating the known word from other known words, while at the same time it allows for sufficient generalisations over features such as voice quality. Abstraction, on the other hand, is crucial for the processing of unknown words. To quote Munson and colleagues, the development of abstract phonological categories "allows children to interpret novel word-forms as combinations of known categories. This ability then allows children to form representations for novel strings more efficiently than if these strings were interpreted solely relative to existing articulatory and auditory representations" (Munson et al. 2012: 299). In other words, abstraction process facilitates new word learning by facilitating novel word processing. This idea is supported by studies such as the one by de Jong and colleagues (2000). Finally, chunking, understood as the development of abstracted representations of common sound sequences, helps children learn novel word forms in their native language more efficiently. In the theoretical framework proposed by Storkel (Storkel and Morrisette 2002; Storkel 2001), this happens because the common speech sequences are more activated in their mind and this pre-activation speeds up forming a representation of word containing these sequences. The computational model proposed by Jones, on the other hand, proposes that the representation of common speech sequences helps remembering new words, because in short-term memory, chunked information takes less space, is more efficient and faster to process than non-chunked information (Jones 2011, 2012).

3.4. Conclusion

To conclude, phonological development and lexical acquisition go hand in hand in children learning their first language. Initial prosodic processing allows for initial word learning. Word learning allows for the creation of abstract phonological representations. Abstract phonological representations (and chunking of them) allows for more efficient novel-word learning. As has been shown in this section, all four processes of phonological development - initial prosodic processing, abstraction, specification and chunking - are intimately connected to lexical development in L1 acquisition. What has largely been unexplored, however, is how this relationship between phonological processing and lexical acquisition would work in L2 learning. As will be shown in the next chapter, L2 phonological acquisition is in many respects similar to L1 phonological acquisition. However, it is also different in that in occurs in older learners, with largely established inventory of phonological representations. The question thus is, whether the phonological development will be as intimately linked to word learning in L2 as it is in L1. The next chapter will try to systematise what is known about L2 phonological development and will describe the data available on phonological processing and word learning in a foreign language. As will be shown, there is a strong theoretical basis for the relationship between phonological and lexical acquisition in L2, but not many empirical studies that would test the theoretical assumptions. Thus, the next chapter will provide the basis for the study presented in this thesis.

Chapter 4: Sounds of another kind: L2 phonology and L2 word learning

4.1. Introduction

The two previous chapters concentrated on establishing a common theory of phonological processing in mature language users and on extending this theory to language learning. It has also been shown that phonological processing and lexical development are largely intertwined in first language acquisition. The following chapter will try to extend these theories into the second language acquisition (L2 acquisition). As has already been mentioned, the subject of the interplay between phonological processing and lexical acquisition is severely under-researched and thus the investigation of this subject is at the core of this thesis. The following chapter will lay the theoretical foundation for the experimental work into phonology and word learning in the second language, presented in the experimental chapter. The chapter will begin with the description of theories pertaining to phonological development in L2. Then studies on the relationship between phonological and lexical development in L2 will be presented. Finally, additional factors that might play a role in these studies will be pointed out to the reader. The chapter will end with description of the current study that is designed to answer questions about the relationship between phonological processing and word learning.

4.2. Theories of phonological acquisition in foreign language learning

The acquisition of L2 is interesting in that it combines the processes present in adult speech processing with those present in L1 acquisition. Individuals learning L2 have developed inventories of speech categories and mature phonological processing strategies. Yet, they are faced with the task of processing a different kind of speech with different prosodic features, different speech sounds and different speech sound combinations. In L1 acquisition, phonological development has been subsumed under the four skills: initial prosodic processing, specification, abstraction and chunking. As will be indicated in this chapter, all of these skills should also take place in L2 learning. One of the problems in L2 acquisition, which has been pointed out, for instance, by Carroll (2004, 2006) is that speakers need to develop some initial strategies to segment the "wall of sound" that is L2 speech. They cannot simply identify L2 sounds and words in the speech stream, because they still do not have the representations of these sounds and words. Therefore, just as in the L1 acquisition babies use initial prosodic processing to help them with the initial segmentation of speech into words, so presumably older learners of L2, need similar strategies to make sense of the new type of speech input. Initial prosodic processing is however not the only process utilised in L2 learning. Also specification, abstraction and chunking need to be employed in the phonological L2 acquisition; for most languages, L1 exemplars, speech categories and chunks do not map well into L2 speech and they need to be adjusted to the new language. In the sections to follow, each of the phonological developmental processes will be described in the context of L2 acquisition. In the next section, the processes occurring during initial contact with L2 (such as initial prosodic processing) will be explored. Then, further developmental processes in L2 acquisition will be described. The theories of speech categories formation (and adaptation) in L2 will be presented, followed by the description of specification and chunking in the foreign language.

4.2.1. The initial contact

There are several studies on the phonological development in the second language (L2), but very few of them examine how listeners deal with the foreign input during very early stages of acquisition or at first exposure to the new language. Consequently, very little is known about how the acquisition of a foreign language is initialised. As pointed out by Carroll (2004, 2006), individuals encountering a completely new language often experience "a wall of sound" - they perceive the foreign speech as a continuous stream of auditory input and have problems even dividing this stream into words. And yet, after some time they start to hear distinct words and units within the stream and they start to acquire the language. Where do the process begin? How does one go about tackling the "wall of sound"?

As indicated in chapter 2, mature speakers of a language analyse the stream of speech by recognising phonological and lexical items and matching them with the representations stored in their minds. But older learners acquiring a new language do not have phonological or lexical representations of this language. They have linguistic representations in their first language (L1) and these might be only partly useful when dealing with a completely new system. In many ways foreign language learners are like infants acquiring their L1 - they are faced with a stream of speech and need to manage it in some way to initiate the learning process. It has been indicated in the previous chapter on L1 development that infants begin the acquisition process by segmenting speech stream into smaller parts using prosodic (along with statistical) cues. These parts are the basis for the first holistic representations of words and speech units. It is probable that a similar mechanism takes place in the foreign language acquisition. Most probably, adults utilise both their L1 knowledge and those more universal strategies (initial prosodic processing) when faced with a foreign language. This seems to be the conclusion of the few studies that deal with the initial speech segmentation strategies during initial exposure to the foreign language (Rast 2010; Shoemaker and Rast 2013; Saffran et al. 1996a; Mirman et al. 2008). These studies indicate that adults learning L2 use, roughly speaking, three strategies: 1) applying L1 categories and segmentation rules to foreign speech, 2) using initial prosodic processing to identify word and segment boundaries and acquire knowledge about the phonological structure of the new language, and 3) using the newly acquired statistical knowledge about phonological structures to find unit boundaries.

The application of L1 knowledge to the processing of a foreign language at the first exposure has limited utility, but nevertheless plays a role in foreign language learning even at the very initial stages. First of all, research shows that speakers use the knowledge about the prosodic structure of L1 to find word boundaries in foreign speech. Second of all, learners will try to identify L1 segments or lexical units in the foreign speech where possible. The first line of evidence comes from studies conducted by Cutler and colleagues (1986, 1992). The basis of this research is the observation that different languages use different prosodic cues as a help for establishing word boundaries. In English, which is a stress-based language, speakers tend to take the beginning of a stressed syllable as the beginning of a word. In French, which has a very clear and unambiguous syllable structure, speakers divide speech into syllables. Cutler's studies indicate that French listeners tend to use the syllable-detection strategy also when they listen to English words. English listeners, on the other hand, are attuned to the stressed syllables even when faced with French words. Those tendencies have been observed not only for the monolinguals faced with a novel speech inputs, but even for French-English bilingual speakers. This means that speakers draw on the L1 segmentation mechanism when analysing the foreign language at both initial and later stages of acquisition.

The second line of evidence comes from studies by Rast (2010) and Shoemaker and Rast (2013). These studies evaluate linguistic processing at first exposure to the foreign language. In both studies, a group of French speakers has been signed into a short-term intensive course of Polish. The authors of the studies tested the participants within the first hours of exposure to Polish as well as before exposure to the foreign language. In one of the studies (Rast 2010), the participants of the study had to engage in a production task - repetition of Polish sentences. The experimenters investigated which words in the sentences were produced more accurately and which less accurately by the speakers. In the other study (Shoemaker and Rast 2013), participants were asked to listen to a set of sentences in Polish and assess after each of them if a particular word occurred in the sentence or not. In both studies, researchers controlled for several factors in the test words - their frequency in the classroom setting, their transparency (if the words resembled existing French words/were cognates of French words) and the position of the word (whether the word occurred at the beginning of the sentences, the middle or the end). In both studies, transparency was a significant predictor of performance on the task. In other words, speakers were better at producing and recognising in the stream of speech those Polish words that resembled French. This finding suggests that learners used their L1 lexical knowledge to process foreign speech and were able to identify the L1 categories in the foreign language. The study by Rast (2010), additionally found that learners produced Polish words better, if they were phonologically similar to French (contained clusters and consonants attested in French). These results further confirm that whenever possible, learners try to use their L1 categories or L1 processing strategies when dealing with foreign speech.

While Rast and Shoemaker's studies have shown the involvement of L1 in foreign language processing, they have also demonstrated the use of other universal strategies for dealing with foreign speech – for instance, pause detection and prosodic processing. Both studies found that the performance on recognition/repetition tasks depended on the position of these words in a sentence. Participants were better at producing and extracting words that occurred at the beginning or at the end of the sentence. The authors assume that these words were easier to extract, since they occurred close to natural pauses in speech. The words occurring in the final position were even easier to detect than the words in the initial position, possibly due to the phrase-final lengthening effect. Speech units at the end of sentences are lengthened and thus easier to extract. These results indicate that when encountering spoken foreign language, learners apply primitive language-universal strategies, such as sensitivity to pauses and to certain prosodic cues to segment the speech and thus gather phonological knowledge.

This phenomenon is similar to the prosodic processing that occurs in infants (Jusczyk and Aslin 1995; Jusczyk et al. 1999). As has been already indicated, children use detection of prosodic patterns to segment speech (Mattys et al. 1999), but prosodic cues are also used by adult speakers to ascertain the boundaries of linguistic units in their L1. As has been indicated in chapter 2, in mature language speakers, using prosodic for segmentation has mostly an auxiliary role. Adults mostly use the post-lexical strategies to segment speech, with prosodic cues providing additional cues about unit boundaries along with segmental and phonotactic cues. (Mattys et al. 2005). However, as pointed out by Mattys (2004), the more noisy or distorted the signal, the greater the role of prosodic cues over other type of cues (phonemic and lexical) in speech segmentation. This type of information seems to be the one that listeners fall back on when faced with complicated and ambiguous signal. Therefore it stands to reason that is should also play an important role in the initial speech processing in a foreign language. The theory of primacy of prosodic cues at the initial stages of L2 acquisition has also been proposed by Hatch (1978), who claimed that during the initial contact with the foreign stream of speech, listeners process is by detecting stress and pause. While stress might not be the best prosodic cue for dealing with the speech in new language, since languages use different means to express stress (length, amplitude etc), speakers might be attuned to the changes in some general acoustic features - such as tone (pitch) (Carroll 2006). This hypothesis is partially confirmed by the data showing that when faced with a foreign language, listeners treat the fall in intonation (pitch) as a marker of sentence end (Henderson and Nelms 1980). All in all, several strands of evidence point to the fact that speakers encountering a stream of foreign speech use initial prosodic processing to help them detect initial boundaries in speech.

It is probable that with time, speakers might acquire more specific knowledge of phonological patterns in a particular language and use these patterns to process speech. With time, they will also presumably gain more general knowledge about the phonology of the new language - including the knowledge of segments. Shoemaker and Rast (2013) concluded in their study that the participants needed only 6.5 hours of exposure to Polish to gain some generalisable phonological knowledge of Polish. As has been mentioned before, during the first contact with the language, speakers were able to recognise Polish words mainly if they were placed at the beginning or end of the sentence, presumably using prosodic cues such as tone change, phrase lengthening or pause. Words in the medial position were recognised at almost a chance level (56.3% of the time). However, after 6.5 hours of exposure to Polish the recognition rate jumped up to 75.2%. This means that in the course of the few hours some learning occurred that facilitated recognition of Polish units in the stream of speech. The authors argue that this learning was phonological, which means that the participants learned about phonological patterns that occur in Polish and used them to segment speech, but that they did not gain and consequently use lexical knowledge in the foreign language. In other words, it is not the case that participants learned the word forms during the 6.5 hours of Polish class and therefore recognised them in speech. The authors made sure that this was not the case by carefully controlling the frequency with which the stimuli words occurred during the Polish classes. Half of the words used in the detection task were frequently used during the course, the other half was infrequently used. If participants were to internalise any word form during the course, it would be the frequent ones, and consequently the frequent words should be easier for them to extract in speech. However, this was not the case. There was no difference in recognition rate between the frequent and infrequent words, which indicates that the learning that occurred in speakers was phonological and not lexical.

In fact, the study by Rast (2010) suggests that listeners start to learn the words and use the lexical information for phonological processing only after 8 hours of exposure. In this study participants were tested on a Polish sentence repetition task after 4 hours of the Polish course and then after 8 hours of the course. The experimenters assessed participants' accuracy of production for frequent and infrequent Polish words in both time conditions. Only after 8 hours of exposure was there an effect of frequency on performance. This means that around this time learners started to develop lexical representation (the representation of word forms) in Polish and use them to process speech. All in all, the data collected in both studies (Rast 2010; Shoemaker and Rast 2013) suggest that after a relatively short period of reliance on general prosodic cues, language learners start to acquire knowledge about the phonological structure of the foreign language and use it to process speech. Then they start to acquire lexical representations and use these for analysing the foreign language input.

Among the knowledge that the speakers gain about the language during exposure is the statistical information about common speech sound combinations and combinations marking word boundaries. As has been indicated chapter 2, among the auxiliary cues used to ascertain word boundaries in speech are transitional probabilities. In other words, speakers use statistical information about the relative frequency of particular sound sequences to judge whether a combination of two sounds in the stream of speech occurs across word boundaries or within word boundaries. Using this kind of knowledge has been evidenced in children as young as 8-months old (Saffran et al. 1996b). It seems that adults learning a new language acquire this information very quickly and use it to segment speech. In a famous experiment by Saffran and colleagues (1996a), a group of participants was listening for 20 minutes to a stream of 6 trisyllabic "words", composed of simple English CV sequences (for example *babupu*, *patubi*), and produced without any pauses by a speech synthesiser. After 20 minutes of such priming, the participants were given a task, in which the artificial "words" from the stream were paired with foils that were composed of the same syllables but in different configurations. The participants had to indicate which item in the pair seemed more word-like. The majority of participants chose the "words" encountered in the stream, showing that they internalised the transitional probabilities of the new language within minutes of exposure and used them to establish word boundaries in the language. A similar study (Mirman et al. 2008) has shown that participants can be given as little as 7 minutes of exposure to the stream of speech and still learn about the distributional probabilities in this stream. In this study, the participants were exposed to bisyllabic artificial words. Then they performed a task, in which they had to learn the association between four words and four made-up objects. Part of participants learned "words" from the speech stream, part learned "part-words", that is the bisyllabic items, for which the transitional probabilities between the two syllables were typical for word boundary. The participants performed much better when they were assigned to the "word" condition rather than the "part-word" condition. To conclude, both studies indicate that the adult participants were able to extract the statistical information about the phonological sequences in a new artificial language in a matter of minutes.

Of course, in both studies, the stimuli used speech sounds from participants' L1 and conformed to L1 phonotactic cues, therefore the results cannot be that easily generalisable to real-life learning of an L2. Usually, when speakers encounter a new language, they have to familiarise themselves with a foreign accent and a different phonological inventory on top of the transitional probability statistics. Moreover, the knowledge of their L1 can often interfere with acquisition of L2 transitional probabilities. This is what has been shown in a study by Finn and Hudson-Kam (2008). The experimenters created two sets of artificial words, beginning with double consonant clusters. In one set, the words were conforming to the rules of English phonotactics, i.e. contained only sound combinations that are possible in English (for instance, /plodu/). In the other set, the consonant initial consonant clusters were illicit in English (for instance, /tfobu/). Then the participants were given a test, in which they had to indicate which items from a pair was more word-like. Two kinds of pairs were created. One was "word" vs. "non-word", where the "non-words" were created by taking the first syllable from "word" and matching it with a completely different syllable. The other kind of pair was "word" vs. "split-word", where the "split-words" were created by splitting the exposure stimuli across the consonant clusters (so if the exposure stream contained "words" /kmodu/, /θmʌre/ and /tfobu/, the "split-words" were /moduθ/, /mʌret/ and so on). It turned out that participants who did the tasks with the illicit "words", did not have a clear preference for the "words" over "split-words". In other words, L1 phonotactics interfered with their acquisition and use of the new languages' transitional probability.

This interference was present even when the participants were exposed to the language stream for a longer period of time and was averted only when additional cues (pauses) were introduced to the exposure stream to mark the "word" boundaries. These results indicate that statistical information at the initial stage of new language acquisition works best as a factor in phonological processing when it is combined with other cues - such as pauses. This finding is further supported by data from Experiment 2 by Saffran et al. (1996a), indicating that people are much better at choosing "words" over foils if the end of the word is marked by a prosodic cue that is vowel lengthening. Overall, it seems that using transitional probabilities and other statistical information constitutes only one of crucial ingredients of processing foreign speech during the initial stages of acquisition and it is greatly facilitated by initial prosodic processing.

To conclude, the evidence presented in this section, although sparse and to some degree speculative, indicates three ingredients that play a role at the initial stages of language acquisition when the adult language learners try to break through the "wall of sound" of the foreign speech. First of all, learners try to use their knowledge about L1 (representation, processing strategies) to make sense of the L2 input whenever it is possible. Second of all, they try to utilise initial prosodic processing to extract words and phonological structure of the new language from speech. Third of all, they will establish transitional probabilities of the language to decide where the word boundaries are likely to occur. These three strategies are probably used together, and form a basis for further phonological development (as well as lexical development, further down the line). As has been mentioned in this section, there are reasons to believe that the phonological development in the second language occurs very quickly - even within hours of the initial exposure. Part of this development is probably the acquisition of transitional probabilities, but it also seems that the listeners need some adjustment to their categorical perception of L2 speech sounds in order to fully process the foreign stream of speech. The next section will focus on the development of this categorical perception and other aspects of L2 phonological development following the initial exposure.

4.2.2. Developing L2 phonology

As already indicated, phonological development in L2 should follow three paths: specification, abstraction and chunking. However, those processes will not be the same as the ones occurring in children learning their first language. In the case of adult language learning, the presence of another phonological system (L1) will significantly impact the development.

The specification and abstraction occurring in adult language learning has best been described by the two most famous models of L2 phonological development: the Speech Learning Model (SLM) by Flege (1995, 2002) and the Perceptual Assimilation Model by Best (Best 1995; Best and Tyler 2008). Both models share an assumption that has been considered groundbreaking when Flege presented it in his first papers (Flege 1995): that human ability for learning phonology remains relatively stable over time. Flege makes the point very explicit in his articles and claims that while late L2 speakers usually have problems acquiring native-like pronunciation and speech perception, this does not stem from maturational constraints. The problem with native-like accent and perception in adult foreign language learners result from the fact that these people already have a phonological system in place and that any new categories acquired in the course of foreign language learning will occupy essentially the same space (Flege 2002). Thus phonological development in L2 involves rewiring the current phonological system (L1) already present in the mind. The foreign accent (and non-native perception) comes from the interaction between the two language systems and depends on how much each of the languages are used. A compelling evidence for this claim has been provided in the study conducted by Flege, Yeni-Komshian and Liu (1999) on Korean-English bilinguals living in the USA. The results of this study indicate that the more years the participants spent in the USA, the more native-like their English seemed. At the same time, the better their English, the more foreign accent they displayed in Korean. Flege concludes that this is due to both phonological systems sharing the phonological space and influencing each other.

In the section to follow SLM and PAM will be used to describe how the common phonological system is changed under the influence of L2. The two models will be discussed within the framework proposed in this thesis, i.e. a framework assuming that there are two levels of phonological representations: exemplars and abstracted categories. It must be noted that these assumptions are not native SLM and PAM. SLM assumes one level of representation - the level of phonetic categories, while PAM assumes the existence of two levels, but conceptualises them differently than the author of this thesis. Best differentiates between the the phonetic level and the phonemic level, by pointing out that learners might notice a difference between two sounds, yet still find them equivalent (Best and Tyler 2008). For instance, learner might hear the difference between English retroflex liquid and Polish alveolar trill and still categorise them as one sound: /r/. The terms "phonological level" and "phonological category" are used in Best's paper heuristically though, because the author does not believe that there are actual phonological categories in people's mind. The phonological category for her is more of a formal term, denoting minimal lexical difference. However, both PAM and SLM can be very easily reformulated in terms of exemplars and abstracted categories and can be used to provide a solid theory of L2 phonological acquisition within this framework.

The investigation of both models, but especially PAM (since this model explicitly names two levels of phonological representations) would lead one to the conclusion that the exposure to foreign language influences both exemplars and abstract categories (phonetic and phonological levels in Best's terms). The two influences are quite difficult to tear apart, since the change in exemplars would change the boundaries of abstract categories. This means that specification (development in the domain of exemplars) and abstraction (the development in the domain of the abstract phonological categories) in L2 acquisition go hand in hand. Nevertheless, the sections to follow will try to tear these two apart. The next section will focus on the formation of new categories under the influence of foreign language input (abstraction), then the attention will be switched to the changes in general perception of acquisition cues and changes within phonological categories (specification). Finally, the discussion of phonological development in L2 will finish with the discussion of chunking.

4.2.2.1. Abstraction

Abstraction in L1 involves the formation of phonological categories. In L2, however, it involves either formation of new categories or changing the old L1 categories to fit L2 input, depending on how similar the L2 sounds are to the L1 sounds. In Flege's SLM, these two types of process in L2 phonological development are called dissimilation and assimilation. Dissimilation entails creating a new phonological category for a sound. This is a process similar to category formation in children acquiring their first language. However, since in adults the process takes place in the common phonological space already occupied by L1, it can result in shifting the boundaries of the existing L1 categories. Essentially, since there are more speech sounds packed in the same space, learners might shift the boundaries of the L1 categories to make the differences between the sounds more distinct. In other words, the new exemplars of the L2 speech cause the restructuring of phonetic space around them. The L2 exemplars are placed together in a new common category, while existing categories shift away from the new exemplars and the new category. This might have the effect of L2 speakers exaggerating some features of a particular L1 speech sound in their pronunciation - to make them different from L2 sounds. This effect has been shown, for instance, by Italian-English bilinguals producing English vowels (Flege 2002).

Assimilation, the other process, occurs when the learner encounters an L2 sound that fits relatively well into a particular L1 category. To use the terms of exemplars and categories, L2 sound is treated as an exemplar of a given L1 category. This has two consequences for the language learners. First of all, they will treat the assimilated L2 sound as equivalent to a particular L1 sound and might not perceive the differences between the two. As a result, whenever they produce the L2 sound, they are likely to articulate it similarly to the equivalent L1 sound, which would contribute to their foreign accent in the second language. Second of all, since new exemplars are added to the L1 category, and the categories are shaped by the exemplars, the L1 category is likely to change and reflect both the L1 and L2 input. Depending on how much input in L1 and L2 the learners received, the merged category will represent the sound in one or the other sound to a greater extent. This, in principle, means, that learners might have a degree of foreign accent when articulating the L1 sound, in this sound is merged with the L2 category and if they have received so much L2 input that this category reflects the L2 sound to a significant degree. The assimilation effect has been shown, for instance, for the acquisition of voicing in stops (/p/, /t/, /k/, /b/, /d/, /g/). Usually the degree of voicelessness in voiceless stops (/p/, /t/, /k/) or voicing in voiced stops (/b/, /d/, /g/), measured in the voice onset time (VOT), differs across different languages. For instance, English voiceless stops are more voiceless (have a longer VOT) than their French equivalents. Studies have shown that in French-English bilinguals, the VOT for producing voiceless stops in both language is actually similar and somewhere in between the usual VOT observed in French monolinguals producing French voiceless stops and the VOT observed in English monolinguals producing English voiceless stops. This shows that in the bilinguals, the categories for voiceless stops in the two languages merged (Flege 1987; see also Flege 2002 for other examples).

The two processes proposed by Flege - assimilation of the L2 sound to a new category or creating a new category - are also features in Best's PAM. Both theories have similar assumptions, however while SLM focuses more on advanced L2 learners, PAM concentrates more on beginners (Best and Tyler 2008). In the 2008 paper, Best and Tyler create, however, a framework for phonological development in L2 that encompasses both processes in beginner and advanced learners, as well as the differences between the two. The crucial difference from SLM in this model is an emphasis on the distinction between the phonological and phonetic level in L2 learning and acknowledging that the listeners might perceive differences on the phonetic levels within a given sound categories. Perceiving such differences and learning to attune to these differences allows the L2 beginners to further develop or even restructure their newly created L2 categories.

The classic PAM (Best 1995) focuses on the likelihood that the L2 learner will perceive two L2 sounds as different depending on their similarity to L1 categories. This is a reasonable goal for a theory focusing on L2 beginners, for whom perceiving the distinction between L2 sound might be assumed as the most important aspect of phonological development. However, in the 2008 paper, Best and Tyler formulate more general rules pertaining to the creation of L2 categories. There are three different scenarios that can influence the category formation in language learners.

The first scenario occurs when there is only one L2 sound that can be classified by the language learner as an exemplar of the L1 category. When this sound is a good (prototypical) exemplar of the category, then there is a very small likelihood of the two sounds being discriminated and a new category for L2 being created. This is because developing such a category would be very effortful (the sounds are very similar to each other) and not particularly useful. Since there is just one L2 sounds that is assimilated to the L1 category, the learner will be perfectly able discriminate it from other L2 sounds.

The cost of entangling it from the L1 category does not bring any significant gains in terms of understanding the new language. When the L2 sound is not a very good exemplar for the L1 category, then this sound will be discriminated from L1 sound, but they will still be treated as equivalent. In other words, according to Best, no new category for the L2 sound will be created.

Another scenario would occur when there are two L2 sounds that are treated by learners as exemplars of one L1 category. There are two possible developments in this scenario. If one of the L2 sounds is a prototypical exemplar of the L1 category and the other one is a poor exemplar, then a new category will be created for the deviant sound. The good exemplar will be treated as both phonetically and phonologically equivalent to the L1 sound - i.e. listeners will not discriminate between the good exemplar and the L1 sound. If both sounds are actually good exemplars of one native categories - then they will not be discriminated properly, at least at the beginning of L2 acquisition. Both will be assimilated to the L1 category. The learners will neither perceive the contrast between the two L2 sounds, nor between the L2 sounds and the L1 sound. This is where, however, the difference between the beginners and more advanced learners can surface. More advanced learners can learn to see the within-category differences between the two L2 sounds and then start to create new categories for them. This will happen when the difference between the sound is really crucial, for example, when there are minimal pairs of the L2 sounds (two words differing in just the sound at hand) and when both of these words are frequent. Since the lack of distinction can cause communication problems, there is a motivation for learning the contrast. Moreover, high frequency minimal pairs offer input for learning the difference. For this reason, advanced learners might actually develop the two categories.

Finally, the third scenario occurs when the two L2 sounds are not assimilable to any L1 category. In this situation new L2 categories for these sounds will be created. However, depending on how similar the sounds are, learners might create either two categories for them or just one. In the latter case, again, the differences between the beginners and the advanced learners might surface. The advanced learners might learn to tune into the within-category differences to acquire the contrast between the two sounds.

As can be seen, while Best's predictions are more specific than Flege's, both theories are in fact very similar. They indicate that foreign language learners develop new, abstract categories (like children learning their first language), but only if the L2 sounds are sufficiently perceptually distinct form L1 sounds to warrant such categories. When the foreign sounds are similar to native ones, assimilation occurs. However, when creating a new category for the L2 sound benefits understanding, more advanced learners can start tune in their perception to the within-category differences and develop such a category. The within-category perception while contributing to abstraction, in fact is the part of specification process in phonological development. The next section will focus on that very process.

4.2.2.2. Specification

In L1 acquisition, specification has been defined as the development of exemplar representations of the learner. As a result of specification, listeners become more attuned to relevant phonetic details in a given language, which aids recognition of known words. In L2 speakers, specification is an important part of phonological development. In their 2008 paper, Best and Tyler compare naive L2 learners to more advanced L2 speakers and the majority of these differences might be subsumed under the specification processes in advanced L2 speakers. As learners are exposed to more exemplars of the foreign speech, they start to use different acoustic cues to process particular speech sounds. There are several studies on the non-native perception of vowels evidencing this. In one such study (Bohn and Flege 1990), native English speakers, experienced German learners of English (7.5 year of exposure to L2) and inexperienced German learners of English (0.6 year of exposure) were asked to categorise artificial vowels, which formed a continuum between the English DRESS vowel /e/ and a TRAP vowel /æ/ (which has no equivalent in German). The artificial vowels on the continuum differed in formant values (vowel quality) and in duration. The participants were asked to indicate whether a given vowel was /e/ or /æ/. The native speakers classified the artificial vowel on the basis of formant values, i.e. vowel quality. The inexperienced German speakers of English had problems with this task and used duration cues to classify the vowels. However, the experienced German speakers of English were attuned to the formant values and classified the vowels on the basis of that cue, similarly to native speakers. This finding is the evidence of specification, since it indicates attunement to language-specific cues for perception and identification of foreign language stimuli.

Another study shows similar findings for the English-Spanish bilingual speakers (Fox et al. 1995). In this study, English monolinguals, Spanish monolinguals, proficient Spanish speakers of English and non-proficient Spanish speakers of English assessed pairs of vowels (three Spanish vowels and seven English vowels, all vowels were paired with each other) and judged how similar the vowels were to each other. Then the researchers used a technique called Multidimensional Scaling to judge what cues (dimensions) did the participants use to assess the similarity of the vowel pairs. The results indicate that Spanish speakers differ significantly from the English speakers, in the number and types of cues used to process vowels. However, proficient Spanish speakers of English were more similar to English monolinguals in their perceptions than the non-proficient Spanish speakers. Spanish participants tended to cluster all the vowels (English and Spanish) into three categories - category of vowels similar to /i/, category similar to /e/ and category similar to /a/. English speakers organised the vowels around much more dimensions. The proficient Spanish speakers of English also clustered the vowels into three categories, but their categories were more dispersed, which made the proficient bilinguals more like English monolinguals. This study provides evidence for specification in L2 speakers, since it shows that adult foreign language learners start to perceive within-category differences between the speech sounds as a result of experience.

The studies on vowel perception like the ones presented, along with several studies on the effect of training on perception of non-native contrasts (see Zhang and Wang 2007 for review) provide evidence for specification in adult foreign language learners. Another evidence for the specification process is graded perceptual learning (Sebastián Gallés and Díaz 2012; Sebastian-Galles and Baus 2005; Díaz et al. 2012). Graded perceptual learning means that while some adult learners will perceive L2 contrasts in simple contexts, they will have more problems doing so in the noisier, more complex environment. The evidence for the phenomenon comes from studies involving assessing the same phonetic contrast in three kinds of perceptual tasks differing in complexity (Sebastian-Galles and Baus 2005; Díaz et al. 2012). In the study by Díaz (2012), which featured adult Dutch learners of English, the participants had to distinguish between the English /e/ and /æ/ vowel. The study began with a simple categorisation task, in which participants heard isolated synthetic vowels forming a continuum between /e/ and /æ/. The task was to match the artificial vowels to one or the other category (i.e. indicate whether the vowel is /e/ and /æ/). Bilinguals as a group did not perform on this task as well as monolinguals, however there was a significant portion of participants (about 44 per cent) who performed within the monolingual range. The next task involved lexical decision. The participants heard English words containing the /e/ or /æ/ vowel or non-words created by switching the vowels in question in an existing word (for instance, a non-word /dæsk/ was created by changing /e/ into /æ/ in the word "desk"). This task was more difficult, since it involved extraction of the sound from the complex stimuli, and dealing with coarticulation. Consequently, here only about 13 percent of Dutch learners performed as well as monolinguals. Finally, in the last task, the participants had to distinguish between English minimal pairs differing in one vowel, which was either /e/ or /æ/ (for instance, 'kettle' /kɛtl/ and 'cattle' /kætl/). During each trial, the participants were auditorily presented with a sentence containing a word from a minimal pair. The sentence was followed by the presentation of two pictures presenting both items from the minimal pairs (in this example - the picture of a kettle and of cattle). The tasks of the participants was to point to the picture representing the word they heard. This task is even more complex than lexical decision, since it involves extraction of the sound from a longer sentence that taxes working memory to a more significant degree. Also, the word embedded in the naturally spoken sentence might be more distorted. As a result, here only 9 per cent of the participants performed within a native range. Overall, the study indicates that the more complex the context for a phonetic contrast, the more problems L2 learners have processing it. A similar pattern of results was observed in a study where participants were Spanish speakers of Catalan (Sebastian-Galles and Baus 2005). These studies provides strong evidence for specification - the speakers, at least those, who performed well on the categorisation task, have established the phonological categories in the foreign language. However, not all of them had sufficiently developed exemplar representation that would facilitate word recognition and thus help them perform the discrimination tasks in more complex and noisy settings. In short, the difference between the more successful learners and their less successful peers was in the degree of L2 specification.

The extent of specification in foreign language learners will depend on several factors. One set of factors is connected with environment and exposure. Immersion into the L2 community generally improves learners' perception of L2-relevant speech cues (Jia et al. 2006). In general, upon arriving into a foreign country, the adult foreign language learners experience quick development in perception, but with time progress becomes stalled (Aoyama et al. 2004, 2008; Best and Tyler 2008). According to Best and Tyler (2008), the greatest progress occurs within five to twelve months of immersion of an individual in a foreign language setting, then is stalled. Another set of factors influencing specification to a large extent is related to the features of the speaker's L1. This has been touched upon in the previous section on abstraction - it is difficult to be more attuned to L2-specific phonetic cues in a speech sound that has a good equivalent in L1. In other words, L1 representations can affect the specification in L2. Similarly, L1 phonotactic, allophonic and coarticulation patterns will influence this process (Best and Tyler 2008). Finally, there are several personal factors that influence the degree of specification that takes place in adult foreign language learners. In general, in the field of foreign language acquisition, factors such as age, personality, attitude towards the language and motivation have been named as the predictors of achievement in language learning (Bialystok and Fröhlich 1978; Gardner and Lambert 1972). A review article by Piske, devoted specifically to establishing the predictors in the acquisition of L2 phonology (operationalised as a degree of foreign accent), concludes that of all the personal factors, age of L2 acquisition and the amount of continued L1 use are the most influential in L2 phonological development (Piske et al. 2001).

Apart from the above-mentioned set of predictors, there also seems to be one particular factor that influences the degree of specification in foreign language learners - the perception of speech-related acoustic cues. Some ERP studies (Díaz et al. 2008; Jakoby et al. 2011) have found this factor to be related to L2 achievement in general and the development of L2 speech perception specifically. In one of these studies, Díaz and colleagues (2008) examined Spanish learners of Catalan who have been divided into good L2 learners and poor L2 learners, depending on their scores in the speech perception task. The good L2 learners were those who scored within the native range on three L2 vowel discrimination tasks of varied complexity, while the poor learners were those who scored below the native range on all there tasks. The bilinguals in both groups did not differ significantly in terms of L2 use, age of acquisition or exposure. The ERP study showed, however, a significant difference between them. When the participants were presented with a vowel contrast in both their native language and a completely foreign language, the brains of the good L2 learners were much better at perceiving the contrast. In one of the ERP tasks, the participants were presented with several repetitions of the Spanish vowel /o/, which were intermixed with the occasional iterations of the Spanish /e/. During the presentation of the /e/ vowel, the ERP signal, or more specifically the MMN (Mismatch negativity) component, which is connected with detection change in auditory stimuli, was recorded. In another task, the iterations of the /o/ vowel was intermixed with production of a Finnish /ö/ vowel and again, the MMN to the incongruous vowel was measured. In both tasks, the good L2 learners displayed a much greater MMN reaction to the odd stimuli than the poor L2 learners. When similar tasks were conducted with non-speech stimuli, the MMN reaction to odd sound was not different in both groups: the good learners were no better than poor learners at perceiving changes in duration, frequency or sound patterns in pure tones. This means that there is a particular language-universal perceptual ability, which predicts an individual's phonological development (specification) in a new language. This ability cannot be pinned down to sensitivity to particular acoustic cues, but it facilitates detection of acoustic cues relevant for speech processing. Similar findings have also been found in the study on late, less-skilled learners (Díaz, Mitterer, Broersma, & Sebastián-Gallés, in preparation, as quoted in Sebastián-Gallés and Díaz 2012). Yet another study (Jakoby et al. 2011), indicates that this perceptual ability is connected with general level of attainment in the foreign language learning and not merely with phonological development in this language.

To conclude, the specification process, defined as a development in exemplar speech representations takes place not only in the native speech learning, but also in L2 learning. The evidence for this process comes from studies showing that the perception of speech sounds in an L2 changes in learners after some period of exposure to the language. The learners start to use new perceptual cues to process foreign sounds and their perception of within-category differences is developed. The degree of specification in L2 learners differs. Some learners will learn to perceive new contrasts and recognise the L2 speech sounds even in more complex and noisy environment. Other might develop within category perception but will recognise the new sounds only in very simple contexts. Yet other learners might never develop their exemplar representations sufficiently and will not recognise the new speech sounds even in very simple environments. The degree of specification in adult learners depends on several factors: the environment and amount of exposure to the new language, the characteristics of the L1 and the personal factors, such as age or L1 use. There is also one specific factor predicting the degree of specification, which is connected with the universal ability to perceive speech-related acoustic cues. According to some studies, this factor predicts also general level of attainment in the new language. The next section will deal with the last type of phonological development, that is chunking, and present how this kind of development takes place in foreign language learners.

4.2.2.3. Chunking

In the section devoted to native language acquisition, chunking was defined as a process of creating abstracted representations of common phonological sequences. As such it was one of the crucial processes of phonological development in children, allowing them to learn verbal material more efficiently. However, chunking is not a mechanism limited to first language acquisition or even to phonological sequence learning. Chunking, understood as sequence learning or combining elements into larger structures is a basic learning mechanism in Newell's unified theory of cognition (Newell 1994). It seems that human beings in general learn things by organising them into larger units. This is a strategy that helps them use their memory resources more efficiently. As mentioned in chapter 1, short-term memory is limited in capacity and can contain only a small number of items. As a consequence, only asmall number of items is carried into the long-term memory unless it is packed more densely into larger chunks. And thus, chunking has been named by several researchers a basic L2 learning strategy (Ellis 1996, 2008; MacWhinney 2004). To learn L2 is to learn sequences occurring in L2. This is also true for L2 phonological sequences.

In chapter 2 it was argued that chunking in L1 learners is a process that follows abstraction. What children "chunked" was abstract phonological representations, as evidenced by the fact that the effects of chunking could be observed in the later course of acquisition (between the age of 3 and 5). While there is evidence that children learn about transitional probabilities as early as in the 8^{th} month of age, this ability was not considered chunking in the previous section of the thesis, because it was not assumed that children had abstract phonological representation at the age of 8 months. However, in the second language acquisition literature, the learner's ability to extract phonological patterns such as transitional probabilities from foreign speech *is* referred to in literature as "chunking" or "sequence learning" (Speciale et al. 2004; Ellis 2008). It might seem illogical to classify the same process (extraction of transitional probabilities) differently in L1 and L2. However, it has been indicated in the previous sections that L2 learners assimilate L2 sounds to their native phonological categories very early on. That means that from the earliest moment of L2 acquisition, they will probably operate on certain kind of abstract phonological representation and therefore the pattern extraction that occurs in L2 learners might be called "chunking". If one takes this perspective, one must conclude that chunking in L2 occurs very early in the course of acquisition. The transitional probabilities experiments (Saffran et al. 1996a) indicate that people gather knowledge about phonological sequences in a particular language in the matter of minutes from the beginning of exposure. Further evidence for this hypothesis comes also from word association studies (Ellis 1996). In these studies, participants are presented with a word and asked to provide another word that comes to their head in response. These studies show that beginner L2 speakers often confuse the foreign words with different foreign words having similar phonological structure (for instance L2 learners of French will confuse the word *béton* with *bâton* or *breton*) (Meara 1984). These findings indicate not only that learners internalise the knowledge about phonological sequences in L2 very early on, but that they organise their L2 lexicon around this knowledge.

However, since L2 learners already have a wide repertoire of native chunks in their phonological inventories, the development of L2 chunks will be widely influenced by the phonotactic patterns in their L1. This is because the perception of speech in older learners is guided by the phonotactics of their L1. Speakers have problems perceiving and learning sequences that are not possible in their L1. This is what has been shown in the already discussed article by Finn and Hudson-Kam (2008). In this study, adults who were exposed to artificial language violating the phonotactic rules of their L1 did not learn the phonological sequences of this language correctly. The effect of L1 phonotactics on perception of non-native sequences has been reported in many other experiments as well. For instance, when native speakers of Japanese listen to a speech sequence that contains consonant clusters which are illegal in Japanese, they are likely to perceive phantom /u/ vowels between the consonants of those clusters (Dupoux et al. 1999, 2001). It is also typical for learners to insert vowels in between the phonological sequences that are not permitted in their native language when producing words of a foreign language (Hancin-Bhatt and Bhatt 1997; Davidson 2006). Although, as pointed out by Davidson (2006), some of these productions might stem from inability to articulate the structure, some of these errors stem from inaccurate perceptions. All in all, it seems that L2 chunking is significantly affected by L1 chunks already stored in learners' minds.

To conclude, the data presented in this section suggest that the chunking process takes place during the L2 phonological development from the very early stages of acquisition, however it is affected by an interference from the L1 phonological system, which shapes the perception of L2 speech. For these reasons it is probable that L2 speakers' repertoire of chunks in a given language will differ from that of native speakers of this language. It is also possible that the L2 chunks of the learners will develop as a result of language exposure. First of all, since the phonological categories of the L2 speakers are likely to be reorganised as a result of exposure to the foreign language, so will the structure of their chunks change. Moreover, since speakers are likely to encounter more L2 material with time, their inventory of the most frequent L2 chunks is likely to be refined by experience.

This previous sections provided an overview of phonological development in L2 on the basis of the theoretical framework presented in this thesis. In particular, this chapter followed on the proposition put forward in chapter 3, that phonological development consists of several processes, including initial prosodic processing, abstraction, specification and chunking. It therefore attempted to show how these processes are reflected in the acquisition of L2 phonology. It transpired that they are reflected very well. If fact, barring the interference from L1 phonology, phonological development in L2 is remarkably similar to the development of phonology in children acquiring their L2. This observation has very interesting consequences. In chapter 3 it was noted that all the processes involved in L1 phonological development are also important for the lexical development of the L1 learners. If the same processes - initial prosodic processing, abstraction, specification and chunking are also present in L2 acquisition it is logical to assume that they would be of importance for the lexical development in L2. The next section will explore this hypothesis in more detail.

4.3. Phonological acquisition and foreign word learning

Chapter 3 described how phonological acquisition helps word learning in the L1. In infants, the initial prosodic processing helps extract words from the stream of speech and thus triggers the word learning process. Specification then helps with the recognition of the known words in the stream of speech. Abstraction (creating abstract phonological categories) allows for processing the phonological structure of novel words and thus learning them more efficiently. Finally, chunking facilitates word learning by freeing the learners' cognitive resources. Overall, both the theory and empirical research point to the relationship between phonological development and word learning in native language acquisition.

In second language acquisition, however, there is a marked lack of empirical data that would show a similar relationship. There is only a handful of studies that investigate the correlation between phonological processing and foreign language learning. A 2008 review article by Walley (2008) devoted to the Lexical Restructuring Model in L2 learners quotes only one study that explores the relationship between phonology and lexicon in L2 acquisition (Imai et al. 2005). This in itself indicates, how severely under-researched the area is. However, seeing that L2 phonological development is in many respects similar to the development in native language, it is reasonable to assume that it should also have similar effects on vocabulary learning. As has been demonstrated so far, in L2 learning, just like in the acquisition of native language, one can distinguish the initial prosodic processing, specification, abstraction and chunking. All four of these processes should have effects on the lexical development in the foreign language.

To date the most convincing evidence for involvement of phonological processes in L2 word learning has been provided for chunking. Chunking in L1 acquisition is a process that allows for more efficient learning of novel words. This is because, as indicated by Jones (2010), packing verbal material into bigger chunks allows for storing more of it in the phonological STM. As a result of more efficient storage of new words in phonological STM, language users learn new words more quickly. It is probable that this rule obtains not only for children, but also for adult learners, and that just as chunking facilitates L1 vocabulary learning, it should also facilitate L2 vocabulary learning. And this is indeed what has been found in two experiments that investigated the effects of novel sequence learning on foreign word acquisition (Speciale et al. 2004). In one of the experiments, 38 English undergraduate students were given a task measuring their sequence learning ability, and a word learning test. In the sequence learning task, the participants were presented with 12 non-words and 12 foils. Then on subsequent trials they were presented with either the 12 old non-words or completely new foils and they had to indicate whether they have heard a particular non-word or not. Overall, the 12 non-words were presented in the task 8 times and there were 96 additional non-words that were foils and that appeared only once in the task. The experimenters measured how quickly the participants could remember the 12 non-words and used is as an indicator of how quickly the participants could learn new sound sequences. In the word learning test, the participants had to learn 24 German words over the course of two learning task (none of students knew German prior to the experiment). Each task comprised of a presentation phase, in which the participants saw an English word on the screen and heard the German equivalent of the word produced twice. 12 items were presented that way, then the procedure was repeated, so every German word was presented to the participants four times. Following the presentation phase was the reception phase, in which the students heard all the German words one by one and had to type their English equivalents into the dialogue box. Finally, there was a productive phase, in which the participants saw the English word on the screen and had to say a German equivalent of the word to the microphone. Once all the 12 items were tested in that way, the participants, were given another task with other 12 German words. Apart from the phonological sequence learning task and the two German learning task, the participants also did a non-word repetition task with items that were low on word-likeness, i.e. did not tap considerably into the knowledge of phonological sequences of the speakers. According to the authors, this ensured that the task was a purer measure of phonological STM. The results of the study indicate that the phonological sequence learning ability correlated moderately with both the receptive and productive German vocabulary scores, even when the non-word repetition scores were controlled for.

The other experiment reported in the paper replicated these findings in a more naturalistic setting. Here experimenters tracked he progress of 44 students during a Spanish course at novice level organised for the purpose of the study. The participants attended Spanish classes twice a week for 10 weeks. At the beginning of the course they were tested on phonological sequence learning, non-word repetition and Spanish vocabulary. 5 weeks after the course has ended, the students sat an exam in Spanish, which consisted of several comprehension tasks, as well as receptive vocabulary tasks. The phonological sequence learning task score correlated moderately with the receptive vocabulary scores obtained during the final exam, even when the initial Spanish vocabulary scores and non-word repetition scores were factored out. In other words, students who were better at sequence learning learned Spanish words during the course more efficiently than their peers with lower sequence learning ability. To conclude, both studies indicate that chunking plays a role in L2 vocabulary acquisition. Additional data confirming the hypothesis come from the study by Mirman and colleagues (2008), already mentioned in the section 4.2.1. In this study, participants were exposed to a stream of speech with novel words and then asked to perform a word-learning paired associates task, with either the words that occurred within the stream of speech, or "part-words" that contained sound sequences typical for word boundaries in the exposure stream speech. The participants were faster to learn the words from the stream of speech than the "part-words", which indicates that learning phonological sequences typical for a language should facilitate learning vocabulary of this language. All in all, the empirical research seems to confirm the hypothesis that chunking, or phonological sequence learning, aids new word learning.

When it comes to the relationship between the other processes involved in phonological development and L2 word learning, however, there is considerable lack of research. Most of the empirical evidence for such a relationship is very indirect. For instance the only study somewhat related to the issue of initial prosodic processing and L2 word learning seems to be the already mentioned research by Shoemaker and Rast (2012). The initial prosodic processing helps initialise word learning in infants. Children use prosodic cues (combined with transition probabilities) to extract the word forms that will become the basis of their first lexical representations. One can assume that in L2 learners, the initial prosodic processing along with the extraction of transition probabilities is also a necessary trigger for word learning. However, Shoemaker and Rast's data gives only very indirect evidence that this might be the case by showing that learners might use prosodic cues for recognising words, which is the first step to learning them. So far there appears to be no studies that would show the link between initial prosodic processing of speech stream and word learning directly.

When it comes to specification, there is slightly more evidence for the process's relationship with L2 lexical development. The specification aspect of phonological development is connected to better recognition of known words in children acquiring their first language. L2 specification should therefore facilitate better word recognition in older L2 learners. Indirect evidence for the link comes from the study by Imai, Walley and Metsala (2005), in which monolingual English speakers, proficient Spanish speakers of English and non-proficient Spanish speakers of English were asked to perform a word recognition tasks with high and low neighbourhood density items (i.e. words with many phonological neighbours and few phonological neighbours). The task involved, listening to the English words recorded by a native speaker and writing them down[2]. The results indicate that the proficient Spanish speakers were as good at recognising the words in sparse neighbourhood as native speakers, while the low proficiency speakers performed poorly in both conditions. According to Walley (2008), the advantage of the more proficient speakers comes from the fact that their phonological system is more developed, however there is not much in the study itself to support this claim. A direct evidence for this hypothesis, however, can be found in another study by Meador, Flege and Mackay (2000), examining 18 English monolinguals and 72 Italian speakers of English (divided into four groups on the basis of L2 proficiency and length of residence in the English-speaking countries). The participants were presented with 10 sentences in noise and asked to repeat as many words as they could. Additionally, they were given two phonological perception tasks. In one of the tasks they had to recognise English consonants placed at the beginning or end of words (both in noise and in the clear), in the other they had to assess the contrast between different English vowels. The results indicate that the ability to recognise English speech sounds contributed significantly to the word identification scores, even when other predictors such as length of residence in an English-speaking country have been factored out. These findings provide convincing evidence that phonological specification in L2 indeed is connected with improved lexical processing and helps speakers recognise words in the stream of speech.

Unfortunately, it is more difficult to find empirical support for the relationship between the abstraction aspect of phonological development, and the learning of new vocabulary items. In L1 acquisition the creation of phonological categories is connected with faster and more efficient word learning (de Jong et al. 2000). It seems therefore

[2] In the task, the participants had actually to recognise English word as spoken by a native speaker and Spanish-accented speaker, but the findings related to the Spanish-accented words do not seem to bear much relevance to the topic at hand.

that creation of new categories in L2 (or at least successful assimilation to L1 categories) should facilitate learning new words in the foreign language. However, it is difficult to find any empirical confirmation of this hypothesis. There is one study that gives very indirect evidence for this claim (Ganschow and Sparks 1995) and three studies by Hu, indicating the relationship between the scores on phonological awareness in L1 and the speed of word learning in L2 (Hu 2003, 2008; Hu and Schuele 2005).

The study by Ganschow and Sparks (1995) examined the effects of explicit phonological instruction on the language aptitude in a group of teenage girls (aged 14-16). 14 participants had foreign language learning difficulties and 19 had no problems with language learning. The 19 girls without learning difficulties took part in a normal language course that took a year. The girls with foreign language difficulty took a one-year-long course based on multisensory structured language methodology, which is largely focued on improving learners' awareness of the phonological structure in L2. The teaching method involves slow introduction of new material, explicitly pointing to the structure of the verbal material, phonetic drills, and "multisensory practice (simultaneous saying, seeing, hearing and writing) of the sounds and symbols of the language" (Ganschow and Sparks 1995: 110). Both groups took a pre-test (before the course) and post-test (at the end of the course) composed of a battery of phonology and orthography measures, as well as a the five subtests of the Modern Language Aptitude Test (MLAT – Carroll and Sapon 1959). One of the tasks performed as a part of MLAT was a paired associates word learning task, in which the participants had to learn new words of an invented language. The experimental group displayed greater improvements than the controls on the tested abilities. The course improved their performance on some of the phonological tests and on most MLAT subtests. What is important is that there was a great improvement on the paired-associates task, which tested the foreign word learning ability. This study indicates that greater phonological awareness indeed helps word learning. However, the number of possible confounds in the study (no random assignment to the groups, the possibility of Hawthorne effect, treatment introducing many techniques and variables) does not allow to fully accept this conclusion.

The three studies by Hu focused specifically on the relationship between phonological abilities, phonological and L2 word learning and thus might provide some clues as to the role of abstraction in word learning. In one of the studies (Hu 2003), 48 Mandarin-speaking children learning English at school on a daily basis were tested in four sessions across the span of two years. The participants were aged 4-5 at the time of the first test. Among the tasks performed during the four testing sessions was a phonological STM task (the children had to repeat sets of 3 bisyllabic non-words) and three phonological tasks. In one of these tasks, vowel substitution, the children had to replace the vowel /a/ with /u/ in the words they heard. The second task, syllable substitution, involved replacing a syllable /feng/ with a syllable /dou/ in presented words. In the third task, vowel identification, children had to indicate whether a monosyllabic word contained a vowel /a/. In session 3 and 4 of the experiment, the participants were given an additional word learning task. They were shown three pictures depicting conger, minnow and triton, and asked to repeat these words. Then they were shown the pictures in random order and had to name them with the new words. The maximum number of trials was nine. The results of the study indicate significant correlations between almost all the phonological tasks scores, and the word learning scores in session 4. There was also a relationship between phonological STM and word learning at both session 3 and 4. To investigate the unique effect of the phonological abilities on word learning efficiency, the author of the study conducted a hierarchical regression analysis. This analysis revealed that even when researchers controlled for the effects of phonological STM, there was a strong effect of phonological awareness on word learning scores during the fourth session. This result provide evidence for the relationship between phonological skills and word learning in L2. While the participants in this study were relatively young, Hu managed to replicate his results also for older children, whose L1 was presumably already mature and established.

Both replications (Hu and Schuele 2005; Hu 2008) were part of a longitudinal project investigating the effects of phonological awareness on lexical acquisition in a group of students. The participants were 74 third-graders (age range 8;2 - 9;3), native speakers of Mandarin Chinese learning English at school, who were classified into either high phonological ability group (N=37) or low phonological ability group (N=37) based on the performance on three phonological tasks. In one task the participants had to spell wordlike stimuli in *zhuyin fuhao*, which is phonetic writing system used for transcription of Chinese. Another task involved deletion of initial consonants in native words. In the third task, participants had to point out the word in a group of three which did not begin with the same phoneme (onset oddity), or which did not rhyme with the rest (rhyme oddity). The groups did not differ in terms of native vocabulary scores or digit span, although they did differ on measures of English vocabulary and phonological STM as measured with a Chinese non-word repetition task. The phonological tests were administered to the participants in the third grade.

A year later (in grade 4) the participants were asked to perform four learning tasks: familiar name learning, native name learning, nonnative name learning and visual learning (Hu and Schuele 2005). In the familiar name learning, children had to memorise the association between three pictures and three words they knew (Chinese names for fruits) within 10 trials. In the native name learning, they learned the association between three pictures and three Chinese novel names created by using common Chinese syllables. In the nonnative name learning, the participant had to learn the associations between three pictures and three English non-words created by interchanging syllables across three common names (i.e. from the names Patsy, Roger and Ellis, experimenters created non-words such as Rotsy, Elger and Palis). Finally, in the visual learning task, participants had to learn the associations between three pictures and three other pictures. The results indicate that the high phonological ability group was much better than the low ability group at the native and non-native word learning, but not at the familiar name and visual learning. When the effects of phonological STM and English vocabulary size were controlled for in both groups, the phonological abilities were significant predictors of the non-native learning scores. This result indicates that phonological skills are of importance precisely for learning novel, nonnative words - that means words of another language.

This finding was replicated in the examination of the same students two years later (in the fifth grade) (Hu 2008). Here the students (30 participants in the low phonological ability group and 35 in the high phonological ability group) were given a rapid naming task and an English word learning task. The rapid naming task tested the speed of lexical retrieval. It involved naming in English four colours presented to the students (yellow, blue, green, red) as quickly as possible. Both speed and accuracy were assessed in this task. In the word learning task the participants were presented with 8 English colour terms in a carrier phrase "This is a X hat" along with appropriate referents (pictures of different-coloured hats). Four of the learned items were familiar colour terms (yellow, blue, green, red), the other four were unfamiliar (rust, ecru, sage and spruce). Following the presentation, there was a production phase in which the participants had to name all the eight colours, then a recognition phase, in which the participants listened to the experimenter producing the name and had to point to the appropriate colours. The cycle of presentation, production and recognition was repeated eight times. The results indicate that the children with lower phonological skills scored lower on the rapid naming task (both in terms of speed and accuracy of production) and on the word learning task. In the learning task there was a significant difference between the groups when it comes to the learning of the new colour terms, but not the known colour terms. Moreover, the differences are visible only for the production trials, not the recognition trials. In other words, children who have better phonological skills in the first language, have also better lexical retrieval for the second language and are better at actively learning the new word forms in another language. Overall, the studies presented clearly indicate that phonological factors play a role in the acquisition of novel words. The author of the research concludes that poorer phonological processing skills make it more difficult for the speakers to create the phonological representations of the new words and, as a result, memorise them.

While Hu's research provides compelling evidence for the relationships between phonological processing and word learning it invites two questions. One is the question about the actual type of phonological processing that facilitates the vocabulary acquisition. The other, related question, is why would L1 phonological task scores be related to L2 word learning. There are at least four possible answers to those questions. The first one is that the kind of processing tapped by Hu's L1 phonological tasks is the initial prosodic processing that allows learner to find unit boundaries in the stream of speech. This is the possibility suggested by Hu himself in his 2003 paper. The issue with this interpretation is that initial prosodic processing should operate mainly at the initial stages of acquisition and thus it should not be utilised to a great extent in L1 phonological tasks by children with developed L1 phonological systems.

The second possibility is that Hu tapped into some language-universal sensitivity to speech-specific acoustic cues that can facilitate the L2 specification process. The existence of such sensitivity and its connection to word learning is indicated by the results obtained by Jakoby and colleagues (2011). Their research investigated phonological skills of a group of Hebrew learners of English characterised by high English vocabulary scores and good comprehension of English, as well as a group of learners with lower vocabulary and comprehension scores. Both groups performed a task tapping into sensitivity to speech-specific acoustic cues. They were presented with a foreign

(French) vowel contrast: they heard several instances of a vowel /u/ with occasional instances of the French vowel /y/ and their EEG response (MMN) to the /y/ vowel was recorded. The participants with higher vocabulary and comprehension scores had a much greater reaction to the vowel contrast. This indicates a relationship between speech-specific auditory sensitivity and the efficiency of vocabulary learning. The problem with fitting this interpretation to Hu's results is that the phonological tasks he used in his studies did not measure sensitivity to minute auditory differences, but rather the knowledge and understanding of L1 phonological system. Hu's tasks did not involve processing of foreign speech sounds, recording brain response to speech stimuli, nor recognition of speech sounds in noise. These tasks tapped into the abstraction rather than specification aspect of phonological development.

The third possible interpretation of Hu's results is that the participants who were better at L1 phonological tasks, had certain ability that allowed them to create abstract phonological categories easily. This ability helped both L1 and L2 abstraction of the speakers. The L2 abstraction in turn allowed for a better L2 word learning. This interpretation seems very likely. The problem is, however, that since the studies did not contain any L2 phonological tasks it cannot be confirmed by Hu's data. A variation of this explanation is that students who did better on L1 phonological tasks had more established phonological categories in L1 and as a result were better at assimilating L2 sounds to these categories. This interpretation has been suggested by Hu in his 2008 paper.

The fourth interpretation of Hu's results is that the L1 task tapped into chunking abilities which facilitated L2 word learning efficiency. The chunking hypothesis could in fact be responsible for part of the effect. However, this process would be tapped mostly by the non-word repetition task. As mentioned previously, chunking works mainly by improving phonological memory. Hu used in his studies phonological memory task involving remembering words composed of typical Chinese speech sequences, which suggests that this task would be loaded very strongly by chunking. Therefore, if chunking was the main mechanism within the phonological tasks that facilitates word learning, the effect of phonological scores on word learning should not be significant in hierarchical regressions after the non-word repetition scores are factored out.

All things considered, the most probable explanations to the phonological processing effect observed in Hu's research is that his phonological tasks measured the ability to create phonological categories, which influence both L1 and L2 abstraction, or that they reflected assimilation to L1 phonological categories. The problem is that without data on L2 phonological abilities, none of these hypotheses can be confirmed. The other problem is that without data on L2 phonological development, it is not really possible to determine how phonological and lexical acquisition are intertwined in L2. This is why there is a need of new research, which would examine the phonological development in both languages and investigate how it is related to lexical acquisition in L2.

4.4. Conclusion: integrating memory, phonology and word learning in L2 acquisition - the current study

The previous chapters of this thesis have indicated that there should be an intimate relationship between phonology and lexicon in both L1 and L2 acquisition. The first chapter focused on a popular theory that these two aspects of language processing are connected via a memory module called phonological STM. Phonological STM has been defined as a short-term store for verbal information, which utilised phonological coding and helps with learning novel word forms. However, at the end of the chapter it has been suggested that there might be a more direct relationship between phonological processing and lexicon. After proposing a working definition of phonological processing in chapter 2, it has been shown that indeed phonology and lexical development seem to be intimately linked in L1 acquisition (chapter 3). On the basis of research finding from L1 development, it has been proposed that also in L2 there should be a relationship between phonological processing and vocabulary acquisition.

On the basis of the proposed theoretical framework and past research, one can make specific hypothesis about the effects of different types of phonological developmental processes on vocabulary learning. The initial prosodic processing should initialise word learning in the L2, thus should be of particular importance for the lexical acquisition in beginner learners. Abstraction should facilitate acquisition of L2 lexical items in more advanced learners. Specification should aid word recognition, but, as indicated by Jakoby et al. (2011) it is also important for word acquisition, perhaps due to the fact that it is tightly connected to abstraction. Chunking should facilitate word learning by helping learners use their phonological STM more efficiently. The role of specification and chunking in lexical acquisition has been supported by empirical research. The role of other kinds of phonological developmental processes have, however, been left largely unexplored in research. The aim of the study presented is to bridge this gap and explore how phonological skills in both the first and the second language influence vocabulary learning. The study presented is probably the first one to test for phonological development in both the first and the second language and correlate the test scores in both languages with word learning tasks.

The participants of the study were Polish 9-year-olds who learned English at school. They were given batteries of phonological tasks in both their L1 - Polish and in their L2 - English. The English battery was used to tap into their L2 phonological development. The Polish tasks were included as a way to extract phonological factors that were important for phonological performance in both languages. In this way the study aimed to tap into the language universal factors that might underlie phonological development (such as general abstraction skills). Additionally, the participants were asked to perform several phonological STM tasks. These tasks involved repeating words and non-words that were composed of both frequent and unusual speech sound combinations. These were used to tap into the phonological STM as well as possible chunking mechanisms. The phonological tasks scores and the phonological STM scores were used as predictor variables in the study. The outcome variable in this research was the efficiency of novel word learning. This was measured by testing the English vocabulary size of participants at the beginning and at the end of the school year and by means of experimental novel word learning tasks. Each child was asked to perform four such tasks. In one they had to learn three pairs of Polish words (L1 words), in the second task they were asked to learn three non-words that were similar in structure and accent to Polish words (L1 non-words), in the third one they learned three non-words resembling in English words structure and accent (L2 non-words) and in the fourth task they learned three non-words that had unusual structure and was pronounced with a foreign accent (LX non-words). These tasks were used to investigate the effect of phonological processing on learning words of a language at different stages of acquisition. The task with L1 non-words tapped into the ability to learn words of a known language, in which the participants had developed speech representations. The task with L2 non-words tapped into lexical acquisition of the language that the participants did not know very well. Finally, the task with LX non-words tapped into the ability to learn a completely new language, in which the participants had no phonological representations.

Chapter 5: Words, sounds and memory. The study on the relationship between phonological factors, phonological STM and foreign vocabulary learning in Polish school children

5.1. Introduction

The four previous chapters presented the theoretical basis for the current study by introducing the literature on the relationships between phonological memory and word learning, as well as the relationship between the phonological development and word learning. The notion of phonological development in L1 and L2 was examined and divided into four separate processes (initial prosodic processing, abstraction, specification, chunking). The literature review demonstrated that the topic of relationship between L2 phonological and lexical acquisition is severely underreserached, thus providing a justification for the study at hand. This chapter will present the experimental research investigating the entwining of L2 phonological development and foreign word learning. Firstly, the aims of the study will be briefly reminded, then the methodology of the study will be presented, along with the methods of statistical analyses. The chapter will end with the results of the study.

5.2. Aims of the study

The aim of the current study is to investigate the relationship between phonological factors (phonological processing and phonological STM) and foreign word learning in an orderly, systematic way. Due to lack of previous research on the effects of L2 phonological acquisition and word learning, the study has a largely exploratory character. The investigations are organized around four research questions. For each of those questions, general and specific hypothesis have been put forward on the basis of the theoretical framework presented in the previous chapters.

Research Question 1: Does phonological development in L2 influence word learning in this language (both in experimental tasks and in natural language learning environment)? If so, in what ways? Which L2 phonological processes facilitate word learning?

As indicated in chapter 3, phonological development influences lexical development in L1 (de Jong et al. 2000; Storkel 2001; Windfuhr and Snowling 2001; Ramachandra et al. 2011). Since, as argued in chapter 3, L2 acquisition is in many respects similar to L1 acquisition, it is hypothesized that phonological and lexical development should go hand in hand also in L2. **Thus hypothesis 1A states that phonological development in L2 should influence L2 in word learning, both in experimental tasks and in natural learning environment.** More specifically, it is expected that the performance on phonological tasks in L2 should correlate with the increase in L2 vocabulary in study participants. It should, moreover, correlate with the results of experimental word learning tasks in L2. The above statement constitutes the main hypothesis of this study.

In chapter 3 and 4 it has been proposed that phonological development in both children acquiring their L1 and in L2 learners could be divided into four processes: initial prosodic processing, specification, abstraction and chunking. It has been noted that each of those four processes should be distinctly linked to word learning. Firstly, initial prosodic processing of the foreign speech stream should jumpstart the word learning process. Secondly, creating abstract phonological representations (and possibly specifying them) in L2 should facilitate encoding new L2 word forms. Thirdly, chunking process should make it easier to store new phonological forms of the words in phonological STM and thus make word learning more efficient. **Thus hypothesis 1B states that there are different phonological factors influencing word learning**. This should be reflected in the study by correlations between different kinds of phonological tasks and word learning. For instance, since blending tasks and ISR tasks with Polish non-words should tap into chunking and chunking should make word learning more efficient, there should be a correlation between those tasks and novel word learning scores. Elision tasks primarily tap into segmentation, which might stem either from initial prosodic processing (in less advanced speakers) or abstraction and specification. Thus the correlation between elision tasks and word learning scores should point to the involvement of these factors in lexical acquisition.

Research Question 2: Are there language-universal factors underlying L1 and L2 phonological development and do these factors facilitate novel word learning, as suggested by Hu and colleagues (Hu 2003, 2008; Hu and Schuele 2005)? If so, what are they?

One clear conclusion that should be drawn from the studies by Hu is that there is a common factor underlying phonological development in any language. **Thus hypothesis 2 states that there will be common variance (factor) in phonological tasks and that this common variance will influence novel word learning**. To test this hypothesis, an exploratory factor analysis will be conducted on all the phonological tasks. Should there be a factor that is loaded by phonological tasks in both L1 and L2 and should this factor predict novel word learning, the hypothesis will be confirmed. Depending on the specific loadings, the nature of this factor can be further specified. For instance, if this factor is loaded mainly by blending tasks, it could be identified as chunking. Should this factor be loaded by elision tasks, it can be identified as certain ability connected with segmentation. This can be either the ability to create abstract phonological categories or initial prosodic processing. Should there be greater involvement of L1 phonological tasks in the factor – the ability to abstract would be the more probable explanation. Should there be greater involvement of the L2 phonological tasks – it can be hypothesised that this factor is initial prosodic processing.

Research Question 3: Do phonological processing and phonological STM interact to facilitate word learning?

Previous research suggests that both phonological processing and phonological STM might provide contributions to word learning. It is also possible that there is some interplay between them (Bowey 2001). Specifically, it is reasonable to assume that chunking (which is part of phonological development) facilitates vocabulary acquisition by making the use of phonological STM more efficient in word learning tasks. **Thus hypothesis 3 states that both phonological processing and phonological STM might**

contribute to novel word learning. This hypothesis will be confirmed by results showing the correlations between phonological STM and word learning, as well as the correlations between phonological tasks scores and word learning. There might also by correlations between phonological task scores and phonological STM scores, which would imply an interaction between the two factors.

Research Question 4: Do phonological factors influence word learning differentially for different languages, i.e. will the influence on native word learning be different than on foreign word learning?

In chapter 3 and 4 it was suggested that phonological development consists of four processes and that each of those processes plays a specific role in lexical acquisition. It has been suggested that these roles might be dependent on the stage of language acquisition. For instance, initial prosodic processing – the process responsible for initial segmentation of speech – occurs mostly at the first stages of acquisition in small children or when the learner is faced with an unusually noisy signal (Mattys et al. 1999, Mattys et al. 2005). It is a factor triggering word learning. In L1, processes such as abstraction and specification occur later in the development ant thus play a role in word learning at later stages of acquisition. The same can be hypothesized for L2 language development. **Thus the hypothesis 4 states that the effect of phonological factors on word learning will probably differ depending on the language, and, more specifically, the stage of acquisition of that language.** Learning words of a completely foreign language would probably be facilitated mostly by initial prosodic processing. Learning words of language that is not known very well will be most probably facilitated by formation of phonological categories in this language (abstraction) and chunking. Learning words of a known (native) language might be facilitated by chunking and perhaps specification. How exactly those hypotheses will be reflected in the results, will depend on the exploratory factor analyses of the phonological tasks. If there is a separate factor that could be associated with the initial prosodic processing, it should correlate mostly with the word learning tasks in the foreign languages (L2 and LX). Chunking, which will be reflected most probably by blending tasks, should correlate with word learning task in known languages (L1, and to some extent L2). If abstraction processes is distilled from the data in the exploratory factor analysis, it will probably be most important for the L2 word learning.

As can be judged from the above presentation, the study has a largely exploratory character. While it makes certain specific predictions about the possible results, it is mostly aimed at gathering initial data about the influence of phonological processing on word learning in L2, in light of the severe lack of other studies in this area.

5.3. Participants

49 third-graders from 10 public primary schools in Poznań, Poland were recruited for the study. This age group was chosen, because at the age of 9 children should already have fully developed representations of their L1 and could be considered mature native speakers. At the same time, they should not have much experience with other languages apart from English which is taught at Polish schools from the first grade. The English exposure of the participants is likely to be more uniform in this age group than in older speakers.

Parents were given detailed information about the studies and asked to fill in the background questionnaire for their children. The questions concerned children's exposure to English and other foreign languages, history of speech and language disorders and prolonged stays abroad (Appendix A). Among the children chosen for the study none were diagnosed with hearing impairments, dyslexia and other global language impairments and none had lived abroad for longer than a month. All children had normal to high IQ range. Four children did not complete all of the tasks and thus were excluded from the study, one was excluded on account of his above-average exposure to English (the child was enrolled in regular extra-curricular English since 2006). Thus overall data from 44 children were gathered during the first part of the study. All of the participants had English classes at school twice a week (each class took 45 minutes). 12 children had participated in additional English classes in the course of their life. These were usually tutoring session aimed at helping children catch up with the material presented in the classroom. The mean age of participants in 9;2 yrs, range: 8;2 - 10;8, SD: 6.75 months.

Of the 44 initial participants, 30 were recruited for the second part of the study. These were the students who did not attend extracurricular classes in English from November 2013 till June 2014. Here the mean age of participants (at time 1) was also 9;2

yrs, range: 8;2 - 10;0, SD: 6.48 months. Both children and their parents gave their written consent to participate in the study.

5.4. Design

The study was designed as a correlational, longitudinal research, in which phonological task scores in L1 and L2, as well as phonological STM scores were predictor variables, while scores on experimental word learning tasks and increase in English vocabulary scores over the school year were outcome variables.

The children were tested at the beginning of the school year (T1) and at the end of the school year (T2). During T1 (November 2013 to January 2014) each child was tested with English vocabulary tests, phonological STM tests, phonological tests in L1 and L2 and experimental word learning task with L1 word pairs, L1 non-words, L2 non-words and LX non-words. Additionally, children's non-verbal intelligence was measured with Polish version of the Raven's Coloured Progressive Matrices (Raven et al. 2003)[3]. This data session was aimed at gathering data about the relationships between phonological processing/phonological STM and word learning in an experimental setting.

During T2 (May to June 2014), children's English vocabulary was measured again to record the progress made by the participants and investigate how it related to their performance on phonological measures. In other words, this part of the study provided data about English learning in naturalistic setting, rendering more ecological validity to the study. Children were also tested once again on phonological STM tasks and phonological tasks in both languages. Collecting vocabulary and phonological data at both T1 and T2 can provide some information about the direction of influence between the vocabulary and phonology in language learners. If there is an increase in vocabulary, but not phonological scores over the period of school year, it is more probable that phonological factors influence word learning and not vice-versa.

[3] Raven's Coloured Progressive Matrices (Raven et al. 2003) – standardized intelligence test, used in several studies on phonological memory and word acquisition (Gathercole and Baddeley 1990; Gathercole et al. 1997). The participants see colourful designs and need to choose from a range of options the most appropriate element that complements a given design.

5.5. Materials and methods

The following section will present all the tests used in the study. First, L2 vocabulary tests will be presented, followed by the description of the novel word learning tasks. Then the reader will be directed to the description of phonological tasks and phonological STM. Within each subsection, the stimuli will be presented first, followed by the description of procedures.

5.5.1. L2 vocabulary

Two tests were used to test the participants' expressive vocabulary in English. One of them was a standardised American test used for assessing the vocabulary level of people living in the USA. This test had the advantage of being very comprehensive and widely used by American clinicians, however, it also had the disadvantage of not being suited to the type of vocabulary the children were likely to acquire at school. Therefore an additional test was devised, which was based on the vocabulary items taught in Polish handbooks.

5.5.1.1. Test 1: Expressive One-Word Picture Vocabulary Test

Stimuli:
The American test was Expressive One-Word Picture Vocabulary Test - Fourth Edition (EOWPVT-4) (Martin and Brownell 2010). The test is composed of 190 coloured line drawings depicting objects or groups of objects. The drawings are organized from the easiest ones (depicting most common objects) to the most difficult ones (requiring knowledge of more complex terms).

Procedures:
In this task the child is shown one by one a series of pictures and is required to name the objects (or states) depicted in them. For each correctly named object in the picture the child receives one point. The task is terminated once the child gives six incorrect an-

swers in a row. The administration procedures of EOWPVT-4 are based on norms for American speakers aged 2.0 to 80+. In the standard procedures these norms define the first picture presented to the participant. Younger participants start with the first picture in the test, while older ones start with picture depicting a more complex concept and occurring further in the test. Since the test has no norms for Polish participants, children in this study always began the task with the first picture in the set.

5.5.1.2. Test 2: Handbook test

Stimuli:

The test devised for the purpose of the study (henceforth Handbook Test) was created on the basis of four most popular handbooks used in schools in Poznań, including the schools in which the study was later conducted. These handbooks were: Bugs World 2 and 3 (Macmillan), Our Discovery Island 2 and 3 (Pearson), English Adventure 2 and 3 (Pearson), and New Sparks 2 and 3 (Oxford University Press). A list of 46 words occurring in all four handbooks and taught to children in the second and third grade was created. The majority of these words were nouns, since children in primary school are taught very few verbs. A graphic artist created 46 colourful 20cm x 20 cm pictures that were to elicit the production of these words. Then a norming study was conducted to establish the reliability and validity of the test.

Three questions were asked in the norming study: A) Are the pictures understandable to children and do they depict what they are supposed to depict? B) Is the test reliable? And C) Is the test sensitive, i.e. does it reflect differences in vocabulary knowledge between children?

To answer these questions, 167 children were recruited to take part in the study in two primary schools in Toruń. 61 second-graders participated in testing the adequacy of the pictures. Each child was given 16 pictures from the set and was asked name them in Polish. Since five children did not finish the test or did not understand the task, data from only 56 children were used. Overall, the objects and states presented in the pictures were easily recognisable for children. Only four pictures were recognised in less than 75% of cases and these were excluded from the final test.

The rest of participants - 41 third-graders and 65 fourth-graders – were asked to name the pictures in English. Each child was given 16 pictures from the set to name. The number of correct answers for each child was calculated. Then statistical tests were conducted to assess the reliability of the test and to see whether there were statistically significant differences in performance between the children in third grade and those in the fourth grade (who should know more words since they had more exposure to English). Split-half method was used to establish the reliability of the test. For all versions of the test the results of the two halves correlated at statistically significant levels (r_s = 0.64, $p < .05$, r_s = 0.46, $p < .05$ and r_s = 0.67, $p < .05$). Therefore the test could be considered reliable. When it comes to sensitivity, statistical analyses have shown that there were clear differences between children when it comes to performance on the test. The mean number of items named correctly was 10.11 with *SD* = 3.65 (averaged across all version of the test). The minimal score in one version of the test was 2, the maximal score was 16. However, fourth-graders were not significantly better than third-graders on the test[4].

The final version of the Handbook Test prepared after the norming is presented in Appendix B. The test consists of 42 pictures, ordered from the easiest ones to the most difficult ones.

Procedures:

As in the EOWPVT-4, the child is presented with the pictures one by one and asked to name the depicted object or state. Unlike in EOWPVT-4 the test is not terminated after six incorrect answers - every participant is shown all 42 pictures.

[4] In fact, the fourth-graders did not display any advantage on the test. The mean score for the third-graders was 11.64 (*SD* = 3.27) and for the fourth-graders - 9.42 (*SD* = 3.72). As can be seen the fourth-graders' scores were both much more variable and much lower than the scores of the younger children. T-test has shown the difference in scores to be statistically significant ($t(93.06) = 2.74$, $p < .01$). This can be due to the fact that one class of fourth graders was definitely worse than the other ones. After removing this group from the data (9 participants altogether), the difference was not statistically significant anymore as calculated with Wilcoxon rank sum test ($W = 960.5$ $p = 0.18$).

5.5.2. Novel word learning tasks

In the study there were four tasks that measured how fast children acquired new lexical items. These tasks were based on paired associates paradigms used in several studies that examined the relationship between phonological processing or phonological STM and novel word learning (Gathercole and Baddeley 1990; Gathercole et al. 1997; de Jong et al. 2000; Cheung 1996; Hu 2003, 2008; Jarrold et al. 2009). In each task the children were asked to learn three pairs of words or three pairs consisting of a word and a non-word[5] in 15 or less trials. In the first task, children had to learn pairs of words in L1, in the second - L1 non-words, in the third - L2 non-words, in the fourth – foreign (LX) nonwords. All the words and non-words had the same phonological structure.

Stimuli:

All items in all tasks were CVCCV speech sequences or CCVCV sequences with penultimate stress placement, recorded in a professional studio.

The cue words in the tasks were composed of the most common initial CCV and CV syllables in the Polish language, combined with the most frequent final CCV and CV syllables in the Polish language, as based on the phonotactic corpus of Polish (Dziubalska-Kołaczyk et al. 2013). This was done to ensure that they were all well-formed Polish words. The words were further controlled for Polish phonotactic probability (with a measure called all-ngram[6] frequency), phonological neighbourhood density (with a measure called PLD20[7]) and lexical frequency, all based on the National Corpus

[5] The small number of pairs to learn is normal for this kind of paradigm. Learning non-words is difficult for children and, as we encountered in the pilot study, three pairs constituted an optimal amount of learning material. Tasks with three pairs are difficult enough to distinguish skilled children from children with word learning problems, but are easy enough not to discourage participants.

[6] The all-ngram frequency is a measure of phonotactic probability. It indicates how common in the given languages are the sound sequences used in a particular non-word or word. It is defined as the logarithm of mean frequency of occurrence for all ngrams occurring in a (non)word (bi-grams, tri-grams, and so on), normalized for length. For instance, to calculate all-ngram frequency for the word „pam", we would take the mean from the frequency of the combination of sounds /#p/, /pa/, /am/, /m#/, /#pa/ /pam/, /am#/, /#pam/, /pam#/, and /#pam#/, and then take the natural logarithm of this mean, weighted for the length of the chunk for which the frequency is calculated. The all-ngram frequencies were derived from an automatically phonologized version of the National Corpus of Polish (Przepiórkowski et al. 2012).

[7] PLD20 (Phonological Levenshtein Distance) is a measure of phonological neighbourhood density. It is defined as the minimal number of alterations (substitutions, deletions or insertions of phonemes) necessary to transform one string of phonemes into another (Levenshtein 1966). For instance, the word /spæm/ and the word /pæm/ have the PLD of 1, because only one operation (deletion of a phoneme) is needed to

of Polish (Przepiórkowski et al. 2012). Six words from the set were used for the Polish words task, nine were assigned randomly to the tasks with L1, L2 and LX non-words as cues.

The non-words for the task were created by combining frequent or infrequent CCV and CV syllables of Polish or English into bi-syllabic structures. The CCVs and CVs were taken from phonotactic corpora of Polish and English (Dziubalska-Kołaczyk et al. 2013). All the non-words were then controlled in terms of all-ngram frequency in Polish and English and PLD20 in Polish and in English. The metrics for Polish were based on the National Corpus of Polish (Przepiórkowski et al. 2012), while the metrics for English were based on the CUV3 corpus (Sobkowiak 2006), a corpus derived from CUV2 (Mitton 1992) with American phonological transcription added. These metrics are presented in Table 1. Additionally Polish non-words were tested for word-likeness in an online survey conducted among Polish speakers and English non-words were tested for word-likeness in an online survey conducted among American English speakers. All the stimuli are presented in Appendix C.

Table 1: Mean lexical frequency, all-ngram frequency and PLD20 for all stimuli used in the tasks

	Polish lexical frequency (mean)	**Polish all-ngram frequency (mean)**	**English all-ngram frequency (mean)**	**Polish PLD20 (mean)**	**English PLD20 (mean)**
Polish cue words ($N = 15$)	26.29	5.13	N/A	1.16	N/A
Polish non-words ($N = 3$)	N/A	4.77	2.56	1.26	2.49
English non-words ($N = 3$)	N/A	4.79	3.87	1.75	1.95
LX non-words ($N = 3$)	N/A	2.49	1.94	2.22	2.88

L1 words task: this was a baseline task, in which participants had to memorise three pairs of words in Polish. The metrics of the words used in the study are presented in Table 1 and the pairs are presented in Appendix C. The words were controlled for the log all n-gram frequency and PLD20. The all n-gram frequency for these words was

change one into another. PLD20 is the mean PLD of 20 words that are the closest phonological neighbours to the target word.

between 4.24 and 5.95, the PLD20 between 0.95 and 1.65. The lexical frequency of these words (all forms of the words were included) varied from 0.5 tokens per million (*tarka*) to 49.5 tokens per million (*matka*). The words were produced by two native speakers of Polish.

L1 non-words task: this task was supposed to emulate the acquisition of new native vocabulary. The participants had to memorise three pairs consisting of a Polish word and a Polish non-word. The Polish non-words were created by combining 30 most frequent word-initial Polish CCV structures and 30 most frequent word-initial CV structures with 30 most frequent word-final CCV structures and 30 most frequent word-final CV structures. The non-words used for this task had a very typical structure, as indicated by high all-ngram frequency in Polish (see Table 1). Consequently, they were assessed as highly word-like (rating 5 on a 5-point scale of word-likeness) or word-like (rating 4) by 23 Polish participants in the norming study (M = 3.41, SD = 0.20). Both words and non-words were produced by the same two native speakers of Polish.

L2 non-words task: this task was to simulate the acquisition of vocabulary in the second language. The participants had to memorise three pairs consisting of a Polish word and an English non-word. The English non-words were created by combining 30 most frequent English word-initial CCVs and 20 most frequent English word-initial CVs with 30 most frequent final CVs and 30 most frequent final CCVs. Items containing /θ/ and /ð/, which are complex and developmentally late English consonants not occurring in Polish, were excluded. As evidenced by all-ngram and PLD20 metrics in English the words were well-formed English-like tokens. This was confirmed in the norming study, in which all the items were considered highly word-like by a group of 11 American speakers (mode rating 5 on a 5-point scale of word-likeness, M = 3.91, SD = 0.64). The non-words were produced by two native speakers of American English.

LX non-words task: this task was to simulate learning words of a completely new language, thus it was important that the LX non-words had a phonological structure that was atypical both for Polish and for English and that they were pronounced with an accent that was completely unknown to the participants. At the same time the non-words had to be pronounceable to the children and thus the speech sequences used for this task had to be legal in Polish. Such stimuli were created by combining the least frequent 11 word-initial CV syllables and 10 word-initial CCV syllables with the 14 least frequent word-final CVs and 18 word-final CCVs from the phonotactic corpus of Polish language. All the base syllables conformed to the Syllable Sequencing Generalisation (Selkirk et al. 1984). The non-words created for this task had a structure that was atypical for both Polish and English, as evidenced the low English and Polish all-ngram frequencies. They were also assessed as either non-word-like (rating 2 on the 5 point-scale of word-likeness) or very non-word-like (rating 1) by 23 Polish native speakers, who heard the items recorded with Polish accent for the wordlikeness norming study (M = 2.04, SD = 0.38). All the items have been pronounced by two native speakers of Russian, since Russian was a language none of the participants knew.

Procedures:

Each task had the same structure. The stimuli were presented auditorily via loudspeakers with E-prime software (there was no visual input). At the beginning of each task, the participants were informed that they would learn three new words in a strange language. In this language, the instruction went, things are called differently. For instance a lamp is called /spuva/. The task began with the presentation of the three pairs. The children listened to each pair once and were asked to repeat each pair upon hearing it. Then they were informed that they would be tested on how well they remembered the three pairs and so the test trials ensued. In each test trial the children heard the first item (the cue word) from one pair, and were asked to say the word or non-word that went with the cue. If they answered correctly, they were praised and they listened once again to the correct answer. They also received one point for the correct answer. If they did not answer correctly, they heard from the experimenter that this answer was not correct and then they listened to the correct answer. In such cases they did not receive any points for this pair in this trial. After the first pair was tested, the participants heard the cue word from another pair and the whole procedure was repeated. In each trial each of the three pairs were tested in a randomised order and for each correct answer the participants could receive one point, which gives the maximum of three points per trial.

The task was terminated either if participants went through 15 trials or if they gave three correct answers in each of two consecutive trials (so they had six correct answers in a row). It was assumed that participants who gave correct answers in two consecutive trials learned all the pairs and so they would give only correct answers in the following trials. Thus, for instance, if one participant had three correct answers in the first trial and then three correct answers in the second trial it was assumed that this participant would give only correct answers in the next 13 trials, so a maximum of points was awarded to this child for these 13 trials (39 points). Overall, this child would score 45 points for the whole task (which is the maximal score, since there are 15 trials and in each trial participants could score the maximum of three points). Terminating the task after two successful trials was introduced to minimise the effort of the children and the psychological load placed on them. Repeating the same items several times after the participants learned them is very tedious and discouraging. The paradigm employed ensured that the task was performed relatively quickly and was less exhausting for the children.

Since the experimenters made online decisions about whether the participants' answers were correct or not and children tend to mispronounce non-words, an assessment protocol was introduced for all the tasks in the set involving non-words. In accordance with this protocol, a non-word answer was considered correct if the participants made no more than one mistake (substitution, deletion, epenthesis, metathesis) in a consonant of the non-word (rendition of /mekli/ as /mepli/ or /melki/ is considered correct /melpi/ or /meplın/ - incorrect). The vowels were not assessed, due to unreliability of vowel assessment in on-line judgements. These criteria were developed experimentally in the pilot study conducted on 9 participants in May and June 2013.

All participants performed all four experimental tasks. Before the first task in the series children performed a practice task with two practice pairs. The aim of the practice task was to familiarise the participants with the paradigm. The practice task was similar to the Polish words task, which was performed first.

5.5.3. Phonological processing

The phonological skills of the participants were measured with three tasks in English and three analogical tasks in Polish. The English tasks were taken from a standardised battery *Comprehensive Test of Phonological Processing - Second Edition (CTOPP-2)* (Wagner et al. 2013) used by several clinicians to assess children's phonological abilities (for instance Gorman 2012). The Polish tasks were taken from a Polish standardised battery used for diagnosing children with dyslexia - *Dysleksja 3* (Dysleksja 3 2009). The tasks are described below.

5.5.3.1. Elision

Stimuli:

The English task was *Elision* taken from *CTOPP-2* (Wagner et al. 2013). The task was composed of 34 words, in which the participant was required to delete a morpheme or a phoneme. The result of phoneme/morpheme deletion in each word was another word.

The Polish task was *Usuwanie fonemów* [Phoneme deletion] from *Dysleksja 3* (Szczerbiński and Pelc-Pękała 2009), composed of 23 words, in which the participant was required to delete a phoneme. The result of phoneme deletion in each word was a non-word.

Procedures:

For each task, the participant was tested individually by an experimenter. In the Polish version, the experimenter – native speaker of Polish – produced each word and asked the participant to say it without a given sound, e.g. "Powiedz słowo 'domek' bez 'k'" ([Say the word /ˈdɔmɛk/ [house] without /k/], correct answer /ˈdɔmɛ/). The child received a point for each correct answer. In the English version of the task, the experimenter played the items to the child from the computer. The items were produced with American English accent by a bilingual Polish-American speaker. This procedure was introduced, because the experimenter was not a native speaker of American English and she could have problems producing the words with correct and consistent pronunciation of the items. In the English version of the task, the child first had to produce the whole word (the recorded voice said: "Powiedz słowo 'span'" [Say the word 'span']), so that the experimenter could correct the child's inaccurate pronunciation of the item. Once the participant produced the item correctly, she was encouraged to say this word again without a particular sound (the recorded voice said: "Teraz powiedz słowo 'span' bez 's'" [Now say the word 'span' without 's']; correct answer – "pan"). For each correct answer the child received a point. The point was awarded if all the consonants in the word were produced correctly by the child – the quality of vowels was not assessed. The task was terminated if the child produced three incorrect answers in a row.

5.5.3.2. Blending words

Stimuli:

The English version of the task was *Blending words* task from *CTOPP-2* (Wagner et al. 2013). The task consists of 33 recordings, in which a native speaker of American English produces a sequence of speech sounds that together formed a word (for instance /s/ /ʌ/ /n/).

The Polish version of the task consisted of 33 recording, in which a native speaker of Polish produces a sequence of speech sounds that together formed a word. This task was a translation from the English version, prepared by the author of this thesis. A translation was made since no test equivalent to blending words could be found in Polish. The adaptation was made by providing Polish phonological equivalents to the words used in the English version. For instance, if the word *miss* (/m/-/ɪ/-/s/) was used in the English task, a similar Polish word *miś* (/m/ /i/ /ɕ/) [teddy-bear] was used in the adaptation.

Procedures:

In each of the tasks, the participant heard a sequence of sounds played from the computer and had to produce a word consisting of the sounds. For instance, in the English version, the recorded speaker said: "What word do these sounds make: 'To-i'?" and the child had to answer "toy". A similar question was asked in the Polish version ("Jakie słowo powstanie nam z tych dźwięków: m/ /i/ /ɕ/?" [What word would we get from these sounds: m/ /i/ /ɕ/]). In both versions of the task, the child received one point for each correct answer. In the English task, the answer was considered correct if the child produced all the consonants correctly – the quality of vowels was not assessed. Each task was terminated after the child produced three incorrect answers in a row.

5.5.3.3. Blending non-words:

Stimuli:

For the English version, the *Blending non-words* task from *CTOPP-2* (Wagner et al. 2013) was used. This task is analogical to blending words, but it uses non-word stimuli. It consists of 30 recordings, in which a native speaker of American English produces a series of sounds that together form a pronounceable non-word.

For the Polish version, four tasks were selected tasks from *Nieznany Język* diagnostic battery (Bogdanowicz 2009). The Polish task was slightly different from English, since it involved both creating non-words from sounds (called in Polish *Synteza słuchowa* [Auditory synthesis]), as well as dividing non-words into phonemes and syllables (in Polish *Analiza słuchowa* [Auditory analysis]). There were four components in the task: blending non-words from syllables, blending non-words from phonemes, analysing non-words into syllables and analysing non-words into phonemes. The blending non-words from syllables and blending non-words from phonemes components contained the total 13 sequences of sounds that formed pronounceable non-words. The analysis components contained the total of 13 non-words that had to be divided either to syllables or to phonemes.

Procedures:

For the English task, the child heard the sound sequences played from the computer. The recorded speaker asked, for instance, “What made-up word do these sounds make: 'nim-by'?” and the child had to answer: “nimby”. For each correct answer the child received a point. The answer was considered correct if the child produced all the consonants correctly, while the quality of vowels was not assessed. The task was terminated after the child produced three incorrect answers in a row.

For the blending components of the Polish task, the sound sequences were produced by the experimenter, as required by the standard administration procedures. The child was asked to produce a non-word out of each sequence. For each correct answer the child received a point. The child was asked to do all 13 trials in the task. For the analysis components of the task, the experimenter produced a non-word to the participant and then asked her to say all the syllables or phonemes in the non-word. The child was also required to clap her hands once for each syllable or phoneme. The child was asked to do all 13 trials in the task and for each correct answer she received one point. The answer was considered correct if the child clapped her hands for an appropriate number of times and properly segmented the non-word. In the statistical analysis the blending scores for the Polish non-word tasks and the analysing scores were separated.

5.5.4. Phonological short-term memory

Four phonological STM tasks were used in the study. To ensure that all these tasks were comparable, the same Immediate Serial Recall paradigm (henceforth ISR) has been used in all of them. The first task was a standard digit span task, the second task was ISR task with Polish non-words, the third task was ISR task with English non-words and the fourth task was ISR task with LX non-words. Using tasks with different types of non-words (both word-like and non-word-like) allowed for measuring both pure phonological STM as well as chunking.

5.5.4.1. Digit span

Stimuli:

The task was composed of shorter and longer series of digits taken from the Wechsler intelligence scale for children (Wechsler 1974). The forward digit span consisted of 14 sequences (two sequences of three digits, two of four digits, two of five digits etc. up to sequences of nine digits). The backward digit span also consisted of 14 sequences (two sequences of two digits, two of three digits, etc. up to eight digits). The digits were recorded by a native speaker of Polish in a professional studio. The pauses between the digits were between 450 and 600 ms long.

Procedures:

In the forward digit span the child was played a sequence of digits and was asked to repeat it in the order of presentation. Following the answer, another sequence was presented and the procedure was repeated until the end of the task. The child was awarded one point for each sequence, in which all digits were accurately repeated in the correct order. The task was terminated after two incorrect answers in a row. Forward digit span was followed by backward digit span, in which the child was presented with a sequence of digits and had to repeat it backwards. The procedures were similar to that in forward digit span. For each sequence in which the digits were produced accurately and in the right order, the child received one point. In the statistical analysis, both scores was analyzed jointly.

5.5.4.2. ISR tasks with non-words

Stimuli:

ISR - Polish non-words: the non-words used in this task had the same structure and were created using the same procedures as the stimuli for the L1 non-words learning task. They were recorded by two native speakers of Polish. The non-words were characterised by high all-ngram frequency (between 4.07 and 5.92) and PLD20 (between 1.25 and 1.85) in Polish, as measured against the National Corpus of Polish (Przepiórkowski et al. 2012). They were also tested in a word-likeness study with Polish native speakers and assessed as either wordlike or very wordlike. The stimuli used for the task are presented in Appendix D.

ISR - English non-words: the non-words in this task were created using the same procedures as the stimuli for the L2 non-word learning task. They were recorded by two native speakers of American English. The non-words were characterised by high all-ngram frequency and PLD20 in English. They were assessed either wordlike or very wordlike by native speakers of American English. The stimuli used for the task are presented in Appendix D.

ISR - LX non-words: these non-words were created using the same procedures as the ones in the LX non-word learning task. They were recorded by two native speakers of Russian. The non-words were characterised by low phonotactic probability in Polish and in English, as measured with all-ngram frequency. Phonological neighbourhood density as expressed by PLD20 in Polish and in English was also low. The items were assessed as non-wordlike or very non-wordlike by Polish native speakers.

Procedures:

The tasks were modelled after the digit span. For all tasks, the participant heard a series of pre-recorded words and non-words presented via loudspeakers. Immediately upon hearing the sequence, the participant had to repeat all the non-words in this sequence in the same order in which they were presented. The sequences became gradually longer, posing a greater challenge to the phonological memory of the speaker. If the child made a mistake in two sequences in a row, the task was terminated. For each correct sequence the child received one point. The sequence was considered incorrect if the child repeated the items in the wrong order, did not repeat all the items, or (for the non-word tasks) mispronounced one of the items in the sequence. The participant could ask for the repetition of each sequence once. The criteria for mispronunciation of the items in the non-word tasks were created on the basis of the pilot study and were the same as in the case of the non-word learning task. A mispronunciation occurred if the child made more than one mistake (substitution, deletion, epenthesis, metathesis) in a consonant of the non-word.

All ISR tasks with non-words started with one non-word to repeat. Then another non-word was given for repetition. Then there were two trials in which the child had to repeat sequences of two non-words, two trials with sequences of three non-words, two trials with sequences of four non-words and two trials with sequences of five non-words. The pauses between the non-words in the sequences lasted for about 500 ms.

5.6. General procedures

Each child was tested individually on all the tasks in a quiet room. Testing sessions took place at participants' schools, after classes. Parents could accompany the children during the administration of the tasks, but the majority of children chose to be tested alone. Each session lasted between 30 and 60 minutes. The structure of each session is described below:

T1 (November 2013 - January 2014)

Session 1:

- L1 word learning task
- L1 non-word learning task
- English Vocabulary - EOWPVT-4
- English Vocabulary - Handbook test

Session 2:

- L2 non-word learning task
- Raven's Coloured Progressive Matrices
- Forward and backward digit span
- Polish phonological awareness tasks (Elision, Blending words, *Nieznany Język*)

Session 3:

- LX non-word learning task
- ISR - Polish non-words
- ISR - English non-words
- ISR - LX non-words
- English phonological awareness tasks (Elision, Blending words, Blending non-words)

For most participants there was a week-long break between each testing session. The break never exceeded four weeks.

T2 (May 2014 - June 2014)

Session 1:

- English Vocabulary - EOWPVT-4
- English Vocabulary - Handbook test
- English phonological awareness tasks (Elision, Blending words, Blending non-words)

Session 2:

- Forward and backward digit span
- ISR - Polish non-words

- ISR - English non-words
- ISR - LX non-words
- Polish phonological awareness tasks (Elision, Blending words, *Nieznany Język*)

As with the T2, for most participants there was a week-long break between the testing sessions. The break did not exceed four weeks.

5.7. Statistical analysis

To answer the research questions posed in this study and evaluate the factors involved in foreign word learning, two types of outcome variables were used. The first type of variable comprised of experimental novel word learning task scores. The second type of variable was a measure of L2 vocabulary learning progress after six months of L2 classroom instruction. For each type of variable a separate statistical analysis has been conducted. These analyses are described below.

5.7.1. Experimental novel word learning tasks

First, to extract the factors underlying the performance on phonological tasks in L1 and L2, an Exploratory Factor Analysis was run on all the scores of the phonological task scores (Polish elision, Polish blending words, Polish blending non-words, Polish analysing non-words, English elision, English blending words, English blending non-words). Three factors were extracted in this way (see Table 2), which lend themselves to a relatively easy interpretation. The first factor is loaded primarily by elision task in English (thus it will be referred to henceforth as PhonEN). The second factor is loaded by Elision scores in Polish (thus it will be referred to as PhonPL). The third factor reflects blending scores in both languages (thus the name PhonBlend).

Table 2: Summary of the exploratory factor analysis loadings for the phonological tasks ($N = 44$)

	PhonEN	PhonPL	PhonBlend
Polish elision	0.206	**0.975**	
Polish blending	0.128	**0.435**	**0.421**

Polish non-word blending		0.196	**0.784**
Polish non-word analysis		**0.618**	0.323
English elision	**0.875**	0.254	
English blending	**0.617**	0.216	**0.487**
English non-word blending	**0.713**		**0.401**
Eigenvalues	1.811	1.673	1.299
Proportion of variance	0.299	0.239	0.186

Then, to establish the predictors of the novel word learning tasks, multiple regression analyses were run for each task (L1 word learning, L1 non-word learning, L2 non-word learning, LX non-word learning). For each model the following predictors have been considered: age, raw Raven scores (non-verbal IQ), digit span, ISR task with Polish non-words, ISR task with English non-words, ISR task with LX non-words and the three phonological factors extracted via the Exploratory Factor Analysis. The best model for each task was identified using the best-subsets procedure with leaps package in R: (Lumley and Miller 2004). Then each model was tested to see whether it met the necessary assumptions.

5.7.2. Vocabulary learning progress in L2

The next step in the analysis was calculating the difference between vocabulary score at T2 and vocabulary score at T1 for both vocabulary tests (EOWPVT-4 and Handbook Test). Then for each vocabulary difference a regression analysis was conducted using the same method as in the case of novel word learning tasks. The predictors considered for each model were: age, Raven raw scores, T1 digit span, T1 ISR task with Polish non-words, T1 ISR task with English non-words, T1 ISR task with LX non-words, T1 PhonEN, T1 PhonPL and T1 PhonBlend.

Because the relationship between vocabulary and phonological factors is probably bidirectional, it was also investigated whether the performance on L2 phonological tasks has improved over the year under the influence of richer L2 vocabulary. Therefore the difference between scores obtained at T2 and T1 was calculated for the three phonological tasks in English. For each of these differences, a regression analysis was performed with age, Raven raw scores, both T1 vocabulary scores and all T1 ISR tasks as possible predictors.

5.8. Results

5.8.1. Experimental word learning tasks

Raw data used in these analyses are presented in Appendix E. The descriptive statistics for all the tasks are presented in Table 3. As can be observed from both the raw data and from the table, the ISR tasks with English and LX non-words offered very little variance between the participants and were non-parametric. Therefore they were excluded from further analyses. The correlation matrix for all the measures used in the statistical analysis is presented in Table 4.

Table 3: Descriptive statistics for all the measures at T1

	MEAN	SD	MEDIAN	MIN	MAX
Age (months)	110.43	6.75	110.00	98.00	128.00
Raven raw scores	29.11	3.35	29.00	23.00	36.00
Vocabulary EOWPVT	12.55	8.31	11.50	0.00	44.00
Vocabulary Handbook	21.84	8.10	22.00	0.00	38.00
Word learning – L1 words	36.11	7.42	38.50	12.00	44.00
Word learning – L1 non-words	34.00	8.22	36.00	8.00	45.00
Word learning – L2 non-words	27.98	7.82	29.00	4.00	42.00
Word learning – LX non-words	21.59	10.93	22.00	2.00	43.00
Digit span	9.48	2.31	10.00	4.00	16.00
ISR L1 non-words	6.09	1.27	6.00	3.00	9.00
ISR L2 non-words	4.52	1.02	4.50	3.00	6.00
ISR LX non-words	4.02	0.73	4.00	3.00	6.00
Polish elision	20.18	3.23	21.50	8.00	23.00
Polish blending	23.00	5.21	23.50	11.00	32.00
Polish non-word blending	8.39	2.23	8.00	3.00	13.00

Polish non-word analysis	11.95	1.70	13.00	4.00	13.00
English elision	21.82	9.66	25.50	0.00	33.00
English blending	18.45	5.06	19.50	6.00	27.00
English blending non-words	14.86	5.98	16.00	1.00	25.00

Table 4: Correlation matrix for the measures at T1

	Raven raw scores	**Digit span**	**ISR L1 non-words**	**Pho-nEN**	**PhonPL**	**PhonBlend**	**Word learning L1 words**	**Word learning L1 non-words**	**Word learning L2 non-words**	**Word learning LX non-words**
Age	0.03	-0.05	-0.17	-0.11	-0.15	-0.13	-0.25	0.02	-0.07	-0.06
Raven raw scores		0.22	-0.08	0.18	0.12	0.07	0.19	**0.30***	**0.33***	0.24
Digit span			0.17	0.23	**0.37***	**0.45****	0.28	**0.32***	0.27	0.27
ISR L1 non-words				0.15	-0.19	0.25	0.13	0.09	0.00	0.18
PhonEN					0.03	0.10	0.19	0.24	**0.44****	**0.42****
PhonPL						-0.01	0.26	0.09	0.21	0.09
PhonBlend							-0.11	0.12	0.08	0.27
Word learning L1 words								**0.39****	0.22	**0.39****
Word learning L1 non-words									**0.48*****	**0.50*****
Word learning L2 non-words										**0.63*****

* $p < .05$; *** $p < .01$; *** $p < .001$

5.8.1.1. L1 word learning task

This task was the baseline condition and as could be predicted, no interesting relationships with investigated factors have been observed. This can already be gleaned from the correlation matrix (Table 4), which shows no significant correlations with this task and is further confirmed by the regression analysis. The best model chosen with the all-subset regression is presented in Table 5. The model does not predict any significant variance (R^2 = .08, $R^2_{Adjusted}$ = .06, p = .068).

Table 5: Regression model of the L1 word learning task

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	27.64	4.64	5.95		< 0.001 ***
Digit span	0.89	0.48	1.88	0.28	0.067 .

. p < .01; * p < .05; *** p < .01; *** p < .001

5.8.1.2. L1 non-word learning task

As can be already observed on the basis of the correlation matrix (Table 4), the performance on this task was not related to any phonological factors, but it was related to digit span and Raven scores. The scatterplots in Figure 5 show the relationship between the most important predictors and the native novel word learning. In the regression model, which is presented in Table 6, the only statistically significant predictor of scores on this task was digit span, which partly confirms the hypothesis about the involvement of phonological STM in the word learning process. However, this model explains little variance – as little as 10% (R^2 = .10, $R^2_{Adjusted}$ = .08, p = .037).

Table 6: Regression model for the L1 non-word learning task

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	23.36	5.08	4.60		< 0.001 ***
Digit span	1.12	0.52	2.15	0.32	0.03 *

. p < .01; * p < .05; *** p < .01; *** p < .001

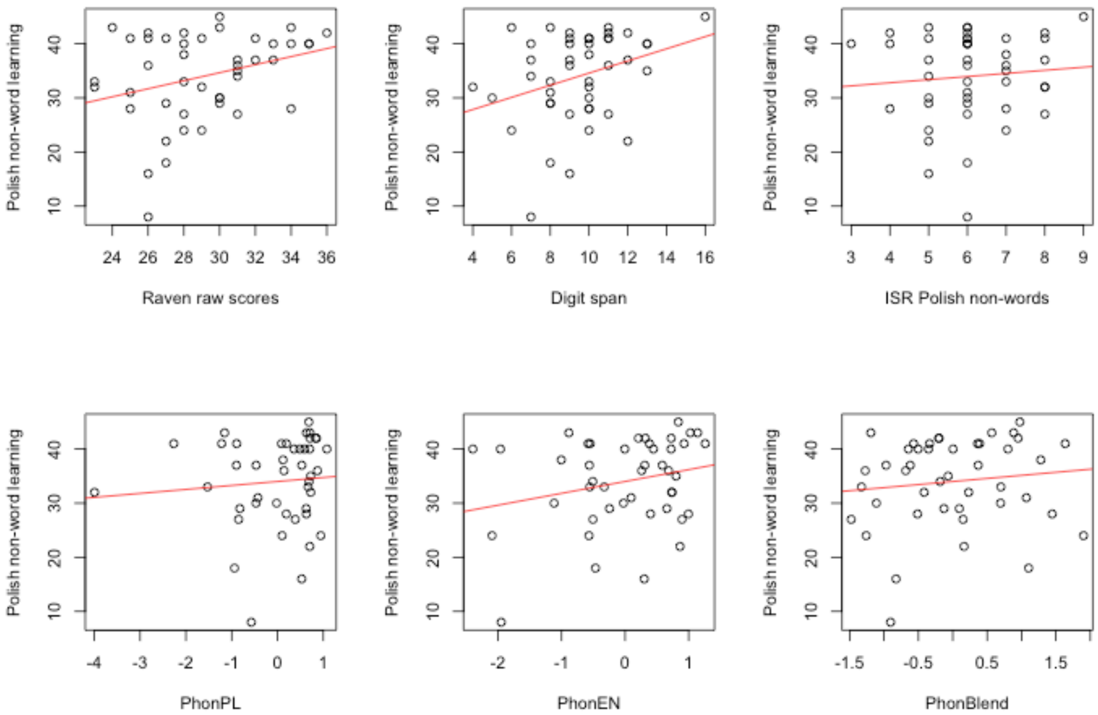

Figure 5: Scatterplots showing the relationships between the most important predictors and the L1 non-word learning scores

5.8.1.3. L2 non-word learning task

As can be seen from the correlation matrix (Table 4), the scatterplots (Figure 6) and the regression model (Table 7), the most significant predictor of L2 non-word learning score turned out to be the PhonEN factor, which confirms the hypothesis about phonological variables playing a crucial role in word learning. Another factor influencing the score was non-verbal intelligence. Overall, this model explains a significant portion of variation in the results – above 20% ($R^2 = .26$, $R^2_{Adjusted} = .22$, $p = .002$).

Table 7: Regression model for the L2 non-word learning task

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	10.07	9.35	1.08		0.287
Raven raw scores	0.62	0.32	1.93	0.26	0.060 .
PhonEN	3.32	1.16	2.85	0.39	0.007 **

. $p < .01$; * $p < .05$; *** $p < .01$; *** $p < .001$

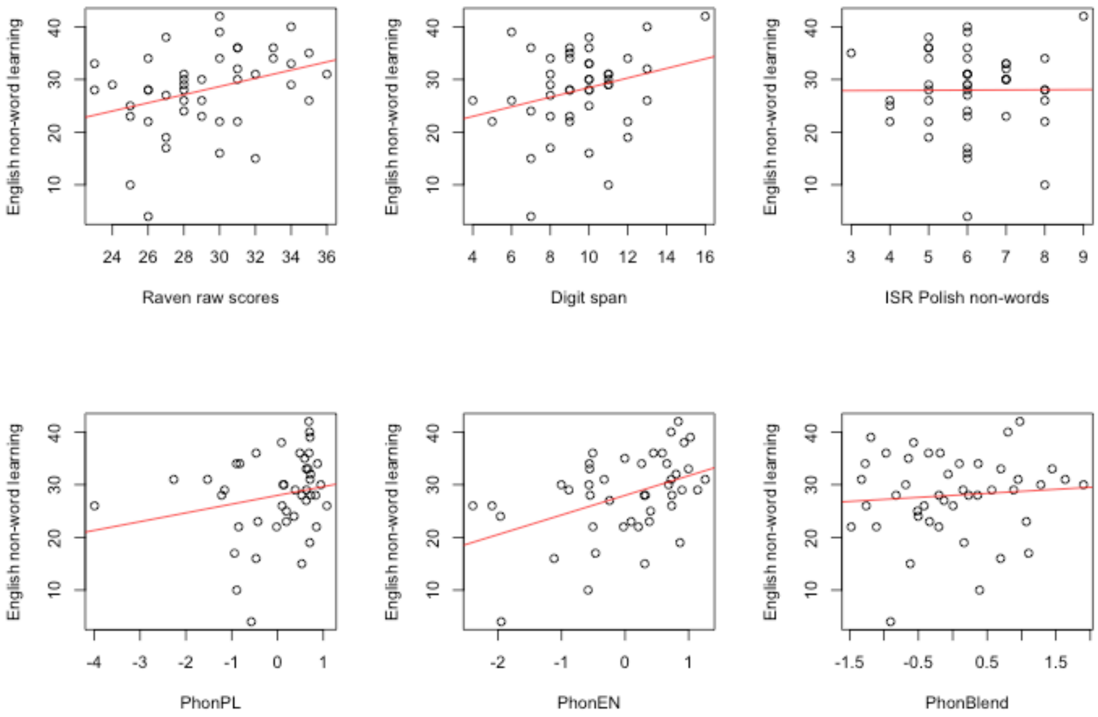

Figure 6: Scatterplots showing the relationships between the most important predictors and the L2 non-word learning task scores

The above analysis supports the hypothesis 1A about the involvement of the phonological factors (PhonEN) in L2 word learning, but invites further questions about the nature of this relationship. In particular, it begs the question of how this relationship is moderated by the proficiency level in L2. In the Aims section (page 126) one of the research questions posed was devoted to the influence of phonological processing on word learning at different stages of acquisition. While all participants of this study were at the advanced stages of acquisition in L1 and at very initial stages of acquisition in LX, they might have varied in terms of L2 acquisition stage. For this reason, an additional analysis was conducted to investigate whether the relationship between PhonEN and L2 non-word learning was obtained for all the students, only the more L2 advanced students or only the less proficient L2 students.

The participants were divided into two groups on the basis of their English proficiency. The scores on *EOWPVT-4* vocabulary test were taken as measures of this proficiency. Participants who scored below the median on *EOWPVT-4* were classified as low proficiency (N = 22), and those who scored above the median were classified as high proficiency (N = 22). Pearson's correlation analyses between the PhonEN factor and L2 non-word learning scores were conducted for both groups. The results show the relationship between PhonEN and L2 non-word learning score for the low proficiency group (r = .47, p = .025, 95% CI [0.07, 0.75]), but not the high proficiency group (r = .06, p = .78, 95% CI [-0.37, 0.47]). This pattern can be clearly observed from the scatterplots presented in Figure 7. These findings indicate that the performance on English elision was a significant predictor of word learning only at the initial stages of acquisition.

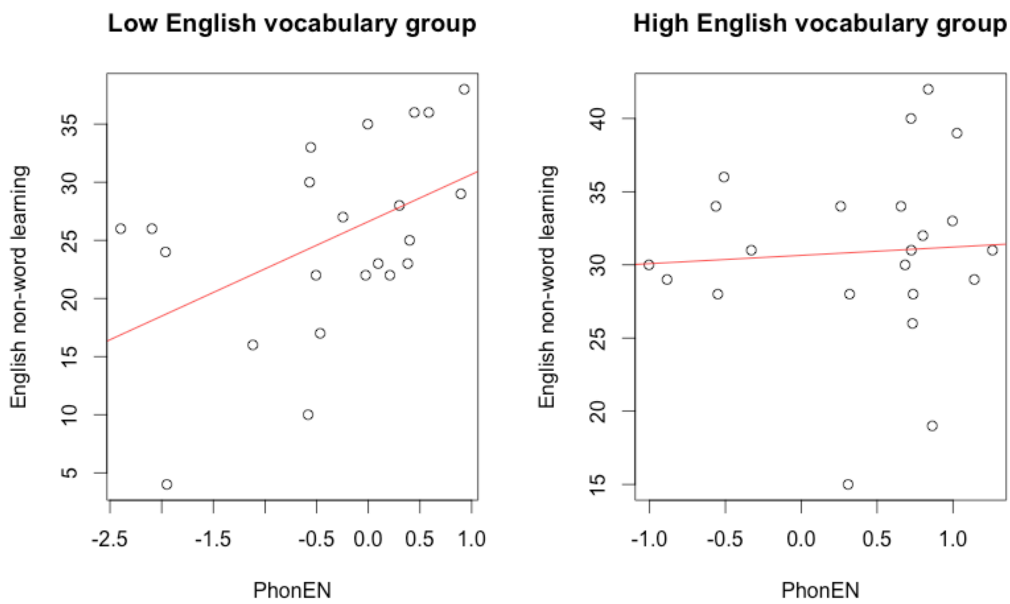

Figure 7: The relationship between PhonEN factor and English non-word learning performance in high and low English proficiency groups

5.8.1.4. LX non-word learning task

As can be gleaned from the correlation matrix (Table 4) and the scatterplots (Figure 8), the one significant predictor of LX non-word learning scores was PhonEN. This result has also been confirmed by the regression analysis, in which PhonEN was the only significant predictor of the non-word learning performance (Table 8). The regression model containing just this predictor explains a significant portion of variation in the results (R^2 = .17, $R^2_{Adjusted}$ = .15, p = .005). Overall, this result confirms the hypothesis that phonological processing tasks are related to learning words in a foreign language.

Table 8: Regression model for the LX non-word learning task

	B	SE B	t value	ß	p
(Intercept)	21.59	1.52	14.5		< 0.001 ***
PhonEN	4.98	1.67	2.99	0.42	0.005 **

. $p < .01$; * $p < .05$; *** $p < .01$; *** $p < .001$

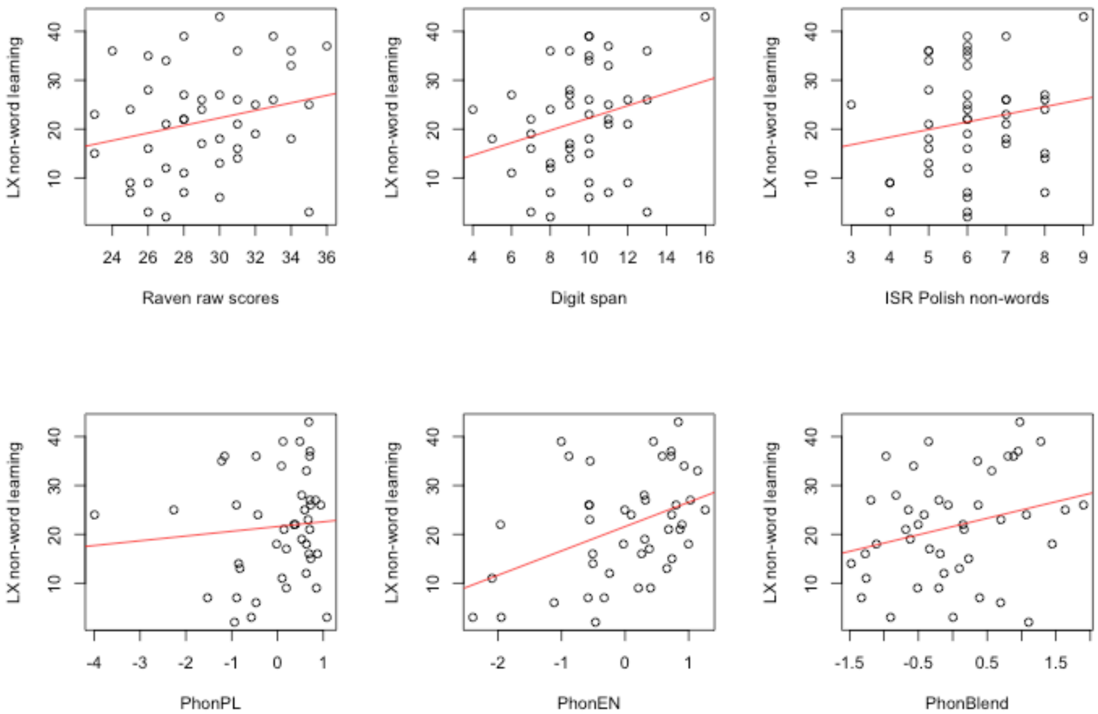

Figure 8: Scatterplots showing the relationships between the most important predictors and the LX non-word learning task scores

5.8.2. Vocabulary learning progress in L2

Raw data used for this set of analyses can be found in Appendix , while descriptive statistics are given in Table 9. As previously, data from the ISR tasks with English and LX non-words were excluded from the regression analyses due to very low variance.

Table 9: Descriptive statistics for children in the T2 group

	MEAN	SD	MEDI-AN	MIN	MAX
Age (months)	109.93	6.48	110.00	98.00	120.00
Raven raw scores	29.30	3.30	29.00	23.00	36.00
Vocabulary EOWPVT T1	12.00	9.20	10.50	0.00	44.00
Vocabulary EOWPVT T2	15.07	11.05	12.00	0.00	48.00
Vocabulary EOWPVT difference (T2-T1)	3.07	4.88	3.00	-9.00	19.00
Vocabulary Handbook T1	21.37	8.28	21.00	0.00	38.00
Vocabulary Handbook T2	24.37	8.76	26.00	1.00	39.00
Vocabulary Handbook difference (T2-T1)	3.00	3.62	3.50	-6.00	9.00
Digit span T1	9.57	2.58	10.00	4.00	16.00
Digit span T2	10.03	2.43	10.00	6.00	18.00
ISR L1 non-words T1	6.20	1.21	6.00	4.00	9.00
ISR L1 non-words T2	6.07	1.23	6.00	4.00	9.00
ISR L2 non-words T1	4.43	1.04	4.00	3.00	6.00
ISR L2 non-words T2	4.20	1.21	4.00	2.00	8.00
ISR LX non-words T1	4.07	0.69	4.00	3.00	6.00
ISR LX non-words T2	3.53	1.01	4.00	2.00	5.00
Polish elision T1	20.10	3.03	21.50	13.00	23.00
Polish elision T2	20.50	2.85	21.50	14.00	23.00
Polish blending T1	22.67	5.44	22.50	11.00	32.00
Polish blending T2	25.43	3.04	26.50	19.00	31.00

Polish non-word blending T1	8.50	2.37	8.00	3.00	13.00
Polish non-word blending T2	9.13	2.47	9.00	4.00	13.00
Polish non-word analysis T1	11.93	1.93	13.00	4.00	13.00
Polish non-word analysis T2	12.33	1.35	13.00	8.00	13.00
English elision T1	20.63	9.56	21.50	0.00	33.00
English elision T2	23.37	8.91	25.00	1.00	33.00
English elision difference (T2-T1)	2.73	8.79	2.00	-15.00	24.00
English blending T1	18.27	5.23	19.00	6.00	27.00
English blending T2	21.17	4.26	21.00	12.00	31.00
English blending difference (T2-T1)	2.90	3.14	2.50	-2.00	11.00
English blending non-words T1	15.40	5.34	16.00	1.00	25.00
English blending non-words T2	17.30	5.21	16.50	4.00	28.00
English blending non-words difference (T2-T1)	1.90	3.97	2.00	-5.00	14.00

5.8.2.1. Handbook vocabulary test

First of all, the difference in vocabulary scores between T2 (M = 24.37, SD = 8.76) and T1 (M = 21.37, SD = 8.28) was calculated. The difference between scores on T1 and T2 was statistically significant, as indicated by a dependent t-test ($t(30) = 4.54$, $p < .001$). On average children learned 3 new words during the school year (95% CI [1.65, 4.35]), although, there was much variance among participants. Some children learned as many 9 new words, while others actually performed worse on the task at T2.

Second of all, multiple regression analysis was conducted for the Handbook vocabulary scores differences. None of the predictors (age, Raven raw scores, digit span measured at T1, Polish non-words ISR at T1, phonological factors measured at T1) were related to the vocabulary increase. The scatterplots in Figure 9 show almost flat regression lines.

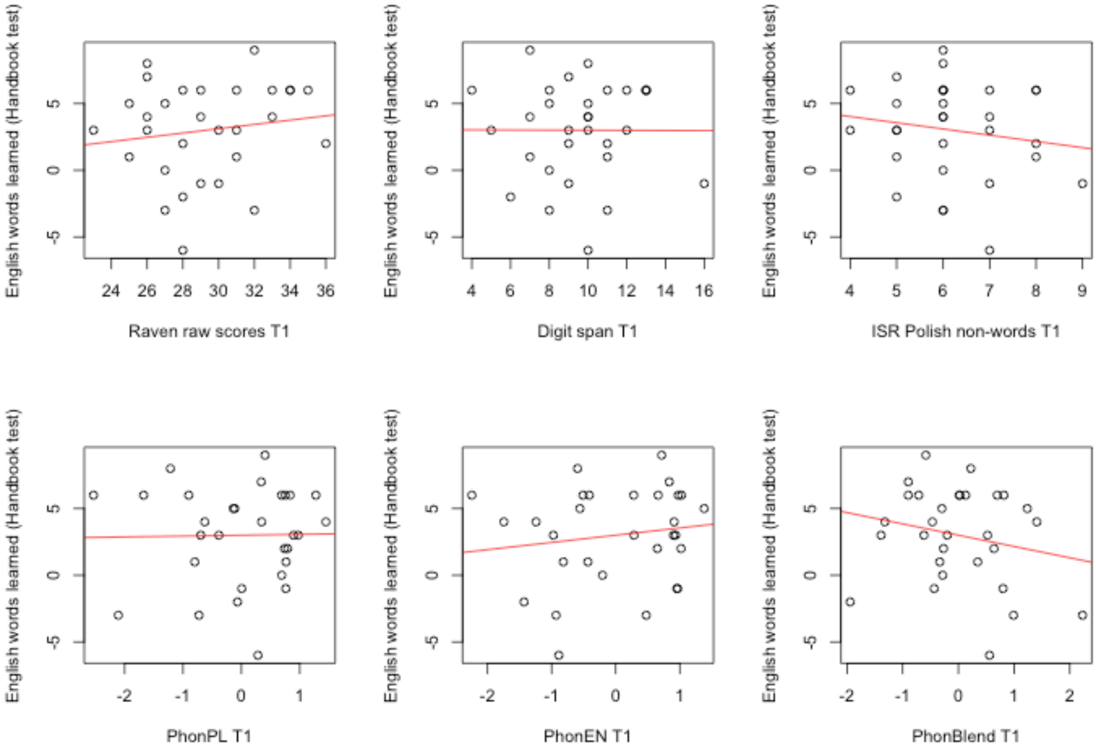

Figure 9: Scatterplots showing the relationships between the most important predictors and the vocabulary increase from T1 to T2 (Handbook test)

The best regression model is presented in Table 10. As can be seen, none of the predictors turned out significant. The best model chosen via the all-subsets method contained age as predictor, but overall it explained insignificant portion of variance (R^2 = .08, $R^2_{Adjusted}$ = .04, p = .14). It seems that despite trying to control for exposure to English by eliminating the students who had extra-curricular classes from the participant pool, there were too many external variables that could influence children's performance to observe any effect of phonological factors.

Table 10: Regression model for the T1 to T2 vocabulary increase as measured with the Handbook test

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	19.90	11.19	1.78		0.086 .
Age	-0.15	0.10	-1.51	-0.28	0.141

. $p < .01$; * $p < .05$; *** $p < .01$; *** $p < .001$

5.8.2.2. EOWPVT-4

A similar pattern of results was observed also for the EOWPVT-4 test. The difference between T2 (M = 15.07, SD = 11.05) and T1 scores (M = 12.00, SD = 9.20) was statistically significant ($t(30) = 3.44$, $p = 0.002$). On average children learned 3.06 new words throughout the year (95% CI [1.25, 4.89]). There were very large differences between children, with some learning as many as 19 words and others forgetting as many as 9.

As presented in the scatterplots (Figure 10), there is no relationship between the vocabulary increase from T1 to T2 and any of the predictors. The same is shown in the regression model (Table 11). As with the Handbook test, no significant predictors were found. The best model contained the PhonPL factor as predictor, but the model itself explained insignificant amount of variance ($R^2 = .05$, $R^2_{Adjusted} = .02$, $p = .22$)

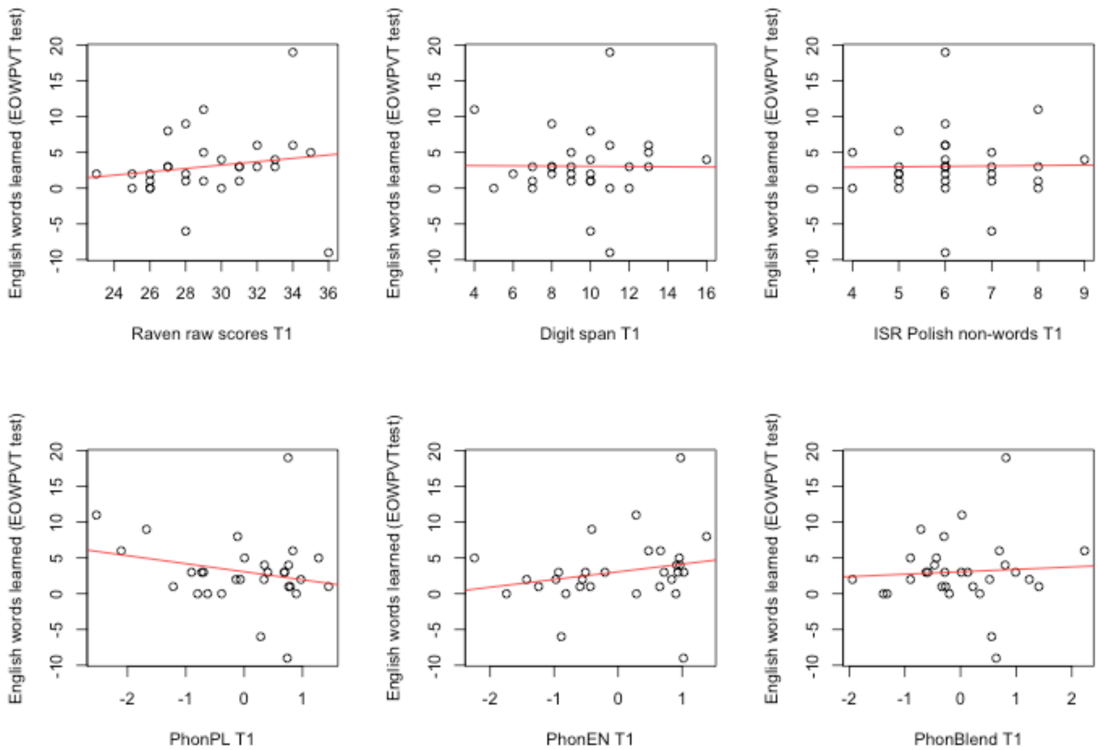

Figure 10: Scatterplots showing the relationships between the most important predictors and the vocabulary increase from T1 to T2 (EOWPVT)

Table 11: Regression model for the vocabulary increase as measured with EOWPVT

	B	SE B	t value	ß	p
(Intercept)	3.07	0.88	3.48		0.002 **

PhonPL	-1.13	0.90	-1.25	-0.23	0.222

. $p < .01$; * $p < .05$; *** $p < .01$; *** $p < .001$

5.8.2.3. The effects of vocabulary on English phonological tasks performance

Because first language acquisition data suggest that the relationship between phonological development and vocabulary development is bi-directional, the statistical analysis contained an investigation of the increase in L2 phonological scores throughout the few months of learning English at school. For all English phonological tasks (elision, blending, blending non-words) there was an average increase in scores from T1 to T2, although not for all tasks was this increase statistically significant. The participants scored on average 2.73 points more on English elision task during T2 (95% CI [-0.55, 6.02]), but the difference was non-significant as shown by a t-test ($t(30) = 1.7$, $p = .09$). For the English word blending task there was an average increase by 2.9 points (95% CI [1.73, 4.07]), which was statistically significant ($t(30) = 5.05$, $p < .001$), while for the English non-word blending tasks there was a statistically significant increase by 1.9 points (95% CI [0.42, 3.38], $t(30) = 2.62$, $p = .013$).

For none of the English phonological tasks was there any relationship with the predictors (age, Raven scores, digit span at T1, ISR with Polish non-words at T1, Handbook vocabulary test at T1, EOWPVT at T1). For the increase in English elision, the scatterplots are presented in Figure 11, while the best regression model is presented in Table 12. The model has very little predictive value and there is no effect of any of the predictors on the difference in elision scores ($R^2 = .07$, $R^2_{Adjusted} = .04$, $p = .143$). For the increase in English blending scores there is a marginally significant effect of vocabulary scores at T1 as measured by the Handbook test (Figure 12), but this relationship is inverse (the greater the vocabulary at T1, the smaller the gain in English blending). The model itself (Table 13) is also marginally significant ($R^2 = .11$, $R^2_{Adjusted} = .09$, $p = .061$). For the increase in the English blending non-word task, the results were non-significant ($R^2 = .11$, $R^2_{Adjusted} = .08$, $p = .069$). The scatterplots for this task are presented in Figure 13, while the regression model is presented in Table 14.

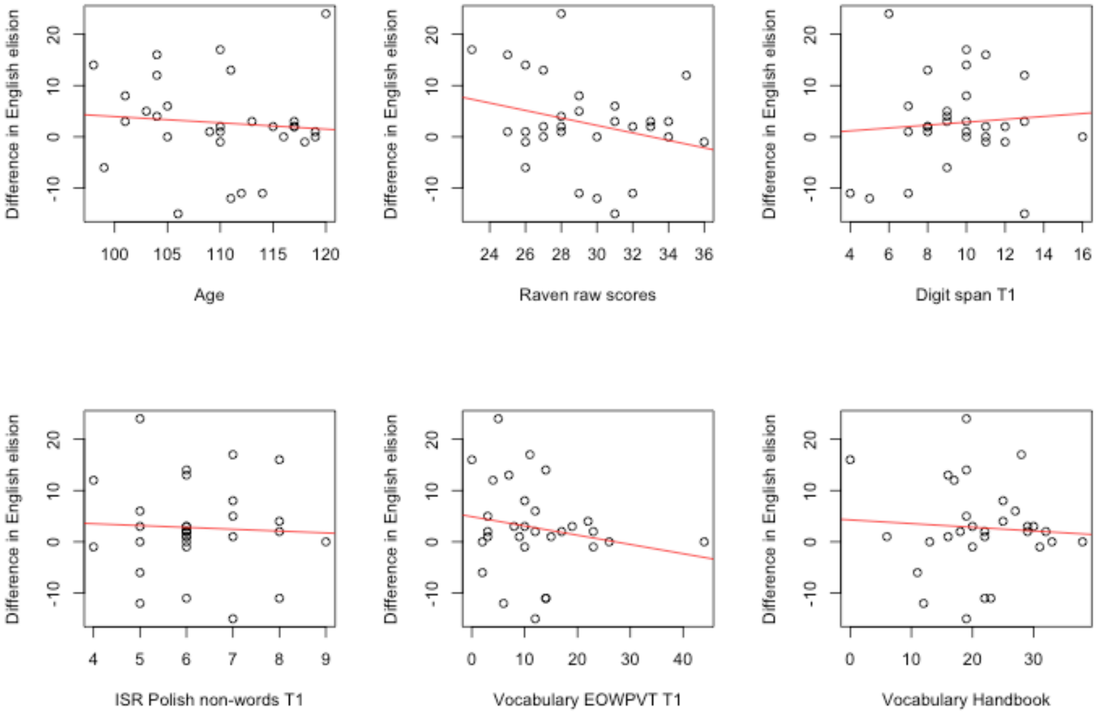

Figure 11: Scatterplots showing the relationships between the predictors and the difference in English elision scores from T1 to T2

Table 12: Regression model for the increase in English elision scores

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	24.09	14.26	1.69		0.102
Raven raw scores	-0.73	0.48	-1.51	-0.27	0.143

. *p* < .01; * *p* < .05; *** *p* < .01; *** *p* < .001

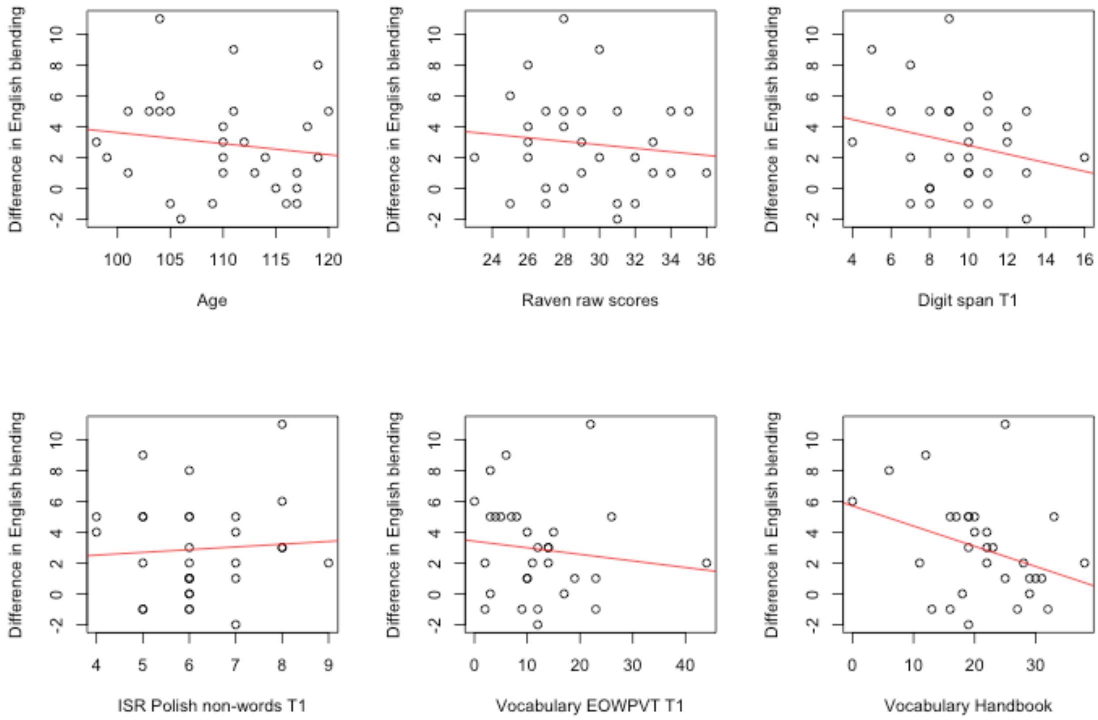

Figure 12: Scatterplots showing the relationships between the predictors and the difference in English blending scores from T1 to T2

Table 13: Regression model for the increase in English blending scores

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	5.71	1.54	3.71		< 0.001 ***
Vocabulary Handbook	-0.13	0.07	-1.95	-0.35	0.061 .

. *p* < .01; * *p* < .05; *** *p* < .01; *** *p* < .001

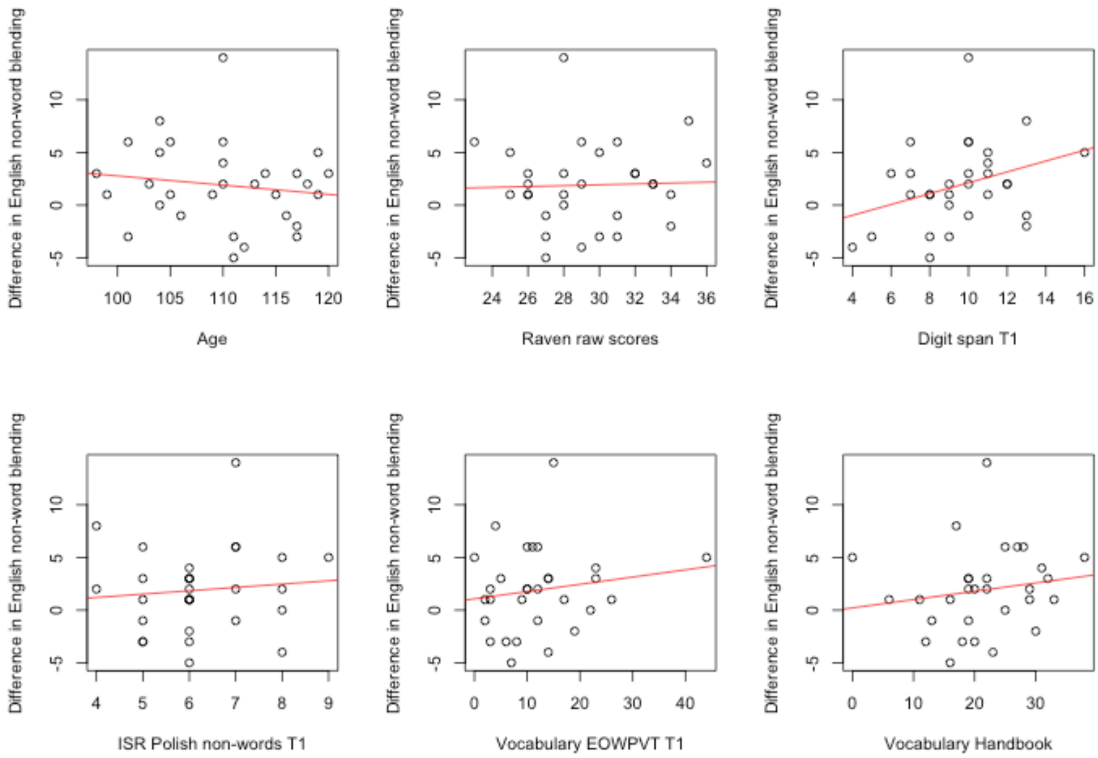

Figure 13: Scatterplots showing the relationships between the predictors and the difference in English non-word blending from T1 to T2

Table 14: Regression model for the increase in English non-word blending

	B	**SE B**	**t value**	**ß**	**p**
(Intercept)	-3.03	2.71	-1.12		0.272
ISR digits	0.52	0.27	1.89	-0.34	0.070 .

. *p* < .01; * *p* < .05; *** *p* < .01; *** *p* < .001

Chapter 6: Discussion

6.1. Introduction

Four research questions have been asked at the beginning of chapter 5. This chapter will be an attempt to answer these questions with data collected for the purpose of this work. The first part of the chapter will be devoted to the phonological factors in foreign word learning in the experimental and natural settings. Then the possibility of a language-universal factors underlying L1 and L2 phonological development will be discussed. The chapter will also attempt to describe the different roles of phonological STM and phonological processing in word learning. Finally, the effect of phonological processing on word learning at different stages of acquisition will be discussed.

6.2. L2 phonological development and L2 word learning

Research question 1: Does phonological development in L2 influence word learning in this language (both in experimental tasks and in natural language learning environment)? If so, in what ways? Which L2 phonological processes facilitate word learning?

At the beginning of the previous chapter it has been hypothesised that phonological development in L2 should go hand in hand with lexical development in L2 (hypothesis

1A). This hypothesis has been put forward on the basis of the findings suggesting an intimate relationship between phonology and lexicon in L1 acquisition (de Jong et al. 2000; Bowey 2001; Storkel 2001; Windfuhr and Snowling 2001; Ramachandra et al. 2011). It was expected that a similar relationship would be observed for L2 both in the word learning experiment and in the natural classroom environment. In other words, it was predicted that in the study participants, L2 phonological tasks should correlate with the performance on the L2 word learning task, as well as with the increase in L2 vocabulary throughout the school year.

This hypothesis was only partly corroborated. The factor related to the performance on English phonological tasks (PhonEN) was a very strong predictor of performance on the L2 non-word learning task. However, it was not a predictor of L2 vocabulary increase over the course of school year. There are two possible explanations to these findings. One explanation is that phonological performance in L2 is related to the speed of word learning, which was tested with the experimental non-word learning task, however it is not related to word retention. That means that learners with better phonological skills are faster to acquire new vocabulary items over short periods, however. this learning efficiency does not translate to long-term knowledge of the learned words.

Another possible explanation is that the relationship between phonological processing and retention of L2 words has not been detected, because there was too much noise in the vocabulary learning data. The study had been designed in such a way as to minimise this noise. Firstly, all the children tested at T2 were exposed to the same amount of English input (45 minutes twice a week) in the period between T1 and T2, since none of them had additional English classes. Secondly, children learned L2 in state schools in accordance with the government-approved curriculum. Thirdly, all participants used one of the four handbooks that were the basis of the Handbook Test of English vocabulary (Bugs World, Our Discovery Island, English Adventure, and New Sparks). Nevertheless, there were many additional factors such as socioeconomic status, motivation and teacher's engagement that could not be controlled for and that could have exerted a significant effect on children's progress in English word learning. It is worth noting that while some children improved significantly in terms of L2 vocabulary size over the school year, others actually performed worse on the second administration of the tests. This variation might reflect differences in learning styles of the participants. It could also reflect teaching methods favoured at schools. It is possible that some of the teachers promoted rote learning and learning for the test and did not practice the material enough. Other teachers might have continually practiced the material with students during subsequent classes and encourage the children to actively engage with the vocabulary by using new words in all sorts of communicative situations. Seeing that the children were recruited across ten different primary schools in Poznań the differences in teaching styles might be a valid explanation for the results.

Despite these problems, the data clearly indicate that there is an effect of phonological processing on L2 word learning. The question now is - what kind of processing is involved in learning. It has been hypothesised that phonological factors should influence foreign vocabulary acquisition in three different ways (Hypothesis 1A). Firstly, initial prosodic processing should initialise the word learning process. Secondly, creation of abstract phonological representation in L2 should facilitate learning words in L2. Thirdly, chunking should make the learning process more efficient. The results indicate that learning novel L2 words might have been influenced by the initial prosodic processing and/or the creation of abstract phonological representations in L2. There is no evidence for the involvement of chunking in the process.

This is because apart from Raven raw scores, which might indicate a general learning capacity, the factor loaded by the English elision task (PhonEN) was the only significant predictor of L2 non-word learning task scores. Elision task involves listening to English words, dividing them into smaller units and then deleting one of the units. As such, this task can reflect two different skills. One of them is the capacity to prosodically process the stream of foreign speech and in this way find possible junctures between units (initial prosodic processing). The other is connected to the knowledge of abstract L2 units - more specifically, it is the ability to extract such units from speech (abstraction). On the basis of the regression model for the L2 non-word learning task it is difficult to establish which of these processes underlies the relationship between English elision and English word learning.

However, this problem has been partially solved by applying the post-hoc analysis to the collected data and checking whether the PhonEN factors influences word learning differently in low proficiency English speakers and high proficiency English speakers. If the relationship between English elision and word learning were accounted for by the initial prosodic processing, then this relationship would have occurred only at the initial stages of acquisition, because prosodic processing facilitates vocabulary development primarily at the first stages of language learning. Therefore, for the participants, who are less advanced speakers of English, there should be a strong correlation between the PhonEN factor and L2 non-word learning. For the more advanced students of English, who presumably do not rely to such an extent on the prosodic processing to find word and unit boundaries, the correlation between PhonEN and L2 non-word learning should be not as significant. If, on the other hand, the relationship between English elision and English non-word learning were due to the process of L2 abstraction, then this relationship should have obtained for the whole sample and in particular the more advanced students. The correlational analyses in section 5.8.1.3 show that PhonEN was a significant predictor of L2 novel word learning only for the low proficiency participants. This result indicates that the effect of PhonEN on L2 word learning speed should be attributed primarily to the initial prosodic processing of foreign sounding speech.

When it comes to the third phonological factor, i.e. chunking, the data collected do not provide convincing evidence for its involvement in word learning. As indicated in the previous chapters, chunking works by making the storage of new words in phonological STM more efficient (Jones 2011, 2012). However, in the regression model for the L2 novel word learning task, none of the phonological STM tasks turned out to be a significant predictor. This indirectly disproves the involvement of chunking in the learning task. Of course, this result should be treated with care. Some of the STM tasks used in the study (including the ISR with English non-words, which was supposed to be the task most sensitive to chunking in English) offered too little variability to be included in the regression model. Nevertheless, the two tasks that were included in the model - digit span and ISR with Polish non-words - did not account for significant portion of variation in the results. The implications of this are considered in more detail in section 6.4 of this chapter.

All in all, the data collected in this study indicate a relationship between phonological and lexical development in L2. Specifically, it seems that the component called initial prosodic processing facilitates L2 word learning at the initial stages of acquisition. The results do not, however, provide much evidence for the involvement of L2 abstraction or chunking in the lexical acquisition process. These findings are based on the experimental word learning data and thus pertain to the efficacy of word learning over short periods of time. It is impossible to tell on the basis of the data collected how phonological processing influences the retention of vocabulary over long periods of time. The data on vocabulary progress throughout the school year, which were supposed to answer this question, did not show the involvement of phonology in the long-term retention of L2 vocabulary items. This, however, might be due to the presence of strong confounds in the study, such as different teachers and different teaching styles.

6.3. Universal phonological factors facilitating word learning

Research question 2: Are there language-universal factors underlying L1 and L2 phonological development and do these factors facilitate novel word learning, as suggested by Hu (Hu 2003, 2008; Hu and Schuele 2005)? If so, what are they?

Hypothesis 2 of this thesis predicts the existence of common variance in L1 and L2 phonological tasks reflecting the existence of common phonological factor facilitating novel word learning. This hypothesis has been put forward on the basis of research conducted by Hu (Hu 2003, 2008; Hu and Schuele 2005), in which the performance on L1 phonological tasks seems to be a significant predictor of learning L2 vocabulary. Different interpretations as to the nature of this factor have been suggested. One interpretation is that this factor is initial prosodic processing – the ability to prosodically process unknown stream of speech and establish tentative boundaries of the speech units. This initial prosodic processing should facilitate both the creation of phonological categories (Goswami et al. 2011) and word learning (Curtin et al. 2005). The second interpretation is that the universal factor is the sensitivity to speech-specific auditory cues which facilitates specification and which has been examined in the ERP study by Díaz et al. (2008). The third interpretation is that the universal factor has to do with the ability to create abstract phonological categories (abstract). Finally, the fourth interpretation sees the universal factor as a chunking mechanism.

To determine whether there is a common phonological factor at all, it is useful to investigate both the Exploratory Factor Analysis performed on all phonological tasks in both languages, as well as the effects of different phonological factors on all novel learning tasks. The Exploratory Factor Analysis was performed mainly to investigate whether there is an overlap between different phonological tasks. What was found is that there is such an overlap, but mainly between blending tasks in both languages. The fact that the PhonBlend factor is loaded by the four blending tasks in Polish and in English suggests the existence of common memory component that is essential to the performance on these tasks. In the blending tasks, the participants hear speech sounds, have to hold them in their memory and then compose a word out of them. This component might reflect either working memory or echoic phonologic memory - an ability to hold brief verbatim memory of speech sounds. It might also be related to chunking, since the tasks involve combining several phonological elements into larger structures. The further evidence for the relationship of this component to memory stems from the fact that it correlates moderately with the digit span tasks. It is important to notice, however, that this component did not play a significant role in any of the experimental novel word learning tasks. Therefore, it does not really fit the description for Hu's underlying phonological factor influencing word learning.

For the other two factors extracted from the phonological tasks there is very little overlap. One factor is loaded almost exclusively by English phonological tasks (notably elision), the other by Polish phonological tasks (again, mainly elision). This finding would suggest that phonological processing is different for the native and the foreign language – that it is based on two separate mechanisms. In other words, this finding does not corroborate the existence of a common factor underlying both L1 and L2 speech processing. However, before one jumps to this conclusion, it is important to bear in mind that this difference in processing might stem from a vast difference in proficiency between participants' L1 and their L2. While all the participants were mature and proficient speakers of Polish, they were not very proficient at English. The majority of participants did not know many verbs and were not able to communicate fluently in the language. They also did not have much contact with native English speech, since teachers of English in Polish primary schools are rarely native speakers of English. Therefore, it might be the case that while phonological processing of L1 in the participants was based on the recognition of abstract phonological categories and chunks, the phonological processing in L2 was in many cases still based on the initial prosodic processing. In other words, the performance on L1 phonological tasks might have reflected the knowledge of L1 phonological categories, while the performance on the L2 phonological tasks might have reflected the ability to prosodically process novel stream of speech.

The hypothesis that the performance on L2 phonological tasks reflected initial prosodic processing finds further confirmation in the results of the study - specifically the regression models for the non-word learning tasks. These models indicate that phonological processing in L2 (the PhonEN factor) was the main predictor of the performance on the experimental LX non-word learning task. The LX task simulated learning a completely new language. LX non-words were produced with an accent that was unknown to any of the participants, which made speech sounds constituting the non-words seem novel. Moreover, the speech sound combinations used in LX items were infrequent and unusual. In other words, it was fairly unlikely that the participants could apply their phonological knowledge of L1 (such as the knowledge of L1 abstract phonological categories and speech sound combinations) to this task. It seems, instead, that they must have used initial prosodic processing to process these LX items. The fact that the performance on the task was strongly connected with the performance on L2 phonological tasks, indicates that L2 phonological tasks tapped into this universal ability to process LX speech stream. To conclude, the data obtained from the LX non-word learning task seem to indicate that the factor underlying performance on English phonological tasks in many participants was the initial prosodic processing. On the other hand, it is probable that the factor underlying the performance on the Polish phonological tasks was the knowledge about Polish phonology – an inventory of well-established abstract L1 phonological representations and the representations of common L1 speech sequences – i.e. abstraction and chunking.

It is easy to notice that the results obtained in this study stand in contradiction to the results obtained by Hu. His research suggests the existence of phonological component that is manifested in L1, but is important for learning L2. The present study does not show this. No component related to L1 influenced any word learning scores in the current investigation. There are two possible explanations for this discrepancy. The first explanation would be that Hu's participants' were much more proficient speakers of L2 that the participant of the current study. As a result, the L2 phonological processing of Hu's participants was based on the extraction of abstract phonological categories from speech. Their L2 word learning was also based on tapping into abstract phonological categories. If one assumes that L1 phonological tasks administered by Hu tapped into the ability to create abstract phonological categories, then the performance on L1 phonological tasks should influence the abstraction-based phonological development and lexical development in L2.

The other explanation is that Hu's L1 phonological tasks tapped into primitive initial prosodic processing and this prosodic processing facilitated L2 word learning. This interpretation suggests, however, that the L1 phonological tasks used in the current study did not tap into this prosodic processing ability (since they have effect whatsoever on L2 word learning). This raises a question of why would Mandarin phonological tasks tap into prosodic processing in Mandarin native speakers, while Polish phonological task would not do the same in Polish native speakers. There is a good explanation for this paradox. Polish is a language with an alphabetic writing system and because of that Polish children are explicitly trained to divide speech into phonemes from the age of six. The Polish phonological tasks, which involved division of words into phonemes, required the children to do what they have been trained in for a number of years. These tasks tapped into phonological representations that have been likely influenced by years training into phoneme recognition. Mandarin, on the other hand is a language without an alphabet and thus Mandarin children most likely do not receive such extensive training in phoneme recognition as Polish children. As a result it is likely that the phonological categories of Mandarin children are not at all phoneme-based (Ziegler and Goswami 2005; Read et al. 1986) and that for them phoneme recognition is a novel and to a large extent unnatural activity. And yet the Mandarin phonological tasks involved mostly division of speech into phonemes (extractions of vowel, recognition of initial consonants). For this reason, it is likely that Hu's Mandarin tasks did not tap into L1 abstract phonological representations of the participants, but rather into the ability to find new unit boundaries in a speech sequence that usually is undivided. To conclude, it is likely that the L1 phonological tasks in Hu's research due to their phoneme-based nature tapped into the primitive prosodic processing factor, which is also of importance for learning new words of a foreign language. This explanation is more consistent with the results of the current study.

To sum up, the data in the current study support the existence of a universal phonological factor (initial prosodic processing) that is of importance for the initial processing of foreign speech stream. This factor seems to be important also for word learning at the initial stages of acquisition. The results of the Exploratory Factor Analysis indicate also that there is a common factor underlying blending tasks (regardless of the language). This factor might be related to chunking or to certain memory components. The data collected do not offer strong support for the existence of a common factor related to phonological category building in L1 and L2, but that may be because the participants of the study were not proficient enough to build phonological representations in L2. The data also do not support the hypothesis of the universal sensitivity to speech-based stimuli, since there is not much overlap between phonological scores in L1 and L2. However, it is possible that a more sensitive measure would be necessary to tap into such universal sensitivity. An example of such a measure would be an EEG response, or, more specifically, MMN response to minimal phonetic changes in the speech sounds of a native and a foreign language (see Díaz et al. 2008).

6.4. Phonological STM and phonological processing in word learning

Research question 3: Do phonological processing and phonological STM interact to facilitate word learning?

In hypothesis 3 it has been proposed that both phonological STM and phonological processing will be involved in novel word learning. This prediction is firmly grounded in the previous literature. In fact certain studies indicate that phonological processing tasks and phonological STM tasks are related and that the common underlying factor of the two variables facilitates new word learning (Bowey 1996, 2001; Metsala 1999; Gathercole 2006). There is ample evidence for the relationship between phonological STM and phonological processing. In many datasets, correlations between measures of the two variables have been found. For example, in Bowey's dataset (Bowey 2001) there is a moderate and statistically significant correlation ($r = 0.51$) between performance on phoneme identity tasks (task in which participants decide whether given words start with the same phoneme) and non-word repetition scores in pre-school children. In the data collected by Jarrold et al. (2009), there has been an even stronger correlation between performance on phonological tasks (elison, identification of words with odd rhymes, identification of words starting with the same sounds) in typically developing children aged 5-8 ($r = 0.58$). Another evidence for the relationship comes from studies Hu and Schuele (2005), in which participants with low phonological scores were also characterised by low phonological STM scores.

When one investigates the correlation matrix for the word learning tasks in the results section (Table 4: Correlation matrix for the measures at T1, one can see that in the present dataset, a similar relationship has been found. There is a moderate correlation (r = 0.37) between phonological performance in Polish (PhonPL factor) and digit span and an even stronger correlation (r = 0.45) between performance on blending tasks (PhonBlend factor) and digit span. It has to be noted that this relationship has been observed only for one of the phonological STM tasks, however this might be due to the fact that the ISR tasks with non-words were not sensitive enough to produce a meaningful correlation.

All in all, it seems that phonological processing and phonological STM are related. However, it is not know what is the common factor underlying both variables and what is the relationship between this factor and words learning. Some researchers claim that the phonological store within the STM is the common factor underlying the contribution of phonological tasks and STM tasks to word learning (Gathercole 2006). Others believe phonological processing to be the factor underlying both the effect of phonological tasks and the effect of STM task on word learning (Bowey 1996, 2001; Metsala 1999; Snowling et al. 1991). Studies investigating this issue provide contradictory findings. In one study on word form learning predictors in children with Down syndrome (Jarrold et al. 2009), it was found that phonological STM had an effect of the speed of word learning, even when the phonological processing scores were controlled for. However, there was no effect of phonological processing when the phonological STM was controlled for. This would mean, basically, that phonological STM would be the main factor influencing word learning and underlying the effects of both the non-word repetition and the phonological tasks. A similar tendency has been found among typically developing children in the same study. Bowey (2001), however, found the exact reverse pattern of results in her investigation of native vocabulary size predictors in typically developing pre-schoolers. Here, there was an effect of phonological scores on vocabulary size even with phonological STM controlled for, but there was no effect of phonological STM with phonological scores controlled for. This would mean that it is the phonological processing that is the main factor in word learning.

The explanation for these discrepancies might be that the two variables while related, might influence word learning differentially. Indirect evidence for this hypothesis comes from the study by Palladino and Ferrari (2008). The results of this study indicate that while people with language learning difficulty (including word learning difficulty) have usually problems with both phonological processing and phonological STM, there is a large portion of separate variances between these two factors. The current dataset provides further confirmation of this hypothesis, by showing that phonological processing and phonological STM influence novel word learning to a different degree depending on the language in which the words are learned. In the present study, there has been a different pattern of results for the L1 non-word learning task and the foreign non-word learning tasks (L2 and LX). For the L1 non-word task, the main predictor of the learning speed was digit span (phonological STM), but not phonological task scores. On the other hand, for both L2 and LX non-words, there was a strong relationship with the English phonological scores (PhonEN), which can be identified as representing the initial prosodic processing of foreign speech stream. There was, however, no relationship with the phonological STM scores of any kind. This would lead to the conclusion that phonological STM and phonological processing have different effects on word learning.

Before, however, one accepts this conclusion, there is one important issue to bear in mind when considering the findings of the study. The result indicating that phonological STM influences native word learning in older speakers does not match the pattern of results obtained in the previous literature. Several previous studies have shown that the effects of phonological STM on L1 vocabulary acquisition dwindle with age and are significantly less pronounced in children after the age of 8 (Gathercole et al. 1992; Baddeley et al. 1998). This, however, does not necessarily undermine the results of the current study. For instance, Gathercole's results were obtained by measuring the native vocabulary size of children with vocabulary tests and by keeping track of the change in this vocabulary size. As indicated by the results of the current investigation, the measure of actual vocabulary size and vocabulary increase in children can be ridden with noise that might be impossible to control. In the current dataset, the noise introduced into the vocabulary learning data by various confounding variables was so strong that it was impossible to detect any of the investigated effects. In the experimental word learning tasks, however, the learning conditions were more controlled and therefore it was possible to detect more subtle effects. Therefore, it is reasonable to assume that there is an effect of phonological STM on native word learning in the older speakers, however this effect is too subtle to be detected with measures of vocabulary increase and can be elicited only in more controlled conditions. All in all, on the basis of the current data it can be claimed that phonological STM influences learning word forms of the known language, while the phonological processing, or more specifically, initial prosodic processing, facilitates learning words in a foreign language.

One issue pertaining to the relationship between phonological STM and phonological processing or development is the issue of chunking. In the current study, the role of chunking was to be investigated through the use of different types of ISR tasks. Unfortunately, the ISR tasks with different types of non-words used in the study did not turn out to be sensitive enough in this particular group of participants. Two problems with this task was insufficient granularity of the task and floor effects. For this reason, the current research does not allow to tell whether the effect of phonological STM on L1 word learning was related purely to phonological STM capacity or to L1 chunking. If the study results had shown correlations between all STM tasks and L1 word learning, it would have been concluded that the effect should be attributed to the phonological STM capacity. If there had been correlations only between digit span and Polish non-word learning, as well as between ISR with Polish non-words and Polish non-word learning, the effect could have been attributed to chunking. However, with data from English and LX ISR displaying so little variability, it is not possible to determine which mechanism or skill underlies the relationship between phonological STM and vocabulary acquisition. It is fairly possible that both large memory capacity and chunking mechanism facilitate vocabulary acquisition.

To sum up, in the current study it has been found that phonological STM and phonological processing are correlated, but that they can influence novel word learning differently depending on the language. There was no effect of any of the phonological STM tasks on the experimental novel learning tasks in L2 or LX. However, for the novel word learning task in L1 there was an effect of digit span. This would suggest that phonological STM and phonological processing provide different contributions to word learning and that while phonological STM helps with learning words of a known language, phonological processing (specifically prosodic processing) facilitates learning words of a new language.

6.5. Phonological factors and word learning in different languages

Research question 4: Do phonological factors influence word learning differentially for different languages, i.e. will the influence on native word learning be different than on foreign word learning?

The differential effects of phonological STM on word learning in L1 and in foreign languages touches upon the last research question asked in this study: the effects of various phonological factors on word learning in different languages. Hypothesis 4 of this study stated that different types of phonological processing should facilitate word learning at particular stages of language acquisition. For completely new languages (i.e. the initial stages of language acquisition), the initial prosodic processing would have the greatest effect on word learning. At this stage of acquisition, learners have no speech representations of the language yet, and thus the prosodic processing is necessary for them to extract linguistic units from the stream of speech and in this way initialise the acquisition of new word form. For a partly unknown language (i.e. intermediate stages of acquisition) the acquisition of vocabulary can be related to the development of abstract phonological categories in this language and to chunking. Finally, for a well-known language (i.e. late stages of acquisition) word learning should be related purely to chunking ability. This is because in this type of word learning there is no great need anymore for initial prosodic processing, so variance in this phonological factor should not affect lexical development. Also, since abstract phonological categories are presumably already established in learners' minds at this point of acquisition, there should not be much variance in this factor that could predict word learning. Overall, is seems that chunking could be the only relevant phonological factor for L1 learning. The current study was designed to test the above hypotheses. The participants were asked to learn novel words of a completely new language (LX non-words learning task), a partly unknown language (L2 non-words learning task) and a well-known language (L1 non-words learning task).

The results of the study seem to partially confirm the hypotheses about differential influence of phonological factors on word learning. The PhonEN factor, which could be associated with initial prosodic processing, was a significant predictor of LX non-word learning and L2 non-word learning in children with lower L2 proficiency.

The L1 non-word learning task was predicted by digit span, which was the indicator of phonological STM, but could be also the indicator of chunking. The data do not allow, however, to strongly support the idea that the process of abstraction (and specification) takes part in vocabulary learning in relatively unknown languages.

To conclude, the current dataset provided some support for the hypotheses put forward at the beginning of chapter 5. Firstly, the data from experimental non-word learning task support hypothesis 1A, stating that there is a relationship between phonological development and lexical development in L2. Performance on English phonological tasks predicted performance on English non-word learning tasks, although not the increase in English vocabulary in the classroom environment. Hypothesis 1B stated that different phonological factors should influence word learning. In fact, among the factors extracted from the phonological tasks with the use of exploratory factor analysis, only the factor PhonEN, associated most probably with initial prosodic processing, predicted L2 non-word learning. Hypothesis 3, stating that both phonological STM and phonological processing could contribute to word learning was corroborated. Interestingly, the findings indicate that phonological STM is crucial for novel word learning in the native language, while phonological processing (or, more specifically, initial prosodic processing) is of importance for learning words of a foreign language. This finding at the same time supports hypothesis 4, stating that phonological factors influence word learning differentially depending on the stage of acquisition of a given language. The two main conclusions from the study findings are that A) phonological processing does influence lexical development in a foreign language and B) the involvement of phonological factors in word learning changes as language acquisition progresses.

Conclusion

The aim of this thesis was to investigate the connections between memory, phonology and lexicon in language acquisition - in particular second language acquisition. The first chapter introduced the topic by presenting research on phonological STM - a memory component responsible for storing phonological information that has been shown to influence vocabulary learning. It was proposed that part of the effect of phonological STM tasks on word learning might be attributed to the facilitative role of phonological processing in novel word acquisition.

Thus in the further chapters the relationship between phonological processing and word learning was explored in more detail. Since the notion of phonological processing seems controversial in itself, chapter 2 was devoted to the investigation of the concept. The issue of phonological processing can be boiled down to four issues. The first issue is segmentation - whether it happens as a first step of phonological processing before any categorisation of units takes place or whether it takes place as a result of unit recognition. The second is abstractness of representation - whether there are abstract phonological categories. The third issue is granularity of representation – i.e. the question of the shapes/units of speech representations. The fourth and final issue is the order of speech processing - whether it proceeds sequentially or whether different level of representations are processed at the same time. The outlook on phonological processing adopted in this thesis is that 1) it is a process of recognising speech units in the speech stream. That 2) those units form two levels of abstractness - the exemplar level and the level of abstract phonological representations, which are created as prototypes of the exemplars. That 3) there are different levels of granularity of speech representations, although it is not certain what these levels might be exactly. There is definitely the level of smaller units and the level of speech sequences. Finally, that 4) all levels of processing occur at the same time. The last point already indicates the relationship between the phonology and lexicon – phonological processing and lexical processing occur simultaneously and they feed into each other.

Once a tentative definition of phonological processing has been provided, chapter 3 moved to investigate the relationship between phonology and lexicon in L1 acquisition. Phonological development in L1 has been divided into several different processes, including A) initial prosodic processing, which is a pre-lexical type of segmentation that happens at the beginning of acquisitions before phonological representations are create, B) specification, which development on the level of exemplars, perceptual attunement to the acoustic cues in L1, C) abstraction, which is the development of abstract phonological categories in L1, and D) chunking, which is sequence learning, putting together phonological categories into speech combinations. Research indicates that each of these processes influences L1 lexical development in slightly different ways. Initial prosodic processing triggers early word learning (Curtin et al. 2005; Mattys et al. 1999). Specification helps with the recognition of known words (Edwards et al. 2002). Abstraction facilitates learning new words (de Jong et al. 2000). Chunking makes word learning more efficient, because it frees the resources of phonological STM during word learning (Jones 2011, 2012; Storkel 2001). The role of phonological processing in L1 lexical acquisition is therefore quite established in research.

What is far less established and far less researched is the effect of phonology on word learning in foreign language and at different stages of acquisition. In chapter 4 a theoretical basis of such investigations has been laid out. It has been shown that the acquisition of L2 phonology, while to a large degree influenced by L1, can be basically described in terms of the same processes as L1 phonological acquisition. There is initial prosodic processing that helps listeners tackle the "wall of sound" that is the speech stream of a completely new language (Carroll 2006). There is specification, that is the perceptual attunement to L2 contrasts (Sebastián Gallés and Díaz 2012; Sebastian-Galles and Baus 2005), and abstraction, which is related to this attunement and allows for the creation of new L2 phonological categories (Best and Tyler 2008). Finally there is chunking, which is L2 sequence learning (Speciale et al. 2004; Ellis 2008). Since these phonological factors are related to lexical development in L1, it stands to reason that they should also influence lexical development in L2. Initial prosodic processing should initialise word learning in L2, then abstraction and specification working hand in hand should facilitate novel L2 word learning, while chunking should make novel L2 word learning more efficient by helping learners use the phonological STM store more economically. This idea was the basis for the study conducted.

Since there is very little research on the relationship between phonological processing and word learning in L2, the study had exploratory character. First of all, it was necessary to establish, whether there is any relationship between phonological factors and lexical acquisition in a foreign language at all. The study asked questions about the role of phonological factors in L1 and L2 on the acquisition of L2 vocabulary in both experimental and naturalistic settings. It also investigated whether there is a common factor underlying phonological development in L1 and L2 that would be important for word learning. Yet another question asked was connected with the interaction of phonological STM and phonological processing in word learning. Finally the study examined whether different phonological and STM factors influenced word learning differently depending on the stage of acquisition.

The research presented did not manage to answer all of the questions asked. In particular the data on L2 vocabulary increase collected to test the hypotheses in more naturalistic settings turned out to be too noisy to draw any definite conclusions. Nevertheless, the study indicates that there is a relationship between phonology and lexicon in L2 word learning. The study also provides evidence for the relationship between phonological STM and word learning, although not in the foreign languages at the initial stages of acquisition. The research conducted was not able to answer all the questions related to the phonology-memory-lexicon interface, but it has shown direction of further research. It has shown which research methods work in the investigation of this topic and provided specific questions to be addressed in further studies.

The first task to be addressed in the future is to create separate measures investigating different kind of processes involved in phonological development (initial prosodic processing, specification, abstraction and chunking). Then these tasks could be used to investigate the relationship between phonology and lexicon at different stages of language acquisition more thoroughly. First of all, it might be beneficial to find physical correlates of initial prosodic processing, which seems to be involved in phonological and lexical development at the initial stages of acquisition. Carroll (2006) suggested that the sensitivity to tone (pitch) might be related to this kind of processing. These proposals have been confirmed by the studies of Henderson (1980) and Richardson (2004). It also seems that sensitivity to amplitude changes might be related to the initial segmentation of speech (Goswami et al. 2002, 2011). Thus, in further studies initial prosodic processing can be operationalized as these basic auditory skills. Sensitivity to pitch and amplitude change can be easily tapped with AXB perceptual tasks. Those tasks can provide a reliable, independent measure of initial prosodic processing. Second of all, EEG measures (in particular MMN) can be used to tap into specification - the perceptual attunement to auditory cues specific to a particular language. MMN is an ERP component that can be used to detect automatic reactions to sound contrasts. It a very sensitive measure of perceptual abilities and it has been used as such in the study by Díaz et al. (2008). To measure abstraction, sound categorisation tasks (ascribing a sound to one or other categories) can be used. Finally, the degree of chunking in a particular language might be measured with lexical decision tasks containing words and non-words with common and unusual sound sequences. All of these tasks can be used alongside the phonological awareness tasks such as elision and blending. In this way, one can investigate different processes that go into the performance of such tasks.

With the phonological tasks in place, one can investigate the relationship between different phonological factors and performance of the experimental word-learning tasks. However, it would be even more beneficial to use other, more ecologically valid methods of assessing vocabulary learning skills. First of all, further studies can use experimental word learning tasks in which words would be learned in context. An example of such paradigms are QUIL tasks (Brackenbury and Fey 2003; Ramachandra et al. 2011), in which the participants are typically presented with a story (film or picture book), introducing novel items and novel names several times. Another way of testing vocabulary learning progress would be to organise a language course, as in the study by Speciale et al. (2004), and measure the learning outcomes of the participants at the end of the course. In this design, the external variable of teaching method could be controlled and thus the data from such study would be characterised by both reliability and ecological validity.

Investigating the relationship between phonological factors and word learning is an important direction of research, because it could help develop better teaching methods and therapies directed at individuals with foreign language learning difficulties. In particular, it can show whether phonological trainings, which are often used for dyslexia treatment, could also help with language learning problems. In the times of globalisation, where bilingualism seems to be the new norm, helping people with learning new languages can have a profound impact.

SUMMARY

The aim of the thesis was to investigate the relationships between phonological processing, phonological short-term memory and novel, especially foreign, word learning. Phonological short-term memory is a memory component responsible for storage of verbal information (or, more specifically, the phonological form of words and sentences) over short periods of time. As described in chapter 1, there is a large body of research indicating that this memory module plays a crucial role in learning novel words of both the native and foreign language. As indicated at the end of the chapter, however, this theory is not without controversy. In particular, it has been suggested that the relationship between certain phonological short-term memory tasks and word learning can be attributed to another factor – phonological processing. However, this hypothesis remains largely under-researched. There is a severe lack of studies especially on the relationship between phonological processing and second language word learning. This thesis aimed to bridge this gap.

To investigate the relationship between phonological processing and word learning it is necessary first to define and understand the former notion. Phonological processing is understood differently by different researchers and thus for the purpose of this thesis it was necessary to provide some unified framework for regarding this process. Arriving at this framework is the subject of chapter 2. In this part of the thesis, phonological processing is described from the point of view of phonology, acoustic phonetics, psycholinguistics and neurobiology. At the end of the chapter, different ideas about the mechanism taken from those areas of research are combined into one approach. Within this approach, phonological processing is understood as the process of recognising in the speech stream the phonological units that are stored in the mind of the listeners. The

process is facilitated by additional segmentation strategies. The phonological units have the form of abstracted phonological categories that are based on speech exemplars.

With a tentative definition of phonological processing, the reader is directed to chapter 3, which describes the relationship between the development of phonological processing and word learning in the first language. It is proposed that phonological development in first language acquisition is composed of four processes: initial prosodic processing, abstraction (creation of abstract phonological categories), specification and chunking. Several studies are presented showing that each of those processes is related to lexical development in the first language.

Chapter 4 is devoted to the relationship between phonological development and lexical development in the foreign language – the main topic of this thesis. It is argued that phonological development in the second language is in many respects similar to the phonological development in the first language and is composed of the same four processes: initial prosodic processing, abstraction, specification and chunking. Since these processes play an important role in native word learning, it is argued, they should also play a role in foreign word learning. However, as indicated by the review of available research on this topic, so far there is a dire lack of studies that would support this hypothesis. Therefore, a new study is proposed investigating the relationship between phonological processing and word learning in the second language.

Chapter 5 presents this study, which investigates word learning in 44 Polish nine-year-olds learning English at school. The children were asked to do a range of phonological tasks in Polish and in English aiming at tapping their phonological development in both languages. They were also asked to perform phonological short-term memory tests. Moreover, the participants took several experimental word learning tasks, in which they were asked to memorise new word forms in their native language (Polish), their second language (English) and a completely foreign language (LX). It was hypothesised that the performance on those tasks should be related to the performance on the phonological tasks. Moreover, the children were tested on their L2 vocabulary at the beginning and at the end of the school year. It was hypothesised that the increase in the vocabulary size over the year should also be positively correlated with the performance on the phonological tasks.

The results of the study – presented at the end of chapter 5 and discussed in chapter 6 – suggest that there is a relationship between phonological processing and foreign word learning. Specifically, the performance on the English phonological tasks correlated with the ability to memorise new word forms of a completely foreign language (LX) and of L2. It was hypothesised that the specific process tapped by English phonological tasks that facilitated this learning was initial prosodic processing – the ability allowing for the segmentation of foreign or noisy speech stream in the absence of ready-made phonological categories.

STRESZCZENIE

Celem poniższej pracy było zbadanie związków pomiędzy przetwarzaniem fonologicznym, fonologiczną pamięcią krótkotrwałą i uczeniem się nowych słów, zwłaszcza słów języka obcego. Krótkotrwała pamięć fonologiczna to rodzaj pamięci odpowiedzialny za krótkoterminowe przechowywanie informacji o charakterze werbalnym, konkretnie fonologicznej formy słów i zdań. Jak przedstawiono w rozdziale pierwszym, istnieją liczne badania wskazujące, że ów moduł pamięci ma kluczowe znaczenie dla uczenia się słów zarówno języka ojczystego, jak i obcego. Jak jednak wskazano pod koniec rozdziału, teoria wpływu fonologicznej pamięci krótkotrwałej na uczenie słów budzi kontrowersje. W literaturze zasugerowano, że związek pomiędzy wynikami niektórych zadań na krótkotrwałą pamięć fonologiczną, a postępami w uczeniu się słownictwa, może wynikać z wpływu innego czynnika - przetwarzania fonologicznego. Niestety, istnieje bardzo niewiele badań podejmujących tematykę związków pomiędzy przetwarzaniem fonologicznym a uczeniem się słownictwa, zwłaszcza w obcym języku. Celem poniższej pracy jest wypełnienie tej luki.

Aby zbadać wpływ przetwarzania fonologicznego na uczenie się słów, konieczne jest zdefiniowanie i zrozumienie tego terminu. Przetwarzanie fonologiczne jest różnie pojmowane przez różnych badaczy, toteż na potrzeby dysertacji konieczne było znalezienie jednolitej teorii definiującej ten mechanizm. Rozdział drugi dokumentuje poszukiwanie takiej teorii. Ta część pracy opisuje przetwarzanie fonologiczne z punktu widzenia fonologii, akustyki, psycholingwistyki i neurobiologii. Rozdział kończy się propozycją jednolitego podejścia do omawianego mechanizmu, opartego o idee zaczerpnięte z różnych dziedzin nauki. W myśl proponowanego podejścia, przetwarzanie

fonologiczne jest procesem rozpoznawania jednostek fonologicznym w strumieniu mowy. Proces ten wspomagany jest dodatkowymi mechanizmami segmentacyjnymi. Jednostki fonologiczne mają formę abstrakcyjnych kategorii tworzonych w oparciu o zbiór tokenów mowy.

Rozdział drugi kończy się zatem ustaleniem wstępnej definicji przetwarzania fonologicznego, tematem rozdziału trzeciego są natomiast związki pomiędzy przetwarzaniem fonologicznym, a uczeniem się nowych słów w języku ojczystym. W rozdziale tym pada propozycja, że rozwój przetwarzania fonologicznego u dzieci składa się z czterech różnych procesów: wstępnego przetwarzania prozodycznego, abstrakcji (tworzenia abstrakcyjnych kategorii fonologicznych), specyfikacji i sekwencjonowania ("chunking"). W rozdziale przedstawione są też liczne prace wskazujące, że każdy z tych procesów jest powiązany z rozwojem leksykalnym u dzieci uczących się pierwszego języka.

Rozdział czwarty poświęcony jest związkowi pomiędzy akwizycją fonologiczną i leksykalną w języku obcym, który to związek stanowi główny temat niniejszej dysertacji. Rozdział zaczyna się propozycją, że akwizycja fonologiczna w drugim języku jest pod wieloma względami podobna do akwizycji fonologicznej w pierwszym języku i że składa się z tych samych czterech procesów: wstępnego przetwarzania prozodycznego, abstrakcji, specyfikacji i sekwencjonowania. W związku z tym, że w języku ojczystym procesy te odgrywają zasadniczą rolę w uczeniu się nowych słów, w rozdziale pada hipoteza, iż taką samą rolę pełnią one i w języku obcym. Jednak, jak wskazuje przegląd literatury, do tej pory brakuje badań, które potwierdziły wpływ przetwarzania fonologicznego na uczenie się słów w drugim języku. Dlatego też rozdział kończy się przedstawieniem nowego eksperymentu, którego celem jest zbadanie tego wpływu.

Rozdział piąty poświęcony jest opisowi badania, którego uczestnikami jest 44 polskich dziewięciolatków uczących się języka angielskiego w szkole. Dzieci poproszone były o wykonanie kilku zadań fonologicznych w języku angielskim i polskim. Zadania ta miały na celu sprawdzenie ich rozwoju fonologicznego w obu językach. Dzieci wykonały też testy sprawdzające pojemność krótkoterminowej pamięci fonologicznej. Ponadto uczestnicy wzięli też udział w serii eksperymentów mierzących uczenie się nowych form słów ich języku ojczystym (polskim), drugim (angielskim) i w zupełnie obcym języku (LX). Przewidywano, że wyniki tych testów będą korelować z wynikami testów fonologicznych. Dzieciom biorącym udział w badaniu sprawdzono też zasób słownictwa drugiego języka na początku i na końcu roku szkolnego. Przewidywano bowiem, iż wzrost zasobu słownictwa w przeciągu roku szkolnego powinien także być powiązany z wynikami testów fonologicznych. Wyniki badania - przedstawione na końcu rozdziału piątego i omówione w rozdziale szóstym - potwierdzają istnienie związku pomiędzy przetwarzaniem fonologicznym a uczeniem się słów języka obcego. Wyniki testów fonologicznych w języku angielskim korelowały z umiejętnością zapamiętywania nowych słów języka zupełnie obcego (LX), jak również języka drugiego u dzieci w początkowych stadiach nauki. Przypuszcza się, że proces, który został uchwycony przez wyniki testów fonologicznych w języku angielskim i który miał wpływ na efektywność uczenia się słów, to wstępne przetwarzanie prozodyczne - umiejętność wspomagająca segmentację obcego lub zaszumionego strumienia mowy w sytuacji braku w pełni wykształconych kategorii fonologicznych.

References

Anderson, Stephen R. 1985. *Phonology in the twentieth century*. Chicago: University of Chicago Press.

Anthony, Jason L., Christopher J. Lonigan, Kimberly Driscoll, Beth M. Phillips and Stephen R. Burgess. 2003. "Phonological sensitivity: a quasi−parallel progression of word structure units and cognitive operations", *Reading Research Quarterly* 38, 4: 470–487.

Aoyama, Katsura, James E. Flege, Susan G. Guion, Reiko Akahane-Yamada and Tsuneo Yamada. 2004. "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/", *Journal of Phonetics* 32, 2: 233–250.

Aoyama, Katsura, Susan G. Guion and James E. Flege. 2008. "The first years in an L2-speaking environment: a comparison of Japanese children and adults learning American English", *International Review of Applied Linguistics* 46, 1: 61–90.

Atkinson, Richard C. and Richard M. Shiffrin. 1968. "Human memory: a proposed system and its control processes", in: Kenneth W. Spence and Janet Taylor Spence (eds.), *Psychology of learning and motivation*. New York: Academic Press, 89–195.

Baddeley, Alan D. 1966. "Short-term memory for word sequences as a function of acoustic, semantic and formal similarity", *Quarterly Journal of Experimental Psychology* 18, 4: 362–365.

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

Baddeley, Alan D. 2003. “Working memory and language: an overview”, *Journal of Communication Disorders* 36, 3: 189–208.

Baddeley, Alan D. 2007. *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, Alan D. and Graham J. Hitch. 1975. “Working memory”, in: Gordon H. Bower (ed.), *The psychology of learning and motivation*. New York: Academic Press, 47–90.

Baddeley, Alan D., Costanza Papagno and Giuseppe Vallar. 1988. “When long-term learning depends on short-term storage”, *Journal of Memory and Language* 27, 5: 586–595.

Baddeley, Alan D., Susan E. Gathercole and Constanza Papagno. 1998. “The phonological loop as a language learning device”, *Psychological Review* 105, 1: 158–173.

Bangasser, Debra A., David E. Waxler, Jessica Santollo and Tracey J. Shors. 2006. “Trace conditioning and the hippocampus: the importance of contiguity”, *The Journal of Neuroscience* 26, 34: 8702–8706.

Beattie, Rachel L. and Franklin R. Manis. 2014. “The relationship between prosodic perception, phonological awareness and vocabulary in emergent literacy”, *Journal of Research in Reading* 37, 2: 119–127.

Beckman, Mary E. and Jan Edwards. 2000. “The ontogeny of phonological categories and the primacy of lexical learning in linguistic development”, *Child Development* 71, 1: 240–249.

Beckman, Mary E., Benjamin Munson and Jan Edwards. 2007. “Vocabulary growth and the developmental expansion of types of phonological knowledge”, in: Jennifer Cole and Jose Ignacio Hualde (eds.), *Laboratory phonology*. Berlin: Mouton de Gruyter, 241–264.

Best, Cathrine T. 1995. “A direct realist view of cross-language speech perception”, in: Winifred Strange (ed.), *Speech perception and linguistic experience*. Timonium: New York Press, 171–204.

Best, Cathrine T. and Micheal D. Tyler. 2008. “Nonnative and second-language speech perception: Commonalities and complementarities”, in: Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language speech learning*. Amsterdam: John Benjamins Publishing Company, 13–34.

Bialystok, Ellen and Maria Fröhlich. 1978. "Variables of classroom achievement in second language learning", *The Modern Language Journal* 62, 7: 327–336.

Bloomfield, Leopold. [1933] 1962. *Language*. New York: Holt.

Boemio, Anthony, Stephen Fromm, Allen Braun and David Poeppel. 2005. "Hierarchical and asymmetric temporal sensitivity in human auditory cortices", *Nature Neuroscience* 8, 3: 389–395.

Bogdanowicz, Marta. 2009. "Nieznany język" [Unknown Language], in: *Dysleksja 3* [Dyslexia 3]. Warszawa: Pracownia Testów Psychologicznych.

Bohn, Ocke-Schwen and James E. Flege. 1990. "Interlingual identification and the role of foreign language experience in L2 vowel perception", *Applied Psycholinguistics* 11, 03: 303–328.

Bowey, Judith A. 1996. "On the association between phonological memory and receptive vocabulary in five-year-olds", *Journal of Experimental Child Psychology* 63, 1: 44–78.

Bowey, Judith A. 2001. "Nonword repetition and young children's receptive vocabulary: a longitudinal study", *Applied Psycholinguistics* 22, 03: 441–469.

Brackenbury, Tim and Marc E. Fey. 2003. "Quick incidental verb learning in 4-year-olds: identification and generalization", *Journal of Speech, Language, and Hearing Research* 46, 2: 313–327.

Bregman, Albert S. 1990. *Auditory scene analysis*. Cambridge, MA: MIT Press.

Bybee, Joan. 2002. "Phonological evidence for exemplar storage of multiword sequences", *Studies in Second Language Acquisition* 24, 02: 215–221.

Caplan, David and Gloria S. Waters. 1995. "On the nature of the phonological output planning processes involved in verbal rehearsal: evidence from aphasia", *Brain and Language* 48: 191–220.

Carroll, Susanne E. 2004. "Segmentation: learning how to 'hear words' in the L2 speech stream", *Transactions of the Philological Society* 102, 2: 227–254.

Carroll, Susanne E. 2006. The 'micro-structure' of a learning problem: prosodic prominence, attention, segmentation and word learning in a second language. (http://ling.uwo.ca/publications/CLA-ACL/CLA-ACL2005.htm) (date of access: 15 Mar. 2015).

Cheung, Him. 1996. "Nonword span as a unique predictor of second-language vocabulary language", *Developmental Psychology* 32, 5: 867–873.

Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Cohen, Neal J. and Larry R. Squire. 1980. "Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that", *Science* 210: 207–210.

Connine, Cynthia M. and Eleni Pinnow. 2006. "Phonological variation in spoken word recognition: episodes and abstractions", *The Linguistic Review* 23, 3: 235–245.

Conrad, R. and A. J. Hull. 1964. "Information, acoustic confusion and memory span", *British Journal of Psychology* 55, 4: 429–432.

Cowan, Nelson. 1988. "Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system", *Psychological Bulletin* 104, 2: 163–191.

Cowan, Nelson, Jeffrey N. Rouder, Christopher L. Blume and J. Scott Saults. 2012. "Models of verbal working memory capacity: what does it take to make them work?", *Psychological Review* 119, 3: 480–499.

Craik, Fergus F. I. 2002. "Levels of processing: past, present... and future?", *Memory* 10, 5-6: 305–318.

Craik, Fergus F. I. and Robert S. Lockhart. 1972. "Levels of processing: a framework for memory research", *Journal of Verbal Learning and Verbal Behavior* 11, 6: 671–684.

Curtin, Suzanne and Janet F. Werker. 2009. "The perceptual foundations of phonological development", in: M. Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 579–600.

Curtin, Suzanne, Toben H. Mintz and Morten H. Christiansen. 2005. "Stress changes the representational landscape: evidence from word segmentation", *Cognition* 96, 3: 233–262.

Cutler, Anne. 1990. "Exploiting prosodic probabilities in speech segmentation.", in: Gerry T. M. Altmann (ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives*. Cambridge: The MIT Press, 105–121.

Cutler, Anne, Jacques Mehler, Dennis Norris and Juan Segui. 1986. "The syllable's differing role in the segmentation of French and English", *Journal of Memory and Language* 25, 4: 385–400.

Cutler, Anne, Jacques Mehler, Dennis Norris and Juan Segui. 1992. “The monolingual nature of speech segmentation by bilinguals”, *Cognitive Psychology* 24, 3: 381–410.

Dahan, Delphine and James S. Magnuson. 2006. “Spoken word recognition”, in: Matthew J. Traxler and Morton A. Gernsbacher (eds.), *Handbook of Psycholinguistics*. London: Elsevier/Academic Press, 249–284.

Damasio, Antonio R. 1989. “Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition”, *Cognition* 33, 1-2: 25–62.

Davidson, Lisa. 2006. “Phonology, phonetics, or frequency: Influences on the production of non-native sequences”, *Journal of Phonetics* 34, 1: 104–137.

de Boysson-Bardies, Benedicte. 2001. *How language comes to children: from birth to two years*. Cambridge: MIT Press.

de Courtenay, Jan Niecisław Baudouin. 1894. *Próba teorii alternacyj fonetycznych* [An Attempt at a Theory of Phonetic Alternations]. Rozprawy Wydziału Filologicznego PAU.

de Jong, Peter F., Marie-Jose J. Seveke and Marjo van Veen. 2000. “Phonological sensitivity and the acquisition of new words in children”, *Journal of Experimental Child Psychology* 76, 4: 275–301.

de Saussure, Ferdinand. [1916] 2013. *Course in general linguistics*. New York: Columbia University Press.

Diehl, Randy. 2011. “On robustness of speech perception”, in: Lee Wai-Sum and Eric Zee (eds.), *Proceedings of the 17the International Congress of Phonetic Sciences*. Hong Kong: City Univeristy of Hong Kong, 1–8.

Díaz, Begoña, Cristina Baus, Carles Escera, Albert Costa and Núria Sebastián Gallés. 2008. “Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language”, *Proceedings of the National Academy of Sciences* 105, 42: 16083–16088.

Díaz, Begoña, Holger Mitterer, Mirjam Broersma and Núria Sebastián Gallés. 2012. “Individual differences in late bilinguals' L2 phonological processes: from acoustic-phonetic analysis to lexical access”, *Learning and Individual Differences* 22, 6: 680–689.

Donegan, Patricia J. and David Stampe. 1979. “The study of Natural Phonology”, in: D.A. Dinnsen, D.A (ed.), *Current approaches to phonological theory*. Bloomington: IUP, 126-173.

Donegan, Patricia and David Stampe. 2009. “Hypotheses of Natural Phonology”, *Poznań Studies in Contemporary Linguistics* 45, 1: 1–31.

Dressler, Wolfgang U. 1984. “Explaining Natural Phonology”, *Phonology* 1: 29–51.

Dressler, Wolfgang U. 1996. “Principles of naturalness in phonology and across components”, in: Bernhard Hurch and Richard Rhodes, *Natural Phonology: the state of the art*. Berlin: Walter de Gruyter, 41-54.

Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi and Jacques Mehler. 2001. “New evidence for prelexical phonological processing in word recognition”, *Language and Cognitive Processes* 16:5-6, 491-505.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier and Jacques Mehler. 1999. “Epenthetic vowels in Japanese: a perceptual illusion?”, *Journal of Experimental Child Psychology* 25, 6: 1568-1578.

Dysleksja 3. 2009. Warszawa: Pracownia Testów Psychologicznych.

Dziubalska-Kołaczyk, Katarzyna. 2002a. *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, Katarzyna. 2002b. “Challenges for Natural Linguistics in the twenty first century: a personal view”, *University of Hawai'i - Working Papers in Linguistics* 33: 15-39.

Dziubalska-Kołaczyk, Katarzyna. 2006. “Modern Natural Phonology: the theory for the future”, in: Jacek Fisiak (ed.), *English Language, Literature and Culture. Selected papers from the 13th PASE conference, Poznań 2004*. Poznań: Uni-Druk, 1–10.

Dziubalska-Kołaczyk, Katarzyna, Paulina Zydorowicz and Michał Jankowski. 2013. “English morphonotactics: a corpus study”, *The Phonetician. A Publication of ISPhS International Society of Phonetic Sciences* 107-108: 53–67.

Edwards, Jan, Mary E. Beckman and Benjamin Munson. 2004. “The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition”, *Journal of speech, language, and hearing research* 47, 2: 421–436.

Eimas, Peter D. and Joanne L. Miller. 1992. “Organization in the perception of speech by young infants”, *Psychological Science* 3, 6: 340–345.

Eimas, Peter D, Einar R. Siqueland, Peter Jusczyk and James Vigorito. 1971. “Speech Perception in Infants”, *Science* 171: 303–306.

Eiter, Brianna M. and Albrecht W. Inhoff. 2010. “Visual word recognition during reading is followed by subvocal articulation”, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, 2: 457–470.

Ellis, Nick C. 1996. “Sequencing in SLA: chunking, phonological memory and points of order”, *Studies in Second Language Acquisition* 18: 91–126.

Ellis, Nick C. 2001. “Memory for language”, in: Peter Robinson (ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 33-68.

Ellis, Nick C. 2008. “Constructions, chunking, and connectionism: the emergence of second language structure”, in: Catherine J. Doughty and Michael H. Long (eds.), *The handbook of second language acquisition*. Cambridge: Blackwell Publishing, 63-103.

Ellis, Nick C. and R. A. Hennelly. 1980. “A bilingual word-length effect: implications for intelligence testing and the relative ease of mental calculation in Welsh and English”, *British Journal of Psychology* 71, 1: 43–51.

Elman, James L. and Jeffrey L. McClelland. 1984. “The Trace model of speech perception”, *Cognitive Psychology* 18: 1–83.

Finn, Amy S. and Carla L. Hudson Kam. 2008. “The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation”, *Cognition* 108, 2: 477–499.

Flege, James E. 1987. “The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification”, *Journal of Phonetics* 15: 47–65.

Flege, James E. 1995. “Second language speech learning: theory, findings, and problems”, in: Winifred Strange (ed.), *Speech perception and linguistic experience*. Timonium: New York Press, 233–277.

Flege, James E. 2002. “Interactions between the native and second-language phonetic systems”, in: Petra Burmeister, Thorsten Piske and Andreas Rhode (eds.), *An integrated view of language development: papers in honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, 217–244.

Flege, James E., Grace H. Yeni-Komshian and Serena Liu. 1999. “Age constraints on second-language acquisition”, *Journal of Memory and Language* 41, 1: 78–104.

Fox, Robert A, .James E. Flege and Murray J. Munro. 1995. “The perception of English and Spanish vowels by native English and Spanish listeners: a multidimensional scaling analysis”, *The Journal of the Acoustical Society of America* 97, 4: 2540–2551.

Gabrieli, John D. 1998. “Cognitive neuroscience of human memory”, *Annual Review of Psychology* 49: 87–115.

Ganger, Jennifer and Michael R. Brent. 2004. “Reexamining the vocabulary spurt”, *Developmental Psychology* 40, 4: 621–632.

Ganong, William F. 1980. “Phonetic categorization in auditory word perception”, *Journal of Experimental Psychology: Human Perception and Performance* 6,1: 125.

Ganschow, Leonore and Richard Sparks. 1995. “Effects of direct instruction in Spanish phonology on the native-language skills and foreign-language aptitude of at-risk foreign-language learners”, *Journal of Learning Disabilities* 28, 2: 107–120.

Gardner, Robert C. and Wallace E. Lambert. 1972. *Attitudes and motivation in second-language learning*. Rowley: Newbury House Publishers.

Gaskell, M. Gareth and William D. Marslen-Wilson. 1997. “Integrating form and meaning: a distributed model of speech perception”, *Language and Cognitive Processes* 12, 5-6: 613–656.

Gathercole, Susan 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, Susan E. 1996. *Models of short-term memory*. Psychology Press.

Gathercole, Susan E. 2006. “Nonword repetition and word learning: the nature of the relationship”, *Applied Psycholinguistics* 27: 513-543.

Gathercole, Susan E. and Alan 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, 2: 200–213.

Gathercole, Susan E. and Alan D. Baddeley. 1990. “The role of phonological memory in vocabulary acquisition: a study of young children learning new names”, *British Journal of Psychology* 81, 4: 439–454.

Gathercole, Susan E. and Elvira V. Masoura. 2003. “Contrasting contributions of phonological short−term memory and long−term knowledge to vocabulary learning in a foreign language”, *Memory* 13, 3-4: 422–429.

Gathercole, Susan E., Catherine S. Willis, Alan D. Baddeley and Hazel Emslie. 1994. “The Children's Test of Nonword Repetition: a test of phonological working memory”, *Memory* 2, 2: 103–127.

Gathercole, Susan E., Catherine S. Willis, Hazel Emslie and Alan D. Baddeley. 1992. “Phonological memory and vocabulary development during the early school years: A longitudinal study”, *Developmental Psychology* 28, 5: 887–898.

Gathercole, Susan E., Claire Tiffany, Josie Briscoe, Annabel Thorn and 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, Susan E., Graham J. Hitch, Elisabet Service and Amanda J .Martin. 1997. “Phonological short-term memory and new word learning in children”, *Developmental Psychology* 33, 6: 966–979.

Girolametto, Luigi, Patsy Steig Pearce and Elaine Weitzman. 1997. “Effects of lexical intervention on the phonology of late talkers”, *Journal of Speech, Language, and Hearing Research* 40, 2: 338–348.

Goldfield, Beverly A. and J. Steven Reznick. 1990. “Early lexical acquisition: rate, content, and the vocabulary spurt”, *Journal of Child Language* 17, 01: 171–183.

Goldinger, Stephen D. 1998. “Echoes of echoes? An episodic theory of lexical access”, *Psychological Review* 105, 2: 251–279.

Goldsmith, John. 1976. “An overview of autosegmental phonology”, *Linguistic Analysis* 2: 23–68.

Gorman, Brenda K. 2012. “Relationships between vocabulary size, working memory, and phonological awareness in Spanish-speaking English language learners”, *American Journal of Speech-Language Pathology* 21, 2: 109–123.

Goswami, Usha, H.-L. Sharon Wang, Alicia Cruz, Tim Fosker, Natasha Mead and Martina Huss. 2011. “Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese”, *Journal of Cognitive Neuroscience* 23, 2: 325–337.

Goswami, Usha, Jennifer Thomson, Ulla Richardson, Rhona Stainthorp, Diana Hughes, Stuart Rosen and Sophie K. Scott. 2002. “Amplitude envelope onsets and developmental dyslexia: A new hypothesis”, *Proceedings of the National Academy of Sciences* 99, 16: 10911–10916.

Grossberg, Stephen. 2003. “Resonant neural dynamics of speech perception”, *Journal of Phonetics* 31, 3-4: 423–445.

Gupta, Prahlad and Brian MacWhinney. 1997. “Vocabulary acquisition and verbal short-term memory: computational and neural bases”, *Brain and Language* 59, 2: 267–333.

Gupta, Prahlad, Brian MacWhinney, Heidi M. Feldman and Kelley Sacco. 2003. “Phonological memory and vocabulary learning in children with focal lesions”, *Brain and Language* 87, 2: 241–252.

Halle, Morris. 1959. *The sound pattern of Russian. A linguistic and acoustical investigation.* The Hague: Mouton.

Hallé, Pierre A. and Bénédicte de Boysson-Bardies. 1996. “The format of representation of recognized words in infants' early receptive lexicon”, *Infant Behavior and Development* 19, 4: 463–481.

Hancin-Bhatt, Barbara and Rakesh M. Bhatt. 1997. “Optimal L2 syllables”, *Studies in Second Language Acquisition* 19: 331–378.

Hatch, Evelyn M. 1978. “Discourse analysis and second language acquisition”, in: Evelyn M. Hatch (ed), *Second language acquisition: a book of readings 2*. Rowley: Newbury House, 383-400.

Hawkins, Sarah. 2010a. “Phonetic variation as communicative system: perception of the particular and the abstract”, *Laboratory Phonology* 10: 1–19.

Hawkins, Sarah. 2010b. “Phonological features, auditory objects, and illusions”, *Journal of Phonetics* 38, 1: 60–89.

Hayes, Bruce. 1981. A metrical theory of stress rules. [Ph.D. dissertation, MIT].

Hazan, Valerie and Sarah Barrett. 2000. “The development of phonemic categorization in children aged 6–12”, *Journal of Phonetics* 28, 4: 377–396.

Henderson, A. I. and S. Nelms. 1980. “Relative salience of intonation fall and pause as cues to the perceptual segmentation of speech in an unfamiliar language”, *Journal of Psycholinguistic Research* 9, 2: 147–159.

Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Houston, Derek M. and Peter W. Jusczyk. 2000. "The role of talker-specific information in word segmentation by infants", *Journal of Experimental Psychology: Human Perception and Performance* 26, 5: 1570–1582.

Hu, Chieh-Fang. 2003. "Phonological memory, phonological awareness, and foreign language word learning", *Language Learning* 53, 3: 429–462.

Hu, Chieh-Fang. 2008. "Rate of acquiring and processing L2 color words in relation to L1 phonological awareness", *The Modern Language Journal* 92, 1: 39–52.

Hu, Chieh-Fang and C Melanie Schuele. 2005. "Learning nonnative names: The effect of poor native phonological awareness", *Applied Psycholinguistics* 26, 03: 343–362.

Imai, Satomi, Amanda C. Walley and James E. Flege. 2005. "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners", *The Journal of the Acoustical Society of America* 117, 2: 896.

Jakobson, Roman. [1939] 1980. *Child language aphasia and phonological universals*. (3rd edition.) The Hague: Mouton Publ.

Jakobson, Roman and Morris Halle. [1956] 1975. *Fundamentals of language*. (2nd edition.) The Hague: Mouton.

Jakoby, Hilla, Abraham Goldstein and Miriam Faust. 2011. "Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment", *Psychophysiology* 48, 11: 1517–1531.

Jarrold, Christopher, Annabel S C Thorn and Emma Stephens. 2009. "The relationships among verbal short-term memory, phonological awareness, and new word learning: evidence from typical development and Down syndrome", *Journal of Experimental Child Psychology* 102, 2: 196–218.

Jia, Gisela, Winifred Strange, Yanhong Wu, Julissa Collado and Qi Guan. 2006. "Perception and production of English vowels by Mandarin speakers: age-related differences vary with amount of L2 exposurea", *The Journal of the Acoustical Society of America* 119, 2: 1118.

Jones, Gary. 2011. "A computational simulation of children's performance across three nonword repetition tests", *Cognitive Systems Research* 12, 2: 113–121.

Jones, Gary. 2012. “Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed”, *Frontiers in Psychology* 3: 1-8.

Jones, Gary, Fernand Gobet and Julian M. Pine. 2007. “Linking working memory and long-term memory: a computational model of the learning of new words”, *Developmental Science* 10, 6: 853–873.

Jonides, John, Richard L. Lewis, Derek Evan Nee, Cindy A. Lustig, Marc G. Berman and Katherine Sledge Moore. 2008. “The mind and brain of short-term memory”, *Annual Review of Psychology* 59, 1: 193–224.

Jusczyk, Peter W. and Richard N. Aslin. 1995. “Infants′ detection of the sound patterns of words in fluent speech”, *Cognitive Psychology* 29, 1: 1–23.

Jusczyk, Peter W., Derek M. Houston and Mary Newsome. 1999. “The beginnings of word segmentation in English-learning infants”, *Cognitive Psychology* 39, 3-4: 159–207.

Kiefte, Michael and Keith R. Kluender. 2005. “The relative importance of spectral tilt in monophthongs and diphthongs”, *The Journal of the Acoustical Society of America* 117, 3: 1395–1404.

Kluender, Keith R. and Andrew J Lotto. 1999. “Virtues and perils of an empiricist approach to speech perception”, *The Journal of the Acoustical Society of America* 105, 1: 503–511.

Kluender, Keith R. and Michael Kiefte. 2006. “Speech Perception within a Biologically Realistic Information-Theoretic Framework “, in: Matthew J. Traxler and Morton A. Gernsbacher (eds.), *Handbook of Psycholinguistics*. London: Elsevier/Academic Press, 153–200.

Kuhl, Patricia K. 1991. “Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not”, *Perception & Psychophysics* 50, 2: 93–107.

Kuhl, Patricia K. 2004. “Early language acquisition: cracking the speech code”, *Nature Reviews Neuroscience* 5: 831-843.

Kuhl, Patricia K., Barbara T. Conboy, Denise Padden, Tobey Nelson and Jessica Pruitt. 2005. “Early speech perception and later language development: implications for the ‘Critical Period’”, *Language Learning and Development* 1, 3-4: 237–264.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens and Björn Lindblom. 1992. "Linguistic experience alters phonetic perception in infants by 6 months of age", *Science* 255, 5044: 606–608.

Kuhl, Patricia K., Barbara T. Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola and Tobey Nelson. 2008. "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)", *Philosophical Transactions of the Royal Society B* 363: 979–1000.

Kuhl, Patricia K., Rey R. Ramírez, Alexis Bosseler, Jo-Fu Lotus Lin and Toshiaki Imada. 2014. "Infants' brain responses to speech suggest analysis by synthesis", *Proceedings of the National Academy of Sciences of the United States of America* 111, 31: 11238–11245.

Larsen, Janet D. and Alan D. Baddeley. 2003. "Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: do they have a common source?", *The Quarterly Journal of Experimental Psychology Section A* 56, 8: 1249–1268.

Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler and Michael Studdert-Kennedy. 1967. "Perception of the speech code", *Psychological Review* 74, 6: 431–461.

Liberman, Mark Yoffe. 1975. The Intonational System of English. [Ph.D. dissertation, MIT].

Locke, John. 1986. "Speech perception and the emergent lexicon: an ethological approach", in: Paul Fletcher and Michael Garman, *Language Acquisition*. (2nd edition.) Cambridge: Cambridge University Press: 240-250.

Luce, Paul A. and David B. Pisoni. 1998. "Recognizing spoken words: the neighborhood activation model", *Ear and Hearing* 19, 1: 1–36.

Luce, Paul A., Stephen D. Goldinger, Edward T. Auer and Michael S. Vitevitch. 2000. "Phonetic priming, neighborhood activation, and PARSYN", *Attention, Perception, & Psychophysics* 62, 3: 615–625.

Lumley, Thomas and Alan Miller. 2004. Leaps: regression subset selection. R package version.

MacWhinney, Brian. 2004. "A unified model of language acquisition", Handbook of bilingualism.

Mandel, Denise R., Peter W. Jusczyk and David B. Pisoni. 1995. "Infants' recognition of the sound patterns of their own names", *Psychological Science* 6, 5: 314–317.

Marslen-Wilson, William D. 1987. "Functional parallelism in spoken word recognition", *Cognition* 25, 1-2: 71–102.

Marslen-Wilson, William D. and Alan Welsh. 1978. "Processing interactions and lexical access during word recognition in continuous speech", *Cognitive Psychology* 10, 1: 29–63.

Martin, Nancy A. and Rick Brownell. 2010. *Expressive One-Word Picture Vocabulary Test - 4th Edition (EOWPVT-4)*. Novato: Academic Therapy Publications.

Masoura, Elvira V. and Susan E. Gathercole. 1999. "Phonological short-term memory and foreign language learning", *International Journal of Psychology* 34, 5-6: 383–388.

Mattingly, Ignatius G., Alvin M. Liberman, Ann K. Syrdal and Terry Halwes. 1971. "Discrimination in speech and nonspeech modes", *Cognitive Psychology* 2, 2: 131–157.

Mattys, Sven L. 2004. "Stress versus coarticulation: toward an integrated approach to explicit speech segmentation", *Journal of Experimental Psychology: Human Perception and Performance* 30, 2: 397–408.

Mattys, Sven L, Laurence White and James F Melhorn. 2005. "Integration of multiple speech segmentation cues: a hierarchical framework", *Journal of Experimental Psychology: General* 134, 4: 477–500.

Mattys, Sven L., Peter W. Jusczyk, Paul A. Luce and James L. Morgan. 1999. "Phonotactic and prosodic effects on word segmentation in infants", *Cognitive Psychology* 38, 4: 465–494.

McCune, Lorraine and Marilyn M. Vihman. 2001. "Early phonetic and lexical development: a productivity approach", *Journal of Speech, Language, and Hearing Research* 44, 3: 670–684.

McQueen, James M. 1998. "Segmentation of continuous speech using phonotactics", *Journal of Memory and Language* 39, 1: 21–46.

Meador, Diane, James E. Flege and Ian R. A. MacKay. 2000. "Factors affecting the recognition of words in a second language", *Bilingualism: Language and Cognition* 3, 1: 55–67.

Meara, Paul. 1984. "The study of lexis in interlanguage", in: Alan Davies, C. Criper, & Anthony Howatt (eds.), *Interlanguage*. Edinburgh: University of Edinburgh Press, 225-235.

Melby-Lervåg, Monica, Arne Lervåg, Solveig-Alma Halaas Lyster, Marianne Klem, Bente Hagtvet and Charles Hulme. 2012. "Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development", *Psychological Science* 23, 10: 1092–1098.

Menn, Lise and Carol Stoel-Gammon. 1995. "Phonological development", in: Paul Fletcher and Brian MacWhinney (eds.), *The handbook of child language*. Oxford: Oxford University Press, 335–360.

Merriman, William E. and Joneen M. Schuster. 1991. "Young children's disambiguation of object name reference", *Child Development* 62: 1288–1301.

Metsala, Jamie L. 1999. "Young children's phonological awareness and nonword repetition as a function of vocabulary development", *Journal of Educational Psychology* 91, 1: 3–19.

Metsala, Jamie L. and Amanda C. Walley. 1998. "Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability", in: Jamie L. Metsala and Linnea C. Ehri (eds.), *Word recognition in beginning literacy*. New York: Routledge, 89–120.

Miller, George A. 1956. "The magical number seven, plus or minus two: some limits on our capacity for processing information", *Psychological Review* 63, 2: 81–97.

Miller, Joanne L and Peter D Eimas. 1995. "Speech perception: from signal to word", *Annual Review of Psychology* 46, 1: 467–492.

Milner, Brenda, Larry R. Squire and Eric R. Kandel. 1998. "Cognitive neuroscience and the study of memory", *Neuron* 20, 3: 445–468.

Milner, Brenda, Suzanne Corkin and H.-L. Teuber. 1968. "Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M.", *Neuropsychologia* 6, 3: 215–234.

Mirman, Daniel, James S. Magnuson, Katherine Graf Estes and James A. Dixon. 2008. "The link between statistical segmentation and word learning in adults", *Cognition* 108, 1: 271–280.

Mitton, Roger. 1992. "A description of a computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of Current English", http://coral. lili.

uni-bielefeld.
de/Classes/Winter98/ExPhon/SynthDBNotes/synthdbnotes/node18. html. (ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/info/CUVOALD_dictionary) (date of access: 10 Jan. 2012).

Morais, Jose, Luz Cary, Jesus Alegria and Paul Bertelson. 1979. "Does awareness of speech as a sequence of phones arise spontaneously?", *Cognition* 7, 4: 323–331.

Munson, Benjamin, Beth A. Kurtz and Jennifer Windsor. 2005. "The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment", *Journal of Speech, Language, and Hearing Research* 48, 5: 1033–1047.

Munson, Benjamin, Jan Edwards and Mary E. Beckman. 2011. "Phonological representations in language acquisition: climbing the ladder of abstraction", in: Abigail C. Cohn, Cécile Fougeron and Marie K. Huffman (eds.), The Oxford Handbook of Laboratory Phonology. Oxford: Oxford University Press, 288-309.

Murray, David J. 1967. "The role of speech responses in short-term memory", *Canadian Journal of Psychology/Revue canadienne de psychologie* 21, 3: 263-276.

Newell, Allen. 1994. *Unified Theories of Cognition*. Harvard University Press.

Norris, Dennis and Anne Cutler. 1985. "Juncture detection", *Linguistics* 23, 5: 689–705.

Norris, Dennis and James M. McQueen. 2008. "Shortlist B: A Bayesian model of continuous speech recognition", *Psychological Review* 115, 2: 357–395.

Norris, Dennis, James M. McQueen and Anne Cutler. 2003. "Perceptual learning in speech*", Cognitive Psychology* 47, 2: 204–238.

Oller, D. Kimbrough and Michelle L. Steffens. 1994. "Syllables and segments in infant vocalisations and young child speech", in: Mehmet S. Yavaş (ed.), *First and second language phonology*. San Diego, California: Singular Pub. Group Inc., 45–62.

Palladino, Paola and Marcella Ferrari. 2008. "Phonological sensitivity and memory in children with a foreign language learning difficulty", *Memory* 16, 6: 604–625.

Palmer, Stephen E. 1999. *Vision Science*. Bradford Books.

Papagno, Constanza and Giuseppe Vallar. 1995. "Verbal short-term memory and vocabulary learning in polyglots", *The Quarterly Journal of Experimental Psychology Section A* 48, 1: 98–107.

Pierrehumbert, Janet B. 2001. "Exemplar dynamics: word frequency, lenition and contrast.", in: Joan L. Bybee and Paul J. Hopper (eds.), *Frequency and emergence of linguistic structure*. Philadelphia: John Benjamins Publishing, 137–157.

Pierrehumbert, Janet B. 2003. "Probabilistic phonology: discrimination and robustness", in: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probabilistic linguistics*. Cambridge: Probabilistic linguistics, 175–228.

Piske, Thorsten, Ian R. A. MacKay and James E. Flege. 2001. "Factors affecting degree of foreign accent in an L2: a review", *Journal of Phonetics* 29, 2: 191–215.

Poeppel, David, William J. Idsardi and Virginie van Wassenhove. 2008. "Speech perception at the interface of neurobiology and linguistics", *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 363, 1493: 1071–1086.

Port, Robert. 2007a. "How are words stored in memory? Beyond phones and phonemes", *New Ideas in Psychology* 25, 2: 143–170.

Port, Robert. 2007b. Phonology is not psychological and speech processing is not linguistic. (Paper presented at Society for Philosophy and Psychology meeting, 14-16 June, 2007.).

Port, Robert. 2008. All is prosody: Phones and phonemes are the ghosts of letters. (Paper presented at the Prosody2008, 6-9 May 2008.).

Prince, Alan and Paul Smolensky. 2008. *Optimality Theory*. Oxford: John Wiley & Sons.

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski and Barbara Lewandowska-Tomaszczyk. 2012. *Narodowy korpus języka polskiego* [National Corpus of Polish Language]. Warszawa: Wydawnictwo Naukowe PWN.

Quené, Hugo. 1992. "Durational cues for word segmentation in Dutch", *Journal of Phonetics* 20, 3: 331–350.

Ramachandra, Vijayachandra, Lynne E. Hewitt and Tim Brackenbury. 2011. "The relationship between phonological memory, phonological sensitivity, and incidental word learning", *Journal of Psycholinguistic Research* 40, 2: 93–109.

Rast, Rebekah. 2010. "The role of linguistic input in the first hours of adult language learning", *Language Learning* 60, s2: 64–84.

Raven, John C., Teresa Szustrowa and Aleksandra Jaworowska. 2003. *Test Matryc Ravena w Wersji Kolorowej (TMK)* [Coloured Progressive Matrices]. Warszawa: Pracownia Testów Psychologicznych.

Read, Charles, Zhang Yun-Fei, Nie Hong-Yin and Ding Bao-Qing. 1986. "The ability to manipulate speech sounds depends on knowing alphabetic writing", *Cognition* 24, 1-2: 31–44.

Redford, Melissa A. and Randy L. Diehl. 1999. "The relative perceptual distinctiveness of initial and final consonants in CVC syllables", *The Journal of the Acoustical Society of America* 106, 3: 1555–1565.

Remez, Robert E. 2008. "Perceptual organization of speech", in: David Pisoni and Robert Remez (eds.), *The Handbook of Speech Perception*. Cambridge: John Wiley & Sons, 28-50.

Remez, Robert E., Philip E. Rubin, Stefanie. M Berns, Jennifer S. Pardo and Jessica M. Lang. 1994. "On the perceptual organization of speech", *Psychological Review* 101, 1: 129–156.

Repovs, Grega and Alan D. Baddeley. 2006. "The multi-component model of working memory: Explorations in experimental cognitive psychology", *Neuroscience* 139, 1: 5–21.

Richardson, Ulla, Jennifer M. Thomson, Sophie K. Scott and Usha Goswami. 2004. "Auditory processing skills and phonological representation in Dyslexic children", *Dyslexia* 10, 3: 215–233.

Saffran, Jenny R. 2001. "Words in a sea of sounds: The output of infant statistical learning", *Cognition* 81, 2: 149–169.

Saffran, Jenny R., Elissa L. Newport and Richard N. Aslin. 1996a. "Word segmentation: the role of distributional cues", *Journal of Memory and Language* 35, 4: 606–621.

Saffran, Jenny R., Richard N. Aslin and Elissa L. Newport. 1996b. "Statistical learning by 8-month-old infants", *Science* 274: 1926-1928.

Salverda, Anne Pier, Delphine Dahan and James M. McQueen. 2003. "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension", *Cognition* 90, 1: 51–89.

Samuel, Arthur G. and Mark A. Pitt. 2003. “Lexical activation (and other factors) can mediate compensation for coarticulation”, *Journal of Memory and Language* 48, 2: 416–434.

Sebastian-Galles, Núria and Cristina Baus. 2005. “On the relationship between perception and production in L2 categories”, in: Anne Cutler (ed.), *Twenty-first century psycholinguistics: four cornerstones*. Mahwah: Lawrence Erlabum Associates, 279–292.

Sebastián Gallés, Núria and Begoña Díaz. 2012. “First and second language speech perception: graded learning*”, Language Learning* 62, s2: 131–147.

Selkirk, Elisabeth. 1984. “On the major class features and syllable theory”, in: Mark Aronoff and Richard T. Oerle, *Language sound structure. Studies in phonology presented to Morris Halle by his teacher and students.* Cambridge: The MIT Press, 107-136.

Service, Elisabet. 1992. “Phonology, working memory, and foreign-language learning”, *The Quarterly Journal of Experimental Psychology Section A* 45, 1: 21–50.

Service, Elisabet and Viljo Kohonen. 1995. “Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition?”, *Applied Psycholinguistics* 16, 02: 155–172.

Shebani, Mustafa F.A., Fons J. R. Van De Vijver and Ype H. Poortinga. 2005. “A strict test of the phonological loop hypothesis with Libyan data”, *Memory & Cognition* 33, 2: 196–202.

Shoemaker, Ellenor and Rebekah Rast. 2013. “Extracting words from the speech stream at first exposure”, *Second Language Research* 29, 2: 165–183.

Singh, Leher, James L. Morgan and Katherine S. White. 2004. “Preference and processing: The role of speech affect in early spoken word recognition”, *Journal of Memory and Language* 51, 2: 173–189.

Smith, Bruce L., Karla K. McGregor and Darcie Demille. 2006. “Phonological development in lexically precocious 2-year-olds”, *Applied Psycholinguistics* 27, 03: 355–375.

Smits, Roel, Natasha Warner, James M. McQueen and Anne Cutler. 2003. “Unfolding of phonetic information over time: a database of Dutch diphone perception”, *The Journal of the Acoustical Society of America* 113, 1: 563.

Snowling, Maggie, Shula Chiat and Charles Hulme. 1991. "Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley", *Applied Psycholinguistics* 12, 03: 369–373.

Sobkowiak, Włodzimierz. 2006. Phonetics of EFL dictionary definitions. Poznań: Wydawnictwo Poznańskie.

Speciale, Giovanna, Nick C. Ellis and Tracey Bywater. 2004. "Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition", *Applied Psycholinguistics* 25, 02: 293–321.

Stoel-Gammon, Carol. 2011. "Relationships between lexical and phonological development in young children", *Journal of Child Language* 38, 01: 1–34.

Storkel, Holly L. 2001. "Learning new words: phonotactic probability in language development", *Journal of Speech, Language, and Hearing Research* 44, 6: 1321–1337.

Storkel, Holly L. 2002. "Restructuring of similarity neighbourhoods in the developing mental lexicon", *Journal of Child Language* 29, 02: 251–274.

Storkel, Holly L. 2004. "The emerging lexicon of children with phonological delays: phonotactic constraints and probability in acquisition", *Journal of Speech, Language, and Hearing Research* 47: 1194–1212.

Storkel, Holly L. and Michele L. Morrisette. 2002. "The lexicon and phonology: interactions in language acquisition", *Language, Speech, and Hearing Services in Schools* 33, 1: 24–37.

Suprenant, Aimee M. and Ian Neath. 2009. "The nine lives of short-term memory", in: Annabel S. C. Thorn and Mike Page (eds.), *Interaction between short-term and long-term memory in the verbal domain*. Hove: Psychology Press, 16-43.

Szczerbiński, Marcin and Olga Pelc-Pękała. 2009. "Usuwanie fonemów dla klasy III" [Phoneme deletion for third-graders], in: *Dysleksja 3* [Dyslexia 3]. Warszawa: Pracownia Testów Psychologicznych.

Trubetzkoy, Nikolai S. [1939] 1970. *Podstawy fonologii* [Principles of phonology]. (Translated by Adam Heinz.) Warszawa: Państwowe Wydawnictwo Naukowe.

Tulving, Endel. 1985. *Elements of Episodic Memory*. Oxford: Oxford University Press.

Van der Hulst, Harry and Norval Smith. 1982. "An overview of Autosegmental and Metrical Phonology", in: Harry Van der Hulst and Norval Smith (eds.), *The structure of phonological representations*. Dordrecht: Foris, 2–45.

Vihman, Marilyn M., Shelley L. Velleman and Lorraine McCune. 1994. "How abstract is child's phonology? Toward an integration of linguistic and psychological approaches", in: Mehmet S Yavaş (ed.), *First and second language phonology*. San Diego: Singular Pub. Group Inc., 9–44.

Vihman, Marilyn May. 1996. *Phonological development*. Cambridge: Blackwell Publishing.

Vihman, Marylin M., Satsuki Nakai, Rory A. DePaolis and Pierre Hallé. 2004. "The role of accentual pattern in early lexical representation", *Journal of Memory and Language* 50, 3: 336–353.

Wagner, Richard K. and Joseph K. Torgesen. 1987. "The nature of phonological processing and its causal role in the acquisition of reading skills", *Psychological Bulletin* 101, 2: 192–212.

Wagner, Richard K., Joseph K. Torgesen, Carol Alexander Rashotte and Nick A. Pearson. 2013. *Comprehensive Test of Phonological Processing-Second Edition: CTOPP-2*. Austin: PRO-ED.

Walley, Amanda C. 2008. "Speech learning, lexical reorganization, and the development of word recognition by native and non-native English speakers", in: Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language speech learning*. Amsterdam: John Benjamins Publishing Company, 315–330.

Wechsler, David. 1974. *Manual for the Wechsler Intelligence Scale for Children*. (Rev.) New York: The Psychological Corporation.

Werker, Janet F. and H. Henny Yeung. 2005. "Infant speech perception bootstraps word learning", *Trends in Cognitive Sciences* 9,11: 519-527.

Werker, Janet F. and Richard C. Tees. 1984. "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life", *Infant Behavior and Development* 7, 1: 49–63.

Werker, Janet F. and Suzanne Curtin. 2005. "PRIMIR: a developmental framework of infant speech processing", *Language Learning and Development* 1, 2: 197–234.

Werker, Janet F., Christopher T. Fennell, Kathleen M. Corcoran and Christine L. Stager. 2002. "Infants' ability to learn phonetically similar words: effects of age and vocabulary zize", *Infancy* 3, 1: 1–30.

Windfuhr, Kirsten L. and Margaret J. Snowling. 2001. “The relationship between paired associate learning and phonological skills in normally developing readers”, *Journal of Experimental Child Psychology* 80, 2: 160–173.

Zhang, Guojun and Herbert A. Simon. 1985. “STM capacity for Chinese words and idioms: chunking and acoustical loop hypotheses”, *Memory & Cognition* 13, 3: 193–201.

Zhang, Yang and Yue Wang. 2007. “Neural plasticity in speech acquisition and learning”, *Bilingualism: Language and Cognition* 10, 02: 147–14.

Ziegler, Johannes C. and Usha Goswami. 2005. “Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory”, *Psychological Bulletin* 131, 1: 3–29.

Appendix A

Parental questionnaire

Imię i nazwisko dziecka

Data urodzin dziecka (dzień, miesiąc i rok)

1. Kiedy dziecko zaczęło uczyć się języka angielskiego?

miesiąc:____________________________rok:__________________

2. Czy dziecko uczy się obecnie języka angielskiego lub uczyło się tego języka również poza szkołą?

a) tak

b) nie (proszę przejść od razu do pytania 4)

3. Dla każdego kursu/cyklu zajęć prywatnych, na które dziecko uczęszczało poza szkołą proszę podać informacje:

Cykl zajęć 1

a) Kiedy dziecko zaczęło uczęszczać na zajęcia?

miesiąc:____________________________rok:__________________

b) Kiedy przestało uczęszczać na zajęcia? (jeśli nie dotyczy, proszę zamiast daty postawić krzyżyk)

miesiąc:____________________________rok:__________________

c) Jak często odbywały/odbywają się zajęcia i ile trwały/trwają?

Liczba zajęć w tygodniu: __________________

Długość zajęć (w minutach): ____________________

d) Jaką formę mają/miały zajęcia?

Indywidualne zajęcia z nauczycielem

Zajęcia w szkole językowej

Inną (jaką?)____

e) Czy nauczyciel, który uczy/uczył dziecka poza szkołą pochodzi z kraju anglojęzycznego (Wielkiej Brytanii, Stanów Zjednoczonych lub Australii)?

tak

nie

nie wiem

Cykl zajęć 2

a) Kiedy dziecko zaczęło uczęszczać na zajęcia?

miesiąc:____________________________rok:__________________

b) Kiedy przestało uczęszczać na zajęcia? (jeśli nie dotyczy, proszę zamiast daty postawić krzyżyk)

miesiąc:____________________________rok:__________________

c) Jak często odbywały/odbywają się zajęcia i ile trwały/trwają?

Liczba zajęć w tygodniu: __________________

Długość zajęć (w minutach): ____________________

d) Jaką formę mają/miały zajęcia?

Indywidualne zajęcia z nauczycielem

Zajęcia w szkole językowej

Inną (jaką?)____

e) Czy nauczyciel, który uczy/uczył dziecka poza szkołą pochodzi z kraju anglojęzycznego (Wielkiej Brytanii, Stanów Zjednoczonych lub Australii)?

tak

nie

nie wiem

Cykl zajęć 3

a) Kiedy dziecko zaczęło uczęszczać na zajęcia?

miesiąc:____________________________rok:__________________

b) Kiedy przestało uczęszczać na zajęcia? (jeśli nie dotyczy, proszę zamiast daty postawić krzyżyk)

miesiąc:____________________________rok:__________________

c) Jak często odbywały/odbywają się zajęcia i ile trwały/trwają?

Liczba zajęć w tygodniu: __________________

Długość zajęć (w minutach): ____________________

d) Jaką formę mają/miały zajęcia?

Indywidualne zajęcia z nauczycielem

Zajęcia w szkole językowej

Inną (jaką?)____

e) Czy nauczyciel, który uczy/uczył dziecka poza szkołą pochodzi z kraju anglojęzycznego (Wielkiej Brytanii, Stanów Zjednoczonych lub Australii)?

tak

nie

nie wiem

4. Czy dziecko kiedykolwiek przebywało w kraju anglojęzycznym (Wielkiej Brytanii, Australii, Stanach Zjednoczonych)?

a) tak (nazwa kraju____)

b) nie

5. Jaki był cel wyjazdu? (kurs językowy, wizyta rodzinna, wycieczka, wakacje – jeżeli wyjazdów było kilka, proszę wymienić je po przecinku)

6. Jak długo dziecko przebywało w kraju anglojęzycznym? (jeżeli wyjazdów było kilka, proszę wymienić je po przecinku)

7. Czy dziecko uczy się lub uczyło innych języków obcych?

a) tak

b) nie

8. Proszę wymienić języki obce (oprócz angielskiego), których uczyło się dziecko i jak długo dziecko się ich uczyło:

a) język 1:____

data rozpoczęcia nauki:____

data zakończenia nauki:____

b) język 2:____

data rozpoczęcia nauki:____

data zakończenia nauki:____

c) język 3:____

data rozpoczęcia nauki:____

data zakończenia nauki:____

9. Czy dziecko kiedykolwiek przebywało obcym kraju nieanglojęzycznym dłużej niż 3 tygodnie?

a) tak (nazwa kraju____)

b) nie

10. Jaki był cel wyjazdu? (kurs językowy, wizyta rodzinna, wycieczka, wakacje – jeżeli wyjazdów było kilka, proszę wymienić je po przecinku)

11. Jak długo dziecko przebywało w obcym kraju? (jeżeli wyjazdów było kilka, proszę wymienić je po przecinku)

12. Ile lat miało dziecko, kiedy zaczęło wypowiadać się pełnymi zdaniami?

13. Ile miesięcy miało dziecko, kiedy zaczęło wypowiadać pierwsze słowa?

14. Czy dziecko w przeszłości przechodziło zapalenia ucha? Jeśli tak, to kiedy i jak długo trwała infekcja?

15. Czy dziecko ma lub kiedykolwiek miało problemy ze słuchem?

16. Czy dziecko kiedykolwiek korzystało z pomocy logopedy? Jeśli tak, to w jakim wieku, jaki był powód interwencji i jak długo ona trwała?

17. Czy u dziecka zdiagnozowano kiedykolwiek dysleksję lub dysgrafię?

Appendix B

Handbook Vocabulary test

1. dog
2. ball
3. cat
4. egg
5. fish
6. T-shirt
7. swim
8. computer
9. dress
10. grandma
11. ice-cream
12. jump
13. chair
14. hat/cap
15. horse
16. house/home
17. salad
18. kitchen
19. run
20. bird
21. mouth/lips
22. leg
23. chicken
24. eye
25. ear
26. fly
27. sun/sunny
28. bathroom/toilet
29. princess
30. food
31. animals/pets
32. spaghetti/pasta/macaroni
33. play football
34. cake
35. living room
36. rain
37. walk
38. eat
39. read
40. sing
41. write
42. paint/draw

Appendix C

Stimuli used in the novel word learning tasks

Polish word learning task:

taśmy /taɕmɨ/ - matka /matka/ [tapes-mother]

tarka /tarka/ - bramy /bramɨ/[grater-gates]

beczki /bet͡ʃki/ - maska /maska/ [barrels-mask]

L1 non-word learning task:

Miska (a bowl) - torka

Graty (things) - jaczki

Polny (field) - wyski

POLISH transcription of non-words	ENGLISH transcription of non-words
tɔrka	torkɑ
jat͡ʃki	jʌt͡ʃki/jɑt͡ʃki
wɨski	vɪski

L2 non-word learning task:

Plany (plans) - muckly

Trawa (grass) - lesty

Kraty (bars) – stamy

POLISH transcription of non-words	ENGLISH transcription of non-words
stemi/stami	stæmi
lesti	lesti
makli	mʌkli

LX non-word learning task:

Sanki (sledge) - bnijo

Gruby (fat) - sznedży

Paczki (packages) – żiwnu

POLISH transcription of non-words	ENGLISH transcription of non-words
szned͡ʒɨ	ʃned͡ʒɪ
ʒiwnu	ʒiːwnuː
bnijɔ	bniːjo

Appendix D

Stimuli used in the phonological STM tasks

L1 ISR

0a: wycka

0b: traby

1a: przyby – zacja

1b: rejna – spuwa

2a: zaska – belna – właby

2b: plawa – wjedy – tocka

3a: prody – taska – zalny – jeczki

3b: tawka – byska – przywa – krała

4a: przeby – miczki – trady – wasta – reska

4b: przyna – traje – Pronia – wycka – traby

L2 ISR

0a: proki

0b: blanuh

1a: denduh – kosty

1b: bintuh – proty

2a: brekuh – kreeluh – seetry

2b: stiny – dilty – beentuh

3a: rintuh – freny – brenuh – mekly

3b: stimuh – rikly – brety – prinuh

4a: preluh – rinduh – proby – blatuh – rilty

4b: blanuh – lesty – muckly – stamy – proki

LX ISR

0a: dżapsa

0b: żimro

1a: dżapłu – bzajo

1b: bnidżo – chłedzu

2a: dżuśmo – szorłu – żispo

2b: gjudża – bzopo – żipsa

3a: bzofu – chłeszy – dżumro – żirłu

3b: dżogłu – bzodży – żiwjo – bnidzu

4a: dżofjo – gjuszi – żikłu – bzadzu – chłedżo

4b: dżeplo – gjufo – źrudzu – dżakłu – bzoszi

Appendix E

Raw data at T1

ID	AGE	Year when started English classes	Additional English classes?	Raven raw scoers	Vocabulary EOWPVT	Vocabulary Handbook	Word-learning Polish words	Word-learning L1 non-words	Word-learning L2 non-words	Word-learning LX non-words	Digit span
1	101	2009	0	31	8	20	39	37	36	36	9
2	113	2011	0	33	10	29	41	40	36	39	10
3	112	2007	1	23	14	34	34	32	28	15	10
4	117	2009	0	27	3	18	12	18	17	2	8
5	110	2011	1	27	23	25	42	22	19	21	12
6	119	2008	0	30	44	38	44	45	42	43	16
7	102	2009	1	30	18	31	24	29	34	13	8
8	110	2009	0	28	15	22	42	38	30	39	10
9	101	2009	0	29	10	25	32	24	30	26	10
10	101	2009	0	26	14	21	37	36	34	16	9
11	113	2007	1	28	11	13	31	40	24	22	7
12	99	2011	0	26	2	11	44	16	28	28	9
13	98	2009	0	26	14	19	42	41	28	35	10
14	104	2010	0	28	22	25	40	42	28	27	9

15	117	2011	0	32	23	32	38	41	31	25	11
16	103	2010	0	29	3	19	41	41	23	17	9
17	118	2008	0	26	10	20	37	42	22	9	12
18	110	2011	0	33	12	22	40	37	34	26	12
19	117	2011	0	34	19	30	42	40	40	36	13
20	104	2010	0	25	0	0	41	41	10	7	11
21	110	2007	0	30	16	27	41	43	39	27	6
22	105	2007	0	34	26	33	36	43	29	33	11
23	102	2010	1	34	24	31	22	28	33	18	10
24	120	2011	0	28	5	19	37	24	26	11	6
25	116	2009	0	27	2	13	38	41	38	34	10
26	118	2007	1	31	9	22	33	27	22	14	9
27	119	2010	0	26	3	6	23	8	4	3	7
28	114	2007	1	32	14	22	42	37	15	19	7
29	111	2010	0	27	7	16	42	29	27	12	8
30	108	2009	1	28	6	12	41	27	29	22	11
31	128	2011	1	35	10	31	32	40	35	25	9
32	113	2011	1	31	19	26	38	36	30	21	11
33	109	2011	0	25	9	16	27	31	23	24	8
34	106	2011	0	31	12	19	43	35	32	26	13
35	110	2010	0	23	11	28	34	33	33	23	10
36	120	2010	1	24	16	23	25	43	29	36	8
37	105	2009	1	31	12	27	44	34	36	16	7
38	104	2010	0	35	4	17	39	40	26	3	13
39	111	2011	0	30	6	12	40	30	22	18	5
40	115	2007	0	25	9	13	30	28	25	9	10

41	110	2007	0	36	23	31	38	42	31	37	11
42	109	2008	1	30	3	11	42	30	16	6	10
43	112	2011	0	29	14	23	39	32	26	24	4
44	115	2008	0	28	17	29	20	33	31	7	8

ID	**ISR L1 non-words**	**ISR L2 non-words**	**ISR LX non-words**	**Polish elision**	**Polish blending**	**Polish non-word blending**	**Polish non-word analysis**	**English elision**	**English blending**	**English non-word blending**
1	5	4	4	19	20	6	11	26	16	18
2	6	3	5	22	13	9	13	28	20	14
3	8	6	6	23	25	9	13	28	23	23
4	6	3	4	17	22	10	13	14	17	19
5	5	3	4	23	27	9	13	31	24	20
6	9	6	5	23	27	12	13	33	25	18
7	5	4	3	18	27	8	12	28	21	16
8	7	4	5	20	32	11	13	15	19	8
9	7	6	4	23	29	13	13	16	24	18
10	6	5	4	23	14	6	13	26	20	11
11	6	5	4	20	20	6	10	0	15	5
12	5	3	3	22	14	8	11	26	18	13
13	6	4	4	16	20	9	10	14	15	13
14	8	5	4	23	28	8	13	27	20	16
15	6	5	5	14	11	12	13	28	27	25
16	7	5	6	21	24	8	13	28	17	15
17	4	4	3	23	23	8	13	24	20	18
18	8	6	5	17	19	9	12	15	14	14

19	6	4	4	23	28	12	10	30	23	20
20	8	3	4	17	20	8	13	13	16	16
21	6	3	3	23	25	6	11	33	22	16
22	6	6	4	23	29	10	13	33	25	24
23	7	5	5	23	30	13	12	32	25	25
24	5	3	4	19	18	5	10	2	7	1
25	5	4	4	21	24	7	13	32	21	17
26	8	5	4	17	17	6	10	21	7	3
27	6	5	4	17	15	5	11	0	7	7
28	6	4	4	22	22	7	13	26	19	15
29	6	5	4	22	23	8	13	19	15	19
30	6	6	4	22	31	9	13	33	21	19
31	3	6	4	22	20	8	11	25	20	7
32	7	5	4	21	26	7	12	31	20	13
33	6	4	3	19	25	10	12	19	24	20
34	7	6	4	23	22	9	13	31	23	18
35	7	5	4	22	29	9	13	15	21	17
36	5	5	4	16	25	9	13	11	14	15
37	5	5	4	22	27	7	13	17	19	13
38	4	3	3	22	27	7	13	0	14	2
39	5	4	3	20	21	7	11	24	6	16
40	4	4	3	21	21	8	13	28	14	17
41	6	6	4	23	31	12	13	33	21	19
42	6	4	3	18	24	10	12	14	18	2
43	8	4	4	8	18	6	4	19	19	16
44	6	4	4	15	19	3	9	12	16	13

Appendix F

Raw data T2

ID	AGE	Year when started English classes	Raven raw scores	Vocabulary EOWPVT T1	Vocabulary Handbook T1	Vocabulary EOWPVT T2	Vocabulary Handbook T2	Word-learning Polish words	Word-learning L1 non-words
1	101	2009	31	8	20	11	23	39	37
2	113	2011	33	10	29	14	33	41	40
4	117	2009	27	3	18	6	15	12	18
6	119	2008	30	44	38	48	37	44	45
8	110	2009	28	15	22	9	16	42	38
9	101	2009	29	10	25	11	29	32	24
12	99	2011	26	2	11	4	18	44	16
13	98	2009	26	14	19	15	27	42	41
14	104	2010	28	22	25	23	27	40	42
15	117	2011	32	23	32	29	29	38	41
16	103	2010	29	3	19	8	18	41	41
17	118	2008	26	10	20	10	23	37	42
18	110	2011	33	12	22	15	28	40	37

19	117	2011	34	19	30	25	36	42	40
20	104	2010	25	0	0	0	1	41	41
22	105	2007	34	26	33	45	39	36	43
24	120	2011	28	5	19	7	17	37	24
25	116	2009	27	2	13	10	18	38	41
27	119	2010	26	3	6	3	10	23	8
28	114	2007	32	14	22	17	31	42	37
29	111	2010	27	7	16	10	16	42	29
33	109	2011	25	9	16	11	21	27	31
34	106	2011	31	12	19	15	25	43	35
35	110	2010	23	11	28	13	31	34	33
37	105	2009	31	12	27	13	28	44	34
38	104	2010	35	4	17	9	23	39	40
39	111	2011	30	6	12	6	15	40	30
41	110	2007	36	23	31	14	33	38	42
43	112	2011	29	14	23	25	29	39	32
44	115	2008	28	17	29	26	35	20	33

ID	**Word-learning L2 non-words**	**Word-learning LX non-words**	**Digit span T1**	**ISR L1 non-words T1**	**ISR L2 T1**	**ISR LX T1**	**Digit span T2**	**ISR L1 T2**	**ISR L2 T2**	**ISR LX T2**
1	36	36	9	5	4	4	9	6	4	4
2	36	39	10	6	3	5	10	8	4	4
4	17	2	8	6	3	4	8	5	3	4
6	42	43	16	9	6	5	18	9	8	5
8	30	39	10	7	4	5	10	8	4	5

9	30	26	10	7	6	4	12	7	4	4
12	28	28	9	5	3	3	11	6	4	3
13	28	35	10	6	4	4	9	6	4	4
14	28	27	9	8	5	4	9	7	4	5
15	31	25	11	6	5	5	13	7	7	3
16	23	17	9	7	5	6	10	7	5	5
17	22	9	12	4	4	3	9	6	4	4
18	34	26	12	8	6	5	13	8	4	4
19	40	36	13	6	4	4	10	5	3	4
20	10	7	11	8	3	4	10	6	6	5
22	29	33	11	6	6	4	13	6	4	4
24	26	11	6	5	3	4	10	4	2	2
25	38	34	10	5	4	4	12	6	4	3
27	4	3	7	6	5	4	10	5	4	2
28	15	19	7	6	4	4	9	5	3	3
29	27	12	8	6	5	4	7	6	4	3
33	23	24	8	6	4	3	8	6	4	4
34	32	26	13	7	6	4	11	6	6	4
35	33	23	10	7	5	4	12	7	5	4
37	36	16	7	5	5	4	9	4	4	2
38	26	3	13	4	3	3	6	4	3	2
39	22	18	5	5	4	3	6	5	4	2
41	31	37	11	6	6	4	11	5	4	3
43	26	24	4	8	4	4	7	6	3	2
44	31	7	8	6	4	4	9	6	4	3

ID	Polish elision T1	Polish blending T1	Polish non-word blending T1	Polish non-word analysis T1	Polish elision T2	Polish blending T2	Polish non-word blending T2	Polish non-word analysis T2	English elision T1	English blending T1	English non-word blending T1	English elision T2	English blending T2	English non-word blending T2
1	19	20	6	11	23	23	6	12	26	16	18	29	21	15
2	22	13	9	13	21	24	12	13	28	20	14	31	21	16
4	17	22	10	13	17	25	9	13	14	17	19	16	17	16
6	23	27	12	13	23	28	13	13	33	25	18	33	27	23
8	20	32	11	13	20	28	12	13	15	19	8	16	23	22
9	23	29	13	13	23	27	12	13	16	24	18	24	25	24
12	22	14	8	11	16	20	7	13	26	18	13	20	20	14
13	16	20	9	10	20	25	8	13	14	15	13	28	18	16
14	23	28	8	13	21	27	11	13	27	20	16	31	31	16
15	14	11	12	13	23	28	11	13	28	27	25	30	26	28
16	21	24	8	13	21	27	9	13	28	17	15	33	22	17
17	23	23	8	13	23	30	9	13	24	20	18	23	24	20
18	17	19	9	12	18	21	10	13	15	14	14	17	17	16
19	23	28	12	10	22	27	10	12	30	23	20	33	24	18
20	17	20	8	13	23	26	7	12	13	16	16	29	22	21
22	23	29	10	13	22	31	11	13	33	25	24	33	30	25
24	19	18	5	10	17	23	6	11	2	7	1	26	12	4
25	21	24	7	13	23	27	11	13	32	21	17	32	20	16
27	17	15	5	11	19	23	4	8	0	7	7	1	15	8
28	22	22	7	13	22	27	7	13	26	19	15	15	21	18
29	22	23	8	13	23	24	7	13	19	15	19	32	20	14
33	19	25	10	12	17	27	13	13	19	24	20	20	23	21
34	23	22	9	13	21	26	9	13	31	23	18	16	21	17

35	22	29	9	13	23	23	11	13	15	21	17	32	23	23
37	22	27	7	13	22	27	10	13	17	19	13	23	18	19
38	22	27	7	13	14	19	6	8	0	14	2	12	19	10
39	20	21	7	11	23	25	8	11	24	6	16	12	15	13
41	23	31	12	13	23	29	12	13	33	21	19	32	22	23
43	13	18	6	4	18	27	8	12	19	19	16	8	22	12
44	15	19	3	9	14	19	5	11	12	16	13	14	16	14