Phonotactics and Its Acquisition, Representation, and Use

An Experimental-Phonological Study

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: *Phonotactics*, by Artjom Avetisyan

ISBN: 978-94-6093-080-5
NUR 616

Copyright © 2012: Natalie Boll-Avetisyan. All rights reserved.

Phonotactics and Its Acquisition, Representation, and Use

An Experimental-Phonological Study

Fonotactische regels en hun verwerving, representatie en gebruik

Een experimenteel-fonologische studie
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 27 april 2012 om 2.30 uur precies

door

Natalie Alexandra Theresia Boll-Avetisyan
geboren op **8 augustus 1977**
te **Bad Schwalbach, Duitsland**

Promotor:
Prof. dr. R.W.J. Kager

Dit proefschrift maakt deel uit van het onderzoeksprogramma “Phonotactic Constraints for Speech Segmentation: The Case of Second Language Acquisition”, dat mede gefinancierd is door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Für Artjom

CONTENTS

CONTENTS vii
ACKNOWLEDGEMENTS xi

CHAPTER 1
GENERAL INTRODUCTION
1.1. Phonotactics, speech segmentation, and lexical acquisition 1
1.2. Phonotactics and its acquisition, representation, and use 4
1.2.1. The form of phonotactic representations in use 5
1.2.2. The acquisition of phonotactic knowledge 7
1.3. Outline of the dissertation 11
1.4. References 13

CHAPTER 2
DOES THE LEXICON BOOTSTRAP PHONOTACTICS, OR VICE VERSA?
2.1. Introduction 17
2.1.1. Phonotactics for speech segmentation 18
2.1.2. Segmentation strategies: trough-based vs. chunk-based segmentation 20
2.1.3. Sequential phonotactics and non-adjacent dependencies 21
2.1.4. Research questions and hypotheses 23
2.1.5. Approach 24
2.2. Experiment 1 27
2.2.1. Method 27
2.2.2. Results and discussion 30
2.3. Experiment 2 34
2.3.1. Method 34
2.3.2. Results and discussion 35
2.4. General discussion 36
2.5. Conclusion 40
2.6. References 40

CHAPTER 3
OCP-PLACE AS A CUE FOR SPEECH SEGMENTATION
3.1 Introduction 49
3.2. OCP-PLACE in Dutch 54

3.2.1. Method 54
3.2.2. Results and discussion 55
3.3. Experiment 1 56
3.3.1. Method 56
3.3.2. Results and discussion 59
3.4. Experiment 2 61
3.4.1. Method 61
3.4.2. Results and discussion 62
3.5. Experiment 3 63
3.5.1. Method 63
3.5.2. Results and discussion 64
3.6. Experiment 4 65
3.6.1. Method 65
3.6.2. Results and discussion 66
3.7. General discussion 66
3.8. Conclusion 69
3.9. References 69

CHAPTER 4
OCP-PLACE FOR SPEECH PROCESSING: A BIAS ON PERCEPTION OR ACQUISITION? THE CASE OF MANDARIN CHINESE
4.1 Introduction 81
4.2. Corpus study: OCP-PLACE in Mandarin Chinese 88
4.2.1. Method 88
4.2.2. Results and discussion 89
4.3. Experiment 1 91
4.3.1. Method 91
4.3.2. Results and discussion 94
4.4. Experiment 2 96
4.4.1. Method 96
4.4.2. Results and discussion 97
4.5. Experiment 3 99
4.5.1. Method 99
4.5.2. Results and discussion 100
4.6. General discussion 101
4.6.1. OCP-PLACE is not a result of channel bias 102
4.6.2. OCP-PLACE as a result of analytic bias? 103

4.6.3. The role of exposure 104
4.7. Conclusion 105
4.8. References 105

CHAPTER 5
PROBABILISTIC PHONOTACTICS IN LEXICAL ACQUISITION: THE ROLE OF SYLLABLE COMPLEXITY
5.1 Introduction 111
5.2. Experiment 1 117
5.2.1. Method 117
5.2.2. Results and discussion 119
5.3. Experiment 2 123
5.3.1. Method 123
5.3.2. Results and discussion 125
5.4. General discussion 128
5.5. Conclusion 130
5.6. References 130

CHAPTER 6
SECOND LANGUAGE PROBABILISTIC PHONOTACTICS AND SYLLABLE STRUCTURE IN SHORT-TERM MEMORY RECOGNITION
6.1. Introduction 139
6.2. Experiment 146
6.2.1. Method 146
6.2.2. Results and discussion 150
6.3. General discussion 156
6.4. Conclusion 159
6.5. References 159

CHAPTER 7
GENERAL CONCLUSIONS
7.1. Summary 167
7.1.1. Segment co-occurrence constraints for speech segmentation 167
7.1.2. Syllable structure constraints for lexical acquisition 169
7.2. Contributions 171
7.2.1. Phonotactic representations and their use 171
7.2.2. Phonotactics and its acquisition 173

7.3. Conclusion 175
7.4. References 176

SAMENVATTING IN HET NEDERLANDS 177

ACKNOWLEDGEMENTS

Whenever I happened to read an interesting linguistic article during my studies at the University of Mainz, it would report on a study that was carried out in the Netherlands. It may not be surprising that this created a kind of desire in me: to become part of this community that does all these interesting things. As things happen, I was lucky enough to find out that René Kager was looking for PhD students for a large project on phonotactics. Half a year later, here I was: in Utrecht in the Netherlands, first as a student of the Research Master's program, and later as a PhD student in René's group. Now, my dissertation project has come to an end, and I can say that my stay at Utrecht University at René's group was a great experience. I would like to acknowledge some people who made this possible.

Most of all, I want to thank René for having invited me to become part of his research program. I learned a lot from René. He is an encyclopaedia of knowledge and always asks the best questions. He taught me to work efficiently by expecting me to finish things in a week that I thought I would need two months for. He taught me how to sharpen my questions and arguments, and taught me writing by letting me rewrite and rewrite my texts ("the hypothesis is still not clear") until they were "Already MUCH BETTER!". Throughout the project, René was always available for questions and discussions, in regular weekly meetings and also in between. We had fun coming up with new artificial languages and listening to them until we knew them by heart.

I owe many thanks to Huub van den Bergh. Huub taught me as much as possible about multi-level models in many intensive get-togethers in the last 1.5 years of my project, in which we also practiced our receptive listening skills: we both spoke our native language. Huub understood how to explain statistics to me by using a more figurative language that I, being a "geesteswetenschapper" and not a mathematician, could understand.

I very much want to thank Frans Adriaans and Tom Lentz, the two other PhD students in René's group. Our endless discussions about work not only in our group meetings, but wherever we met, have clearly inspired this thesis. I really appreciated our honest and direct way of communicating, and that we always helped each other whenever needed. From the start, we were more like friends than like colleagues. I still miss our almost daily lunches at "De Zaak", and your jokes, and hope we will see each other at many conferences in the future. I also want to thank Diana Apoussidou, Marieke Kolkman, Keren Shatzman, Mirjam

Trapman, Chen Ao, Liquan Liu, Marko Simonović and Xiaoli Dong, who were part of the group at various stages, for bringing many opinions and viewpoints to our project. Chen Ao and I ended up collaborating on a study that became Chapter 4 or this thesis which could have never been done without her commitment flying to China for testing a whole bunch of native speakers of Mandarin. Bart Penning de Vries, Tim Schoof, Joost Bastings, Jorinde Timmer, Frits van Brenk, and Mieneke Langberg were student assistants and helped a lot with recoding baby data and testing.

I would also like to thank Juan Toro and Silke Hamann, who were in my support committee and always very helpful. I also thank Alexis Dimitriadis, Esther Janse, Elise de Bree and Annemarie Kerkhoff for always having been available for discussions.

I would like to acknowledge the NWO for providing the financial background that made it possible to visit great conferences and workshops abroad, such as LabPhon in Wellington, the Manchester Phonology Meeting, CCSC in München, and GALANA in Toronto and at the SISSA lab in Trieste, as well as summer schools such as the LSA summer school in Berkeley, and a summer school in Croatia. I would also like to thank Martin Everaert, Maaike Schoorlemmer, the secretaries, and the UiL-OTS as a whole for offering this great environment with plenty of amazing workshops and lecture series, carried out by the staff and the forty PhD students. Thanks also go to LOT for organizing the biannual summer and winter schools throughout the Netherlands with great courses and nice social events.

I also would like to thank the people who participated in the "Artificial language learning reading group": Elise de Bree, Annemarie Kerkhoff, Desiree Capel, Nick Puts, René Kager, Frans Adriaans, Alexis Dimitriadis and Loes Koring. I found these meetings extremely fruitful, as we all came from slightly different research backgrounds and could learn a lot about each other's projects. In fact, in these meetings, Elise and Annemarie raised my interest in infant research, and we ended up collaborating on a study which has become Chapter 2 of this thesis. Alexis, we still will be doing our experiment, too, I promise! I also very much enjoyed our "Elitu Meetings". Thanks to Hannah De Mulder and Roberta Tedeschi for being innovative and reliable Elitu co-organizers.

Thanks to Theo Veenker and Warmolt Tonckens for technical support and last-minute programming. I also thank Alexis Dimitriadis and Gianluca Giorgolo, who often helped me out with writing little or larger scripts that made my work life easier. Anjali Bhatara, Jessica Quinlan, Simon Barthelmé, Joseph DeVeaugh-Geiss, and Bettina Gruber have proof-read parts of this thesis. Tom Lentz and Frans Adriaans have translated my samenvatting into Dutch. I thank them a lot for their help!

I owe many thanks to all the people that made my daily life at UiL-OTS so enjoyable. These were obviously primarily my office-mates, who I would meet every day. During my first year, my office, which I shared with Jacqueline van

Kampen, was in a very quiet building in Achter de Dom. We had a very concentrated work sphere. Later, everyone was moved to Janskerkhof. Roberta Tedeschi, Hannah de Mulder, Loes Koring, Marieke Kolkman, Arnout Koornneef, Arjen Zondervan, Nadya Goldberg, and Marko Simonović: thank you for having been such great office mates in room 005. Sorry I often complained about the noise. Looking back it was great to have a lively office. We had many guests, because our room was just next to the "Ufo", the lunch area. Besides this, we had a foosball table that I was never really into, a toy machine filled with chocolate which you could only get with some "fingerspitzengefühl", and a bunch of mice. Marko, I am still shocked you killed one of them with my suitcase. The only thing we never had, although there was plenty of space, was a couch, because we were afraid that this would inspire someone to add another desk to our office.

I also want to mention the many other people I have met at lunches, coffee breaks or because they were enjoying the foosball table at our office: Bettina Gruber, Ivana Brasileiro, Gaetano Fiorin, Sander van der Harst, Min Que, Bert Le Bruyn, Andreas Pankau, Rianne Schipppers, Marieke Schouwstra, Jakub Dotlacil, Berit Gehrke, Dagmar Schadler, Daria Bahtina, Anna Chernilovskaya, Nicole Grégoire, Naomi Kamoen, Rob Zwitserlood, Rosie van Veen, Huib Kranendonk, Emilienne Ngangoum, Lizet van Ewijk, Christina Unger, Anna Volkova, Ana Aguilar, Cem Keskin, Marie-Elise van der Ziel, Giorgos Spathas, Kate Huddlestone, Sophia Manika, György Rákosi and everyone else.

The two people who are closest to me during my last steps in Utrecht will be Bettina Gruber and Tom Lentz. I thank them both for being my paranymphs.

Regarding my Utrecht life outside UiL-OTS, I would like to thank the Dutch people for having invented broodje haring and for celebrating the Queen's birthday. Thanks to my housemates at Villa Hertekamp. You made this place a home. Thanks Clizia, Anna Roos, Roberta, Virissa, Rolf, Luisa and Bettina for the ridiculous amount of miles we have run around the canals of Utrecht.

By now, I am back in Germany. Here, I would like to thank the people in Potsdam and Berlin, who were supportive and encouraging during the last steps of the thesis writing process: Barbara Höhle, Tom Fritzsche, Annika Unger, Frauke Berger, Hans Trukenbrod and Simon Barthelmé. Bei meinen Freunden in Deutschland möchte ich mich dafür bedanken, dass sie mich haben erfahren lassen, dass Freundschaft keine Grenzen kennt. Mama und Papa, euch danke ich, dass ihr mir beigebracht habt, dass sich Anstrengungen bezahlt machen, und dass man bei der Berufswahl auf sein Herz hören soll. Euch und meinen Schwestern Anne und Christina danke ich dafür, dass ihr immer da seid, wenn man euch braucht.

Artjom, du führst mich durch deine Welt und lässt mir die meine. Für dich war es selbstverständlich, dass ich ins Ausland gehe, um zu promovieren, und hast mich immer unterstützt. Wenn ich nicht bei dir war, hast du mich morgens per Telefon geweckt, damit ich Meetings nicht verschlafe. Wenn ich bei dir war,

hast du mir Kaffee gekocht, damit ich durcharbeiten konnte. Du hast mein Leid geteilt, wenn ein Experiment nicht funktioniert hat, und meine Freude verdoppelt, wenn ich auf eine Konferenz angenommen wurde. Du warst mir die wichtigste moralische Stütze. Dir widme ich dieses Buch.

CHAPTER 1

GENERAL INTRODUCTION

1.1. Phonotactics, speech segmentation, and lexical acquisition

No language allows for a random combination of phonemes within words. Instead, the way words are internally organized is language-specifically restricted. The set of language-specific constraints that define which phonemes may or may not co-occur within words is referred to as *phonotactics*. We have long known that humans have reliable knowledge about the phonotactics of their native language, and this knowledge affects their speech processing (e.g., Bailey & Hahn, 2001; Coleman & Pierrehumbert, 1997; Scholes, 1966; Vitevitch & Luce, 1999, for a more detailed discussion see Chapter 5). Research has shown that phonotactic knowledge is acquired early: at an age before infants have acquired their first words (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994, for a more detailed discussion see Chapter 2). Furthermore, its acquisition is not restricted to a native language. L2 learners have L2 phonotactic knowledge, which improves with increasing L2 proficiency (e.g., Dong, 2008; Lentz & Kager, submitted; Majerus, Poncelet, Van der Linden, & Weekes, 2008; Trapman & Kager, 2009; Weber & Cutler, 2006; see Chapter 4 and 6 for more detail).

This dissertation deals with aspects of the general question of how phonotactic knowledge is acquired and represented, and how it can be used in speech processing. To anticipate: phonotactics gives important cues in speech segmentation and is used to facilitate lexical acquisition. Let us first consider why this may be.

When listening to our native language, it does not immediately occur to us how difficult a task it is to translate the acoustic speech signal into a meaningful utterance. To us, having acquired our language a long time ago, it is obvious that utterances are made up of individual words. Startlingly, we hardly have any difficulty in consciously identifying where a word starts and ends. Orthographic systems reflect this knowledge by using spaces to mark word boundaries. The speech signal, however, contrary from what one might intuitively assume, does not contain any clear-cut cues to word boundaries. The acoustic signal is

continuous (e.g., Cole & Jakimik, 1980). This observation has led to one of the most challenging and interesting research questions in the field of psycholinguistics: the *speech segmentation problem*. When listening to speech, we need to segment it into words or sub-lexical units, i.e., phonemes, biphones, or syllables, which allow us access the words stored in our mental lexicon. How does this take place?

The speech segmentation problem is probably most challenging for language-learning infants. Adults may at least partly rely on feedback from word knowledge when processing the speech signal. For a long time, it was even assumed that speech segmentation is merely a by-product of lexical recognition (e.g., Cole & Jakimik, 1980; Marslen-Wilson & Welsh, 1978). However, then the question arises of how language-learning infants lacking in lexical knowledge would segment the speech signal. If speech segmentation were only possible after lexical recognition, then pre-lexical infants would face an irresolvable task. How would they be able to initiate lexical acquisition, if they lacked lexical knowledge facilitating speech segmentation? Hence, it follows that the capacity to segment speech is a pre-requisite for lexical acquisition for pre-lexical infants.

Among learners of a second language (L2), speech segmentation poses almost just as much of a problem as in first language (L1) acquisition. Adult L2 learners may have one advantage: an initiation of L2 lexical acquisition does not depend on the ability to segment speech of the target language. Particularly in systematic formal language education, adult learners are taught lists of isolated words. Lexical knowledge acquired in such a way may be helpful when segmenting L2 speech. However, anyone who has had foreign language training at school or university may share the experience that understanding foreign speech uttered by native speakers is anything but easy. Only through regular exposure to an L2 spoken by native speakers can we tune in to understanding speech of a non-native language and learn how to segment it.

The learning conditions of both pre-lexical infants and L2-learning adults illustrate the necessity of cues other than lexical information for speech segmentation. It becomes clear that the speech signal must contain other, acoustic or phonological *sub-lexical* cues that point to word boundaries in continuous speech. In fact, since the 1990s, a growing body of studies has gathered evidence that sub-lexical knowledge can be used as a cue for speech segmentation. Sub-lexical cues come for example from prosodic cues, such as pauses, segmental final lengthening, metrical patterns, and intonation contour (e.g., Cutler & Norris, 1988; Shatzman & McQueen, 2006). Another useful sub-lexical cue for speech segmentation comes from phonotactics. Why that is can easily be illustrated with an example. If someone utters the sentence *Five women left*, then we hear it as a continuous stream, e.g., /faɪvwɪmənlɛft/. This utterance contains phonotactic cues to the word boundaries as neither /vw/ in /faɪvwɪmən/

nor /nl/ in /wɪmənlɛft/ form possible word onsets or offsets in English. Research has shown that such phonotactic knowledge can be used as a cue to detect words in speech (e.g., McQueen, 1998; van der Lugt, 2001). There is evidence that at the age of nine months, infants already use their phonotactic knowledge for speech segmentation (e.g., Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999, see chapter 2 for a more detailed discussion). Also, L2 learners can rely on L2 phonotactic cues when segmenting L2 speech (Weber & Cutler, 2006). As will be discussed in Chapters 2, 3 and 4, listeners can even use phonotactic cues from phoneme co-occurrence probabilities of vowels that are non-adjacent. For example, in Finnish a sequence of /y/ followed by a consonant followed by /u/ (i.e., /yCu/) cannot occur within words. This is so because Finnish restricts its lexicon by vowel harmony—a constraint that requires vowels to be similar, for example, in vowel height. Knowledge of vowel harmony can also be used to find word boundaries, in this case before the /u/ (e.g., Kabak, Maniwa, & Kazanina, 2010; Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & de Gelder, 1998).

These studies suggesting a role for phonotactics in speech segmentation also show how phonotactic knowledge indirectly has the potential to facilitate lexical acquisition. In fact, this possibility has already been suggested in computational studies: phonotactics knowledge acquired from the continuous speech signal helps the learner to segment the stream. The product of segmentation can be stored as proto-words (literature will be introduced in more detail below and in Chapter 2).

Beyond this indirect facilitation of lexical acquisition, phonotactics has furthermore been found to directly facilitate the acquisition of words. It may be intuitive to conceive that some words are intrinsically easier to acquire than others, because they are composed of frequently occurring phoneme combinations. Storkel (2001) gives an example to illustrate this. The word *sit*, she reports, might be easier to learn than the word *these*, even though they are structurally much alike. This is because *sit* is combined of frequently occurring phoneme pairs /sɪ/ and /ɪt/, whereas /ðiː/ and /iːz/ in *these* are not very common. In fact, phonotactic knowledge has been shown to facilitate lexical acquisition in children and adults (e.g., Gathercole, Frankish, Pickering, & Peaker, 1999; Storkel, 2001; Storkel, Armbrüster, & Hogan, 2006; Storkel & Rogers, 2000, and see Chapter 5 and 6 for more detail), and there is some evidence that it might also affect lexical acquisition in an L2 (Majerus, et al., 2008, see Chapter 6 for a more detailed discussion).

I will not discuss the background literature at full length in this introduction; all following chapters are stand-alone studies, and provide the reader with a more detailed literature review. However, what should be clear at this point is the following: regarding the problem of how to create the lexicon, there seems to be an intrinsic relation between speech segmentation and lexical acquisition. A lexicon cannot be acquired without knowing how to segment speech.

Phonotactic knowledge seems to be a key to both speech segmentation and lexical acquisition. In the remainder of this introduction, I will discuss the research questions that are the focus of this dissertation.

1.2. Phonotactics and its acquisition, representation, and use

Given the vast evidence for a role of phonotactic knowledge in speech segmentation and lexical acquisition, it may come as a surprise that yet another full dissertation is dedicated to this topic. However, a number of questions have yet to be resolved. In general, the study of language deals with three aspects. It addresses the question of how language is acquired, how it is represented, and how it is used in speech processing. These three aspects of language are intrinsically related. Linguistic representations have to be acquired in order to be used, and, essentially, humans need linguistic representations in order to process speech.

Language is studied by researchers from different backgrounds with different interests: psychologists and linguists. Typically, psychologists, i.e., developmental, cognitive and computational psychologists, will ask questions about how the mind operates in processing, and how the mind learns. The classical interests of linguists, i.e., phonologists, syntacticians and semanticists, refer to the internal structure and complexity of linguistic systems and why certain linguistic patterns are typologically recurrent. Beyond this, they may take an interest in the cognitive biases that lead to typological frequencies and make it possible to acquire complex representations. Both linguists and psychologists have neglected one aspect of language: psychologists often disregard the nature and complexity of linguistic representations, while linguists tend to neglect the consequences of the fact that linguistic representations need to be used in speech processing. With respect to phonotactics, phonological research may have sometimes kept in mind that phonotactics defines the shape of words, and, hence, might play a role in word processing. However, the question of how phonotactic knowledge is actually put to use in speech processing has been heavily neglected.

Knowing this, it may not come as a surprise that in the study of the role of phonotactics in speech segmentation and lexical acquisition—a processing-oriented issue—many questions regarding the complexity of phonotactic representations and their acquisition have remained ignored. Naturally, this is because researchers who have dealt with the topic tend to have a background in developmental, cognitive or computational psychology. To my knowledge, phonologists have left the topic almost untouched (exceptions are studies by Kabak, et al., 2010; Kager & Shatzman, under revision; and Warner, Kim, Davis, & Cutler, 2005).

The present dissertation addresses different issues regarding the acquisition and representation of phonotactic knowledge as it is used in speech processing. It is aimed at combining research interests of phonologists and psychologists in the context of phonotactics for speech segmentation and lexical acquisition. Three research questions form the heart of this thesis. The first deals with the form of phonotactic representations that facilitate speech segmentation and lexical acquisition. The other two address the acquisition of phonotactic knowledge, examining on the one hand the potential role that innate cognitive or functional biases may play when learners discover phonotactic cues for speech segmentation, and on the other the source of phonotactic learning.

1.2.1. The form of phonotactic representations in use

This section will investigate which form of phonotactic representations may facilitate speech segmentation and lexical acquisition. A majority of the studies that have looked at a role for phonotactics in speech segmentation and lexical acquisition have *only* looked at the role of specific *n*-phone probabilities (e.g., Gathercole, et al., 1999; Mattys & Jusczyk, 2001; Mattys, et al., 1999; Storkel, 2001). Exceptions which look into more abstract phonotactic constraints are studies on syllable structure (e.g., McQueen, 1998) and the studies on vowel harmony as a cue for speech segmentation (e.g., Suomi, et al., 1997; Vroomen, et al., 1998; discussed in more detail in Chapter 3). In this dissertation, I will investigate whether phonotactic representations of a complex nature can facilitate speech segmentation and lexical acquisition.

There is considerable evidence from phonological research—mainly supported by data from typology as well as from a few studies on perception and wellformedness intuitions of nonwords (e.g., Albright, 2009; Moreton, 2002, see Chapter 4 and 5 for a more detailed discussion)—suggesting that at least some phonotactic knowledge must be represented in an abstract, symbolic fashion. Phonotactics subsumes abstract structural constraints that can be subdivided into abstractions of segmental (i.e., featural) structure, such as requirements on phoneme co-occurrence patterns such as similarity or dissimilarity, and abstractions of prosodic structure, such as requirements on syllable structure. Many such phonotactic restrictions have been found to systematically affect many genetically and geographically unrelated languages. Phonotactic knowledge of such systematic restrictions on structure is represented by abstract symbolic constraints as part of a phonological grammar (e.g., Prince & Smolensky, 1993/2004).

Featural structure

As mentioned above, previous literature has shown that vowel harmony, a constraint requiring featural identity between neighboring vowels, can be used for speech segmentation. In this dissertation, the focus will be on non-identity constraints instead. Chapters 2, 3 and 4 will investigate the effect of featural non-identity constraints on speech segmentation. As a case of a non-identity constraint, I will focus on OCP-PLACE, a typologically well-attested constraint against consonants with a shared feature [place] within words. In many languages, this constraint takes probabilistic effects: In English, for example, sequences such as /p-m/ in the word *spam* are not forbidden. They are just not very frequent (e.g., Berkley, 1994, see Chapter 3 and 4 for more detail). There is evidence that native Hebrew, Arabic, English and Dutch listeners have knowledge of OCP-PLACE (Berent & Shimron, 1997; Coetzee, 2008; Frisch & Zawaydeh, 2001) and use their knowledge for lexical processing (Berent, Everett, & Shimron, 2001; Kager & Shatzman, 2007, see Chapter 3 and 4 for more detail). The effect of OCP-PLACE manifests itself across intervening vowels, constraining the co-occurrence of consonants that are non-adjacent (although consonants are represented as adjacent on the "consonant tier"; McCarthy, 1986). Hence, it requires a mental representation that abstracts over intervening vowels. As will be discussed below (in section 1.2.2.) and in detail in Chapter 3 and 4, the fact that OCP-Place restricts non-adjacent phonemes that are required to be non-identical makes it particularly interesting from an acquisition perspective. The specific question that will be addressed in the two chapters is whether OCP-PLACE can be used for segmentation, and whether its use is language-specific.

Another question that will be discussed in detail in Chapter 2 is whether high or low phonotactic probability, or both, can be used as a cue for speech segmentation. This question relates to a debate that can be traced back to a computational study by Perruchet & Vinter (1998), which raised the question of whether listeners apply a trough-based segmentation strategy, attending to low transitional probabilities in the stream, or rather a chunk-based segmentation strategy, attending to high transitional probabilities. In this dissertation, it is hypothesized that a trough-based segmentation strategy will be more useful, as listeners need to attend to less information in the speech signal to locate the word boundaries than when applying a chunk-based strategy. The hypothesis will be tested in Chapter 2. This is not to say that knowledge of highly probable sound sequences would be useless in processing. In fact, high probability phonotactics may be used to facilitate lexical acquisition. This aspect will be addressed indirectly in Chapter 6.

Syllable structure

Chapters 5 and 6 will look at a different type of abstract knowledge. There is typological evidence that languages restrict the structure of syllables in a similar way: Within and across languages, syllables containing complex syllable constituents are less common (or more marked) than syllables made up of simple constituents. Complexity of syllable structure has been found to influence speech processing (e.g., Treiman, 1983; Treiman & Danis, 1988, see Chapter 5 and 6 for a more detailed discussion).

The effects of probabilistic phonotactics on the lexical acquisition of complex syllables are worthy of our attention for the following reason: up to this point, studies have only looked at effects of probabilistic phonotactics on the acquisition of words with a simple CVC syllable structure. Results of these studies can never provide a full picture of phonotactic effects on lexical acquisition, as many languages have words that are of a more complex structure including consonant clusters in syllable onsets and codas.

Results from previous studies (e.g., Albright, 2009; Moreton, 2002) suggest that probabilistic phonotactics, represented as specific *n*-phone probabilities, and prosodic structure, represented as abstract, symbolic constraints, are, as they are different by nature, part of two separate knowledge components. By investigating how knowledge of syllable structure interacts with knowledge of phonotactic probabilities in lexical acquisition, Chapter 5 and in particular Chapter 6 will address the question of whether a separate storage will lead to separable effects of the two knowledge components. It was hypothesized that the facilitatory effects of probabilistic phonotactics should be enhanced by structural knowledge.

1.2.2. The acquisition of phonotactic knowledge

With respect to phonotactic acquisition, two questions will be addressed in this dissertation. The first question deals with the potential role that innate cognitive or universal functional biases may play when learners in both L1 acquisition and L2 learning discover phonotactic cues for speech segmentation. More specifically, it will be asked whether knowledge of OCP-PLACE is language-specific, and hence, acquired from the input, or due to some universal perceptual bias, and hence, does not need be acquired. The second question deals with the source of phonotactic learning. More specifically, it will be asked whether phonotactics are acquired from the lexicon or from continuous speech.

The acquisition of phonotactics in an L1

Linguists as well as psychologists have always been fascinated by what makes it possible for children to acquire their native language in such a short time. The

ability to learn to identify words constitutes a part of this learning process. One factor that seems to be crucial in driving language acquisition is (the presence of a sufficient) language input; here the major question arises of how infants use information in the input as a source for internalizing linguistic knowledge. Yet another intriguing question is to what extent the child's innate endowment plays a role in language acquisition. Do children need an innate learning mechanism for discovering certain structures? If so, what kind of mechanisms are involved?

There is consensus among researchers that there must be some innate mechanisms that guide the language-learning infants' attention to information in the input that is relevant for a successful and fast language acquisition process. However, there is discussion about how specialized innate mechanisms must be. Roughly, the positions can be divided into two views. One posits that there is no need for mechanisms other than statistical learning and is defended primarily by psychologists. The other states that there must be additional mechanisms that relate language acquisition to typological data—a position mostly taken by linguists.

The statistical learning mechanism is a domain-general mechanism, which equips humans (but also animals, see Hauser, Newport, & Aslin, 2001) with the ability to extract statistical regularities from a given input. The input can for example come from the continuous stream of speech to which language-learning infants are exposed. In the speech signal, the language-learning child may, for example, find statistical regularities of syllable co-occurrences (Saffran, Newport, & Aslin, 1996) or phoneme co-occurrences (Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997, See also Chapter 2 for a more detailed discussion).

In this dissertation, evidence for probabilistic phonotactic knowledge is provided in all chapters. Hence, it seems clear that some form of statistical learning is important: Language-specific probabilistic phonotactics, in particular *n*-phone probabilities, must be acquired from distributions in the input. However, as we will also see, a part of phonotactic knowledge needs more abstract mental representations than mere *n*-phone probabilities. For that part, the question is raised whether additional learning mechanisms are at play.

As mentioned before, languages are full of systematic restrictions, which are often also typologically well attested. This raises the question of whether learners are naturally biased to prefer structures that are cross-linguistically common (or unmarked) to structures that are cross-linguistically rare (or marked). It has been proposed that there must be a cognitive endowment that enables us to pick up on specific structures. Proponents of this view can be sub-divided into two positions: functionalists and formalists. In a coarse-grained definition, functionalists usually defend the view that humans have an innate endowment to be able to (produce and) perceive certain structures, a factor that continuously affects the course of language change, eventually leading to their

typological frequency. In the perceptual domain, such a bias (also referred to as "channel bias" by Moreton, 2008) does not necessarily need to be unique to the linguistic domain proper, but may be valid for domains other than language, too (e.g., Frisch, Pierrehumbert, & Broe, 2004; Ohala, 1994). Being intrinsic to our perceptual abilities, its influence continues throughout the course of human life.

Formalists mostly argue that there are cognitive biases on acquisition: The assumption is that infants are innately driven to attend to certain aspects of language that help them to extract abstract structural dependencies and create symbolic representations (e.g., Jakobson, 1941; Smolensky, 1996). The aspects they are innately driven to acquire are the same that are typologically recurrent. These cognitive biases (also referred to as "analytic bias" by Moreton, 2008), however, only take an effect in the acquisition process. Once a language system with its language-specific rules and constraints is acquired, the biases will no longer influence perception or production (e.g., Dresher & Kaye, 1990).

It has been proposed that OCP-PLACE—a phonotactic constraint under investigation in this dissertation—is grounded in a functional bias on perception (e.g., Frisch, 2004; see chapter 3 for a detailed discussion). Taken to its extreme, this view assumes that the constraint is completely reduced to the functional bias and may thereby imply that knowledge of OCP-PLACE does not need to be acquired, predicting that the functional bias should universally affect listeners' speech perception regardless of whether their native language supports the OCP constraint or not. However, Chapter 3 and 4 of this dissertation aim at refuting such an extreme approach. Instead, it is hypothesized that knowledge of OCP is language-specific, and hence must be acquired from the input, possibly by cognitive biases on acquisition.

The acquisition of phonotactics in an L2

This dissertation addresses the question of whether L2 learners can acquire and use L2 distributional phonotactic knowledge as well as L2 structural knowledge for speech segmentation and lexical acquisition. It has often been observed that L1 acquisition under normal conditions is mastered without many problems in a relatively short time even based on imperfect input (e.g., Chomsky, 1968), whereas adults are poor learners of L2, and native-like ultimate attainment is rare (e.g., Flege, 1987). In particular, L2 learners have difficulties with acquiring abstract symbolic L2 knowledge (e.g., Odlin, 1989).

In the domain of phonology, an explanation has been proposed to be that L1 abstract phonological representations are so deep-wired that L2 learners cannot possibly re-structure their structural knowledge. Concerning phonotactic learning in an L2, many studies on production (e.g., Broselow, 1984) but also some on perception (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Lentz & Kager, submitted) have found evidence for L1 transfer of structural knowledge: often, non-native consonant clusters are “repaired” to a simpler structure, for

example, by inserting an epenthetic vowel. An explanation for L2 learners' difficulty in mastering abstract structures proposed in the literature is that innate cognitive analytic biases only facilitate language acquisition in children. At later ages, the influence of these biases stops (e.g., Clahsen & Muysken, 1989). However, some studies have found that L2 learners can acquire probabilistic phonotactics of the target language (e.g., Lentz & Kager, submitted; Trapman & Kager, 2009) and use phonotactic knowledge for processing, for example, in speech segmentation (Weber & Cutler, 2006).

A question that will be addressed in this dissertation is whether it could be that also L2 learners have access to cognitive biases on acquisition, but that their influence is less transparent because of L1 structural transfer. Might it be that if L1 structural representations do not run against L2 structure, that L2 phonotactics can be acquired?

In this dissertation, two questions regarding an acquisition of phonotactics in an L2 will be investigated. Chapter 6 asks whether L2 probabilistic phonotactics can be acquired for structures that are not attested in the L1, and whether this knowledge can then be used to facilitate lexical acquisition, independent of knowledge of L2 structural knowledge. This hypothesis is based on the assumption that structural and probabilistic phonotactic representations have separate places of storage.

Chapter 4 will investigate whether L2 learners can acquire phonotactic constraints featural structure of the type OCP-PLACE, and use it as a cue for speech segmentation—if there is no structural constraint that might transfer and, hence, impede the acquisition. It has previously been observed that non-adjacent dependencies are only learnable, if they affect elements with a certain degree of similarity (e.g., Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005; see Chapter 2 and 4 for a more detailed discussion). The question whether L2 learners can acquire such a complex constraint is interesting, as it might suggest that L2 learners receive support from an innate cognitive mechanism.

A further issue dealt with in Chapter 4 is the learning condition under which L2 phonotactic knowledge can ideally be acquired. L1 and L2 acquisition differ with respect to the learning conditions: while the native language is learned from the ambient language’s speech, L2s are not always acquired from native spoken input. This dissertation hypothesizes that the acquisition of L2 phonotactics requires exposure to speech input from native speakers. This hypothesis will be tested in Chapter 4 by comparing learners who live in the country where the target language is spoken and learners who live in their home country.

The source of phonotactic knowledge

A third question relating to the acquisition of phonotactics dealt with in Chapter 2 (and indirectly in Chapter 5) addresses the source of phonotactic knowledge. Phonologists have generally assumed that phonotactics emerges as abstractions over word knowledge (e.g., Hayes & Wilson, 2008; Pierrehumbert, 2003; see Chapter 2 for a more detailed discussion). This view, however, creates a paradox for language-learning infants. It implies that they would have to acquire a sizeable lexicon, one large enough to extract phonotactic knowledge from, before they could start using phonotactic cues that facilitate speech segmentation.

Avoiding this paradox, phonotactic knowledge could be acquired directly from unsegmented, i.e. continuous speech. Computational studies have proposed that this is possible and supported this proposal by means of evidence from simulations with learning models. In such studies, computational learners first learn distributional phonotactics from continuous speech, which they subsequently use to identify word boundaries in continuous speech. The resulting segmented speech stream, these studies argue, can serve to construct a proto-lexicon (e.g., Adriaans & Kager, 2010; Brent & Cartwright, 1996; Cairns, et al., 1997). Up to this point, the question of whether speech segmentation cues are directly acquired from continuous speech has never been specifically addressed in studies with infants (see Chapter 2 for a detailed discussion). In Chapter 2, we will test the hypothesis that infants acquire phonotactic cues for speech segmentation from continuous speech.

As mentioned above, studies that have addressed the role of phonotactic cues for speech segmentation in infancy have merely looked at phonotactics as linear *n*-phone probabilities. These studies offer important support for the assumption that the speech signal contains much distributional phonotactic information that infants might rely on for discovering word boundaries and words. Potentially, infants can also acquire phonotactic knowledge from continuous speech that requires abstract representations (see 1.2.1.). This question is also addressed in Chapter 2 by investigating the effects of knowledge of a sequential restriction on non-adjacent phonemes. More specifically, the experiment will concern a phoneme co-occurrence (i.e., /s/-vowel-/s/) that occurs less often than expected in infant-directed continuous speech, but more often than expected in the infant-directed lexicon.

1.3. Outline of the dissertation

The dissertation is structured as follows. The first three chapters (**Chapters 2, 3 and 4**) address the question of whether phonotactic constraints of a certain

degree of complexity, i.e., phonotactic constraints on non-adjacent dependencies, can be used as a cue for speech segmentation. As a method, these three chapters used the artificial language learning paradigm to assess whether phonotactic knowledge would be transferred when segmenting an artificial language.

Chapter 2 presents a study with infants to address a question concerning the acquisition of phonotactics, namely whether the source of phonotactic learning is continuous speech or the lexicon. In this chapter it is argued that this question can be investigated by looking at whether infants use a trough-based instead of a chunk-based segmentation strategy, as hypothesized here. This line of reasoning follows from the observation that it should be quite difficult for infants to acquire low probability phoneme co-occurrences from their small proto-lexicon. Furthermore, it will be investigated whether infants can also acquire and use constraints on non-adjacent dependencies.

While **Chapter 2** only looks at the use of non-adjacent dependencies on specific phoneme pairs, **Chapter 3** establishes whether OCP-PLACE, an abstract constraint on featural structure, can serve as a cue for speech segmentation. In four experiments with adult native listeners of Dutch, this chapter aims at providing evidence against the extreme functionalist position that entirely reduces OCP-PLACE to a perceptual bias against sequences of similar phonemes, supporting instead the view that the use of the constraint is related to segment distribution in the native language.

In order to further investigate the question of whether effects of OCP-PLACE on speech segmentation are due to perceptual biases or cognitive biases on acquisition, in **Chapter 4**, Mandarin Chinese is brought into play. Mandarin Chinese is a language that does not restrict its lexicon by the constraint OCP-PLACE. It will be tested whether native listeners of Mandarin Chinese and Mandarin Chinese L2 learners of Dutch use OCP-PLACE for segmentation. This chapter furthermore addresses the question of the role of exposure to native spoken input in L2 acquisition by comparing L2 learners who have learned Dutch while living in the Netherlands with L2 learners who have learned Dutch while living in China.

The following two chapters (**Chapters 5 and 6**) look at the role of phonotactic knowledge in lexical acquisition. As a method, they used a short-term memory recognition paradigm. Effects on short-term memory have been found to correlate with effects on lexical acquisition, as will be explained in the two chapters.

Chapter 5 aims at testing the hypothesis that structural phonotactic knowledge and probabilistic phonotactics belong to two separate sub-lexical knowledge components, and that these have an individual, but interacting effect on lexical acquisition. This chapter will present two experiments assessing Dutch adults'

memory of simple and complex syllables manipulated for phonotactic probability.

Chapter 6 will follow up on the question concerning the representation of phonotactics, adding the question of phonotactics acquisition. It is predicted that if probabilistic phonotactics and structural constraints are stored in two separate sub-lexical components, then they should also be acquired independently of each other. Therefore, although it is difficult to learn structural knowledge of an L2, L2 probabilistic phonotactics should be acquired independently and facilitate L2 lexical acquisition. This hypothesis is tested with Spanish and Japanese L2 learners of Dutch using the same stimuli as in Chapter 5. The syllable structures of the stimuli are only partly legal in the respective L1s of the L2 participants.

Chapter 7 will offer a conclusion by summarizing and combining the findings of all studies.

1.4. References

Adriaans, F., & Kager, R. (2010). Adding Generalization to Statistical Learning: The Induction of Phonotactics from Continuous Speech. *Journal of Memory and Language, 62*, 311-331.

Albright, A. (2009). Feature-based generalization as a source of gradient acceptability. *Phonology, 26*(1), 9-41.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighbourhoods? *Journal of Memory and Language 44*, 568-591.

Berent, I., Everett, D. L., & Shimron, J. (2001). Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology, 42*, 1-60.

Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition, 64*, 39-72.

Berkley, D. (1994). The OCP and Gradient Data. *Studies in the Linguistic Sciences, 24*, 59-72.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93-125.

Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics in Language Teaching, 22*, 253-269.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111-153.

Chomsky, N. (1968). *Language and Mind*. New York: Harcourt, Brace & Jovanovich.

Clahsen, H., & Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research, 5*, 1-29.

Coetzee, A. (2008). Grammar is both categorical and gradient. In S. Parker (Ed.), *Phonological Argumentation*. London: Equinox.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Coleman, J., & Pierrehumbert, J. B. (1997). *Stochastic phonological grammars and acceptability.* Paper presented at the 3rd Meeting of the ACL Special Interest Group in Computational Phonology, Somerset, NJ.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance 14*, 113-121.

Dong, X. (2008). *Hidden Phonotactic Knowledge of L2 Listeners.* Utrecht University, Utrecht.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77-91.

Flege, J. 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.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. M. Kirchner & D. Steriade (Eds.), *Phonetically based Phonology* (pp. 346-371). Cambridge: Cambridge University Press.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory 22*, 179-228.

Frisch, S. A., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language Acquisition, 77*, 91-106.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84-95.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53-B64.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39*(3), 379-440.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell. Reprint (1969). Frankfurt am Main: Suhrkamp.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V., & Jusczyk, A. M. (1993). Infants‘ sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402-420.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants’ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630-645.

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Journal of Laboratory Phonology, 1*, 207-224.

Kager, R., & Shatzman, K. (2007). Phonological constraints in speech processing. In B. Los & M. v. Koppen (Eds.), *Linguistics in the Netherlands 2007* (pp. 100-111). Amsterdam: John Benjamins.

Kager, R., & Shatzman, K. (under revision). *Similarity Avoidance in Speech Processing*.Unpublished manuscript.

Lentz, T. O., & Kager, R. W. J. (submitted). *Categorical and Gradient Phonotactics Are Represented Differently: Evidence from Word Recognition*.Unpublished manuscript.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107*, 395-419.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10* 29-63.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91-121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

McCarthy, J. J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry, 17*, 207-263.

McQueen, J. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language, 39*(1), 21-46.

Moreton, E. (2002). Structural constraints in the perception of English stopsonorant clusters. *Cognition, 84*, 55-71.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology 25*(1), 83-127.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology 48*, 127-162.

Odlin, T. (1989). *Language Transfer*. New York: Cambridge University Press.

Ohala, J. (1994). *Towards a universal, phonetically-based theory of vowel harmony.* Paper presented at the Proceedings of the International Conference on Spoken Language Processing (ICSLP).

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*, 225-237.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge MA: MIT press.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*: Rutgers University and University of Colorado at Boulder, 1993, Revised version published by Blackwell, 2004

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Scholes, R. (1966). *Phonotactic Grammaticality*. The Hague: Mouton.

Shatzman, K., & McQueen, J. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. . *Perception and Psychophysics 68*(1), 1-16.

Smolensky, P. (1996). The Initial State and 'Richness of the Base' in Optimality Theory. *154*(1196). Retrieved from http://roa.rutgers.edu/files/154-1196/roa-154-smolensky-2.pdf

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321-1337.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotacic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*, 1175-1192.

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics., 14*, 407-425.

Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.

Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition, 16*, 178-221.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*, 49-74.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 145-152.

van der Lugt, A. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics, 63*, 811-823.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighbourhood Activation in Spoken Word Recognition. *Journal of Memory and Language, 40*, 374-408.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133-149.

Warner, N., Kim, J., Davis, C., & Cutler, A. (2005). Use of complex phonological patterns in speech processing: Evidence from Korean. *Journal of Linguistics 41*, 353-387.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America, 119*, 597-607.

CHAPTER 2

DOES THE LEXICON BOOTSTRAP PHONOTACTICS, OR VICE VERSA?[1]

2.1. Introduction

From an early age onwards, infants start acquiring the sound structure of their ambient language, including the acquisition of prosodic knowledge, the formation of phonemic categories and the acquisition of language-specific phonotactic knowledge about the frequency with which sounds co-occur within words or across word boundaries (for an overview, see Jusczyk, 1997). It has been proposed that phonological knowledge bootstraps syntactic and lexical acquisition (e.g., Höhle, 2009; Morgan & Demuth, 1996; Weissenborn & Höhle, 2001). In this context, it is an intriguing question how infants learn to segment the continuous stream of language that surrounds them.

Speech segmentation is not a trivial task, as speech, unlike written language, does not contain pauses between words. However, the distribution of phonological units such as phonemes, syllables and metrical feet in speech provides important cues to word boundaries. In continuous speech, co-occurrence probabilities of adjacent phonological units are generally higher within words than across words. Previous research suggests that infants do well at extracting multiple distributional acoustic and phonological cues to word boundaries from the speech signal and use it as a cue for segmentation, such as lexical stress (e.g., Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999; Thiessen & Saffran, 2003), allophonic variation (e.g., Jusczyk, Hohne, &

[1] This study, submitted for publication, was carried out in collaboration with René Kager, Elise de Bree, Annemarie Kerkhoff, Huub van den Bergh and Sandra den Boer, Utrecht University, and is submitted for publication. It was funded by two NWO grants, one (277-70-001) awarded to René Kager, and another (360-70-270) awarded to Frank Wijnen. The authors contributed to the study as follows: hypothesis: Natalie Boll-Avetisyan and René Kager; design: Natalie Boll-Avetisyan, Elise de Bree, and Annemarie Kerkhoff; testing: Natalie Boll-Avetisyan, Sandra den Boer, Elise de Bree, and Annemarie Kerkhoff; analysis: Huub van den Bergh and Natalie Boll-Avetisyan, report: Natalie Boll-Avetisyan.

Baumann, 1999), and phonotactic distributions (e.g., Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999; Van Kampen, Parmaksiz, van de Vijver, & Höhle, 2008)

2.1.1. Phonotactics for speech segmentation

The current study focuses on the role of phonotactic knowledge in infant speech segmentation. Phonotactic knowledge develops relatively early, just before the onset of lexical acquisition. For example, 9-month-old infants prefer listening to nonwords that are phonotactically licit over those that are illicit in their ambient language (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Beyond this, infants possess gradient phonotactic knowledge. Nine-month-olds prefer listening to nonwords that are made up of pairs of phonemes of high probability of occurrence within words over nonwords made up of phoneme pairs of low probability (Jusczyk, Luce, & Charles-Luce, 1994). Further support for the possibility that infants extract language-specific phonological knowledge from their input stems from artificial language learning (ALL) experiments, in which infants learn and generalize phonotactic constraints from distributions in artificial languages (e.g., Chambers, Onishi, & Fisher, 2003; Saffran & Thiessen, 2003).

Regarding the role of phonotactics in infant speech segmentation, Mattys and collaborators (Mattys, et al., 1999) found that 9-month-old infants have a preference for CVCCVC items if the medial consonant cluster has a high within-word phonotactic probability in the ambient language. However, if the CVCCVC strings are pre-segmented by inserting subliminal pauses between the consonants in the cluster, infants prefer CVC.CVC items with low within-word phonotactic probability of the consonant cluster. In a related study, 9-month-olds were first trained on nonword strings that were later embedded in nonsense strings. The infants detected the trained nonwords more easily when they were embedded in a phonotactic context that is low in probability than if the phonotactic boundaries were high in probability in the ambient language (Mattys & Jusczyk, 2001).

The ability to segment speech into proto-words has been argued to be a prerequisite for lexical acquisition (e.g., Jusczyk, 1997). Hence, the ability to use phonotactic knowledge as a cue for speech segmentation will facilitate lexical acquisition. It has generally been assumed that phonotactics is derived from lexical knowledge, i.e., from word-entries in the mental lexicon (e.g., Jusczyk, et al., 1994; Pierrehumbert, 2003). From an infant's perspective, however, a paradox arises: Facilitative cues from phonotactics for segmenting speech into proto-words could only be acquired after the onset of lexical acquisition.

Avoiding this paradox, phoneme distributions in continuous speech might be proposed as an alternative source of phonotactic knowledge. Learning phonotactics from continuous speech implies that facilitatory cues for segmentation can be acquired before the onset of lexical acquisition. Another potential benefit when acquiring segmentation cues from continuous speech is that there might be distributions that are harder to acquire from a lexicon, i.e., distributions of low probability of occurrence.

Several computational studies have suggested that segmentation cues can be derived from unsegmented input to bootstrap into word learning (e.g., Adriaans & Kager, 2010; Brent, 1999; Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Perruchet & Vinter, 1998; Swingley, 2005). Some propose that segmentation cues are extracted by attending to utterance boundaries (e.g., Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998; Gambell & Yang, 2005). However, artificial language learning (ALL) studies with humans (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Newport, & Aslin, 1996a, 1996b) have shown that segmentation cues can be extracted from continuous speech without utterance boundaries. Adults as well as already pre-lexical infants can track changes in transitional probabilities (TPs, the probability with which one unit will be followed by another) between syllables in continuous unfamiliar artificial speech streams, and use this distributional information as a cue for segmentation. That is, they prefer inserting word boundaries at troughs in TP between syllables (i.e., between syllables that are unlikely to co-occur). Such results support the hypothesis that infants extract language-specific distributional regularities from speech and use these as a cue for segmentation. This is further supported by the fact that memory of lexical knowledge of the native language is of no use for the task: the distributional information must be computed on-line while listening to the artificial language.

It is difficult to investigate whether infants acquire phonotactic cues for speech segmentation before the onset of lexical acquisition. The literature suggests that lexical knowledge develops shortly after distributional phonotactic knowledge is acquired. Knowledge of phonotactic distributions must be acquired between 6 and 9 months (Friederici & Wessels, 1993; Jusczyk, et al., 1993; Jusczyk, et al., 1994), and lexical acquisition kicks off at 9 months (Fenson et al., 1994). Studies have found a use of phonotactics as a speech segmentation cue at this onset of lexical acquisition (Mattys & Jusczyk, 2001; Mattys, et al., 1999; Van Kampen, et al., 2008). This may indicate that the cues have been acquired pre-lexically. However, their experiments were not designed to disentangle whether phonotactics bootstrap the lexicon, or vice versa. As will be discussed in the following section, it might be possible to shed light on this issue by investigating whether infants focus on cues from high or from low TPs in segmentation.

2.1.2. Segmentation strategies: trough-based vs. chunk-based segmentation

Across languages, phoneme co-occurrences within words are phonotactically constrained, i.e., phoneme combinations do not occur at random, but are subject to positional or purely sequential constraints, either of a categorical or probabilistic nature. This has the consequence that within words, most phoneme transitions are of high probability of occurrence. This increases the predictability of phoneme transitions within words. However, for speech segmentation, it might not be within-word probabilities that are most informative. As compared to words, continuous speech contains many troughs in TP, namely the transitions that reside at word boundaries. Thus, continuous speech contains both information of high probability within-word transition tokens, and relatively fewer transition tokens of low probability that reside at word boundaries. This divergence from randomness presents a potential speech segmentation cue.

The question arises whether the infant uses transitions of high probability or transitions of low probability to segment speech. Saffran and colleagues (1996a) interpret their findings such that humans use a *trough-based segmentation strategy*, attending to troughs in TP to locate word boundaries in speech. However, Perruchet and Vinter (1998) point out that in fact their studies are ambiguous about which segmentation strategy was used. That is, participants might as well have chunked syllables with a high probability of co-occurrence. Just as Saffran et al., also Mattys and colleagues conclude that infants in their study use a trough-based segmentation strategy:

> The fact that infants are sensitive to between-word sequences suggests a sublexical, rather than purely lexical, basis for the effects of probabilistic phonotactics. If infants were merely drawing on information from stored sound patterns of isolated words, it is unlikely that they would demonstrate the sensitivity to between-word phonotactic probabilities observed in the present study (Mattys et al., 1999:483).

This conclusion, however, does not necessarily need to be drawn. The result that infants prefer listening to CVCCVC with a low within-word probability CC-cluster when it is pre-segmented by a subliminal pause could also be explained from a preference for high within-word probability CC-clusters to be chunked. Thus, after all, studies with infants and adults are ambiguous about whether chunking or trough-based segmentation strategies are applied when grouping the speech stream into smaller proto-word units.

Computational studies provide evidence for successful segmentation results for both chunk-based segmentation (e.g., Perruchet & Vinter, 1998; Swingley, 2005) and trough-based segmentation (e.g., Cairns, et al., 1997). Some studies

have even proposed a combined strategy of both finding troughs and chunks (e.g., Adriaans & Kager, 2010). In all these studies, a computational learner learned sequential probabilities from the input (i.e., continuous speech corpora) and subsequently used this knowledge for segmentation.

The question arises which of the cues—high TPs or low TPs—provide the better segmentation cue to the language learning infant. On the one hand, infants are flooded with information about high TPs, i.e., chunks. This could support the prediction that infants acquire high TPs more easily (e.g., Perruchet & Vinter, 1998). However, if a distribution is easier to acquire, it does not necessarily make it a better cue for segmentation. For segmentation cues it might be more important to be useable, i.e., to be targeted at specific information in the speech stream, than to be amply supported by distributional evidence. Hence, if the infant, on the other hand, had knowledge of low TPs between phonemes and applied a trough-based segmentation strategy, it would in sum have to attend to less, i.e., specific, information when processing the speech stream. For this reason, we assume that that a trough-based segmentation strategy as hypothesized in many studies (e.g., Cairns, et al., 1997; Saffran, et al., 1996a, 1996b) is advantageous.

If we were to find that infants preferred trough-based segmentation, this would support the hypothesis that continuous speech is the source of phonotactic learning. Under-representations are largely present in continuous speech from the start, but very difficult to learn from a small young infant's lexicon at the early stages of lexical acquisition. If infants prefer chunk-based segmentation, then both continuous speech and the lexicon could be the source of phonotactic learning, as over-representations are largely present in both. Investigating whether infants use trough-based segmentation or chunking strategies may thus indirectly tell us about the source of phonotactic learning.

The ambiguity whether the lexicon bootstraps phonotactics, or vice versa, can be resolved by investigating infants' segmentation of sequences whose distributional status in the lexicon differs from continuous speech. That is, Mattys et al.'s related notions "within-word probability" and "between-word probability" should be disentangled into on the one hand (high or low) probabilities that are learnable from the lexicon and on the other hand (high or low) probabilities that are learnable from continuous speech. This recognizes the logical independence of two factors for phonotactic learning and segmentation: input source (lexicon, speech) and type of segmentation strategy (chunk-based or trough-based segmentation).

2.1.3. Sequential phonotactics and non-adjacent dependencies

In order to assess whether infants acquire probabilistic phonotactic knowledge from continuous speech, it might be more informative to test sequential

phonotactic knowledge that does not refer to specific syllable positions. Consonant clusters, as used to test infants' knowledge of phonotactics for speech segmentation in Mattys and colleagues' studies, frequently occur in syllable onsets and codas at word edges. Hence, knowledge of the probability of consonant clusters is unlikely to reflect purely sequential phonotactic knowledge. Rather, it reflects positional phonotactic knowledge about specific chunks aligned with syllable onsets or codas.

Usually, cross-linguistically attested phonotactic restrictions on non-adjacent co-occurrence patterns reflect sequential phonotactics, as they are not restricted to specific positions in words. Hence, the current study will investigate whether infants use knowledge of non-adjacent dependency relations between consonants in their input. Systematic dependencies between non-adjacent phonological units are wide-spread cross-linguistically. Many languages, for example, ban sequences of identical or similar consonants—a phenomenon referred to as the *obligatory contour principle* (OCP) (e.g., McCarthy, 1986). Furthermore, constraints that require harmony between segments within words such as vowel harmony (for an overview, see Archangeli & Pulleyblank, 2007) or consonant harmony (for an overview, see Rose & Walker, 2004) are typologically very common.

Adults have been shown to use language-specific knowledge of non-adjacent dependencies between vowels (i.e., vowel harmony) (Kabak, Maniwa, & Kazanina, 2010; Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & de Gelder, 1998) and non-adjacent pairs of consonants (i.e., OCP-PLACE) as a cue for speech segmentation (Boll-Avetisyan, Chen, & Kager, submitted; Boll-Avetisyan & Kager, submitted; Kager & Shatzman, under revision). Hardly anything is known about whether infants have knowledge about non-adjacent dependencies in their native language. To our knowledge, only two studies have dealt with this question. First, van Kampen and colleagues (Van Kampen, et al., 2008) showed that Turkish-learning infants use vowel-harmony as a speech segmentation cue, which reflects distributional knowledge, because the Turkish lexicon is restricted by vowel harmony. Second, a study by Nazzi, Bertoncini, and Bijeljac-Babic (2009) showed that French-learning infants prefer listening to labial-vowel-coronal sequences rather than to coronal-vowel-labial sequences. This knowledge reflects distributions in French, where labial-vowel-coronal sequences are more frequent than coronal-vowel-labial sequences. This suggests it is likely that infants acquire probabilities of non-adjacent consonant co-occurrences and use them for segmentation, too.

Different ALL studies have shown that statistical learning is not restricted to adjacent units. Infants as young as 15 months of age as well as adults are also able to attend to distributional regularities between non-adjacent syllables (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff, de Bree, de Klerk, & Wijnen, 2011; Peña, Bonatti, Nespor, & Mehler, 2002) and phonemes (e.g., Baer-

Henney & Vijver van de, submitted). Furthermore, 11-month-olds can learn rule-like dependencies between non-adjacent vowels (Pons & Toro, 2010). Moreover, adults have been shown to be able to extract non-adjacent dependencies between phonemes from continuous artificial speech and use them as a cue for segmentation in ALL experiments (Bonatti, Peña, Nespor, & Mehler, 2005; Newport & Aslin, 2004). The literature on artificial learning studies suggests that non-adjacent dependencies are much harder to learn than adjacent dependencies (e.g., Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005).

Learning of such probabilities requires particular abstraction skills for ignoring intervening phonemes. It has even been claimed that non-adjacent dependencies between phonemes or other larger units as syllables are only learnable if they are similar at least to some degree. This means that the phonemes both need to be either consonants or both vowels, or they may even need to share the place of articulation (e.g., Newport & Aslin, 2004; Onnis, et al., 2005; Perruchet & Pacton, 2006). Such effects might be caused by a perceptual primitive that guides attention to identity between segments, which has been shown to be relevant in learning non-adjacent dependencies (e.g., Endress, Dehaene-Lambertz, & Mehler, 2007) and to affect newborns in rule-learning (e.g., Gervain, Macagno, Cogoi, Peña, & Mehler, 2008). Other studies, however, suggest that acquisition of vowel harmony (Moreton, 2008) and OCP-PLACE (Boll-Avetisyan, et al., submitted) is not a result of perceptual biases. Rather, they suggest that it is a result of an innate analytic bias on acquisition. This innate analytic bias is also responsible for patterns that are frequently found across languages. What the results of the previous studies suggest is the following: Potentially, the usefulness of under-representations of non-adjacent consonant co-occurrences (i.e., OCP-PLACE) as a speech segmentation cue may come from two factors: First, its acquisition may be facilitated by analytic bias. Second, the fact that OCP-PLACE creates troughs in TP between consonants may additionally facilitate an identification of the phonotactic restriction as a cue for speech segmentation.

Up to this point, it is unknown whether infants have knowledge of under-represented pairs of non-adjacent consonants. Furthermore, it is unclear whether infants would use such knowledge for finding word boundaries in continuous speech.

2.1.4. Research questions and hypotheses

The current study thus addresses the following research questions:

1. What is the source of phonotactic knowledge: Is phonotactic knowledge acquired from the lexicon or from continuous speech? Our hypothesis is that it is acquired from continuous speech.

2. What is the segmentation strategy infants prefer: Do infants prefer a trough-based segmentation strategy, and use phonotactic distributional cues from under-representations of phoneme pairs in the input for locating boundaries? Or do they rely on cues from over-representations of phoneme pairs for chunking phonological units in the speech stream? Or do they do both? We hypothesize that infants prefer the trough-based segmentation strategy.
3. Do infants possess phonotactic knowledge of under-represented pairs of identical non-adjacent consonants in the speech input? We hypothesize that they do, and that they use this knowledge as a cue for trough-based speech segmentation.

2.1.5. Approach

The experimental paradigm

In the current study, we will build on previous studies and combine the advantages of the studies by Mattys and colleagues (2001; 1999) testing infants' phonotactic segmentation strategies based on sequential probabilities in their native language with the advantages of a perfectly controlled ALL setting. In ALL experiments, participants are first familiarized with a highly reduced miniature artificial language, and it is consecutively tested if they have different preferences for items from this artificial language (consistent items) compared with items they should not have acquired (inconsistent items). The value of ALL studies is irrefutable, as these allow for perfectly controlled settings excluding the influence of natural linguistic (e.g., lexical) knowledge. However, learning in a perfectly controlled laboratory setting might not reflect natural language development. Distributional cues in natural languages are much messier and less reliable than in artificial languages irregularities in natural production resulting in reductions (e.g., Johnson & Tyler, 2010). In order to make claims about infants' linguistic knowledge, it is necessary to also test their natural language knowledge.

Rather than implementing statistical cues such that syllable or segment co-occurrences would vary in TP within the artificial stream, we designed languages that did not provide any statistical cues. Instead, the languages were designed to tap into infants' pre-existing knowledge about co-occurrence probabilities of consonant pairs in their native language input. Infants were thus required to "transfer" their native language knowledge to the ALL task. ALL has previously been shown to be affected by phonological knowledge of the native language in experiments with adults (e.g., Boll-Avetisyan & Kager, submitted; Finn & Hudson Kam, 2008; Onnis, et al., 2005; Vroomen, et al., 1998). Also, infants have been found to transfer some knowledge of their native

language, namely that of word order, to an artificial language (Gervain, et al., 2008). Hence, we assume that if infants have sequential phonotactic knowledge of non-adjacent consonant pairs in their native language, they will transfer this knowledge when segmenting an artificial language. We ran two experiments using artificial languages that gave no cue for segmentation besides the cues under test.

Selecting CVC sequences

We tested our hypotheses using stimuli that contained sequences of non-adjacent identical consonants. Several reasons support that decision. First, co-occurrence restrictions on vowels have a different function in speech segmentation than co-occurrence restrictions on consonants. Vowels, being carriers of prosodic information, give cues to syntax, e.g., to phrase boundaries, while consonants give cues to lexical recognition (Bonatti, et al., 2005). Hence, in view of our interest in the role of phonotactics in segmenting the stream into words, it is reasonable to investigate whether infants use knowledge of non-adjacent consonant co-occurrence restrictions. Second, because of the special role for identity in learning non-adjacent dependencies (Endress, et al., 2007; Gervain, et al., 2008), we looked at an infant's use of a non-adjacent dependencies between identical consonants, so that it would be most likely that infants had knowledge of the specific phonotactic restriction under test. Third, natural languages display two types of dependencies: those where non-adjacent consonants have to be similar within words (consonant harmony), causing a high probability of transitions between similar consonants, and those where they need to be dissimilar (OCP-PLACE), resulting in a low probability of transitions between similar consonants. For that reason it is unlikely that one pattern should be more easily acquired than the other.

In order to see what kind of distributional information the infant receives with an input of speech versus lexicon, we calculated the probability of C_1VC_2 sequences in two versions of the same corpus of Dutch infant-directed language: a segmented "lexicon" version, and a "speech" version, in which utterance boundaries were maintained, but all information about word boundaries was eliminated (Van de Weijer, 1998). As a measure, we used the *Observed*/*Expected* (*O*/*E*) ratio (Pierrehumbert, 1993). This measure compares the observed counts of consonant pairs to the counts that would be expected if consonants combined at random (Frisch, Pierrehumbert, & Broe, 2004). *O* is yielded by counting the number of C_1VC_2 sequences in a corpus or lexicon. This is divided by the *E*, i.e., the value to be expected if C_1 and C_2 were combining freely, computed as the probability that C_1 occurs in the initial position of CVC, multiplied by the probability that C_2 occurs in the final position of CVC, which is multiplied by the total number of CVC sequence tokens:

$O/E = NCvC / p(C_1) * p(C_2) * NCvC$

$O/E > 1$: over-represented; < 1: under-represented

The analysis revealed that some CvC sequences with identical Cs (e.g., /pVp/) are over-represented, and some are under-represented (e.g., /vVv/) in both Dutch infant-directed speech and the infant-directed lexicon. Other sequences are over-represented in the infant-directed lexicon, but under-represented in infant-directed speech (e.g., /sVs/) (see full matrix in Appendix A and B). This allowed us to construct stimuli with which we could test our hypothesis with regard to the role of cues from under- and over-representations in speech segmentation. For our experiments, we selected /sVs/ (Experiment 1) and /pVp/ (Experiment 2).

Prediction for Experiment 1

Experiment 1 was run with 15-month-old infants, as they have been found to have already initiated lexical acquisition (Fenson, et al., 1994). The experiment assessed whether infants would apply a trough-based segmentation strategy on the basis of distributional phonotactic knowledge acquired from continuous speech and locate a boundary in the middle of /sVs/, or whether they would chunk /sVs/ due to its over-representation in the lexicon. The artificial language consisted of two syllables starting with a /s/, which were followed by a CV syllable starting with a /x/, combined to a stream of /…sVsVxVsVsVxVsVsVxV…/. This language renders three possible segmentations: /sVxVsV/, /sVsVxV/, or /xVsVsV/. In line with our hypothesis that phonotactic cues for segmentation are induced from continuous speech, our prediction was that if infants locate a boundary between the two /s/ due to the under-representation of /sVs/ in infant-directed continuous speech. Hence, /sVxVsV/, being the only segmentation that separates the two /s/, should be the preferred segmentation. However, if our hypothesis is wrong, and infants rely on the cues from over-represented /sVs/ in the infant-directed lexicon and chunk /sVs/, then /sVxVsV/ should be the dispreferred segmentation.

2.2. Experiment 1

2.2.1. Method

Participants

Nineteen 15 month olds (6 girls, 13 boys, M_{age} = 15 months, 16 days, age range: 15;02-15;28) infants were tested. An additional eight infants (8/26 = 31%) were tested but excluded. Reasons were experimenter error (1), crying (6), and fussiness (1). All infants were growing up in monolingual standard Dutch speaking families, suffered no complications during birth, and had not had more than five ear infections in life. The parents had no history of language or literacy impairments. Parental consent was obtained. The children received a small gift for participation.

Materials

Familiarization stimuli

The artificial language consisted of 6 CV syllables, of which four started with /s/ (sV_1 = {se, so}, sV_2 = {sa, se}) and two with /x/ (xV = {xa, xo}). Each syllable was assigned to a fixed slot (see Table 1). Syllables were concatenated into a speech stream without pauses (…$sV_1sV_2xVsV_1sV_2xVsV_1sV_2xV$…), which resulted into a stream such as /…sesaxasosexosesexasesaxosesexosesaxa.../. Transitional probabilities between syllables were held constant (TP = 0.5).

Experiment	A_1	A_2	B
1 /sVsVxV/	/se/	/sa/	/xa/
	/so/	/se/	/xo/
2 /pVpVtV/	/pe/	/pa/	/ta/
	/po/	/pe/	/to/

Table 1: Syllable inventory of the artificial language used in Experiment 1 and Experiment 2.

The syllables were selected on the basis of the following considerations: We did not use lax vowels (/ɑ, ɛ, ɪ, ɔ, ʏ/), since these cannot occur in syllable or word-final position in Dutch. Furthermore, we excluded diphthongs (/ɑu, ɛi, œy/)

because we feared difficulties in perception of vowel glides. Schwa /ə/ was disregarded because its duration is significantly shorter than that of any other vowel, and because of its strong connection with prosody and morphology. As{/u, i, y/}are tense but short vowels in Dutch, we removed these from the remaining set of seven vowels {/a, e, i, o, u, y, ø/}. Of the remaining four long vowels /a, e, o, ø/, three were selected. In order to use different vowels with each syllable, we would have needed six. To control for the number of identical vowel sequences between segmentations, we needed to balance the number of syllables in our language with the vowel types. Hence, three vowels {/a, e, o/} were chosen. These three vowels were distributed in syllables with the consonants {/s, x/} after a number of controls for lexical statistics using the software PhonotacTools[2] based on the Van de Weijer corpus (1998), excluding further potential segmentation cues from language-experience. The vowels were distributed among the slots such that *O*/*E* values for pairs of non-adjacent vowels (i.e., VCV) were controlled best between segmentations. Furthermore, as positional syllable frequencies might give cues to word boundaries to the child (e.g., a certain syllable might be infrequent in initial but frequent in final position), we selected syllables such that their frequency in infant-directed speech would be even in utterance-initial, -medial and -final position. Similarly, we selected biphones that had even TP values. Finally, we controlled the string for real words of Dutch.

The syllables were concatenated to an artificial stream of speech without pauses. The duration of the stream was three minutes. It contained 216 ABA strings in total. The order of the three-syllabic strings was organized in pseudo-random order as such that no three-syllabic string would occur twice in succession. The artificial language was synthesized using the text-to-speech software MBROLA (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996). We used the Dutch female voice nl3. Pitch was flat (F_0=200). Each syllable had a duration of 278 ms with the duration of the consonants set at 135 ms, and the duration of the vowels at 143 ms. The stream faded in and out for five seconds at its edges to avoid cues for segmentation from the boundaries.

Test stimuli

Four sequences of an ABA pattern, hence /sVxVsV/, and four of an AAB pattern, hence /sVsVxV/, were chosen as test words. The patterns ABA and AAB were chosen to avoid the possibility that infants prefer items on the basis of their initial phoneme. The four items of each type were chosen to be as dissimilar of the items from the other type as possible. Thus, there were no test items that

[2] We thank Frans Adriaans for supplying us with this software.

consisted of exactly the same syllables. For example, we used /saxase/ as a test item, so we did not use /sesaxa/. All test stimuli are listed in Table 2.

Experiment	ABA	AAB
1 /sVsVxV/	/saxase/	/sesaxo/
	/sexose/	/sesexa/
	/saxoso/	/sosaxa/
	/sexaso/	/sosexo/
2 /pVpVtV/	/patape/	/pepato/
	/petope/	/pepeta/
	/patopo/	/popata/
	/petapo/	/popeto/

Table 2: Test stimuli used in Experiment 1 and Experiment 2.

Design, procedure & apparatus

We used a modified version of the Head-turn Preference Procedure (Kemler Nelson et al., 1995). Infants were first familiarized with the artificial language for three minutes. In a consecutive test phase, eight test trials were presented in a random order without allowing for more than two trials from one type (ABA vs. AAB) in succession. Each test item was repeated up to a maximum of 15 times. The dependent measure was looking time (LT) to the test items. Each infant was tested individually in a booth. Before entering the booth, the caretaker was instructed to hold the child on the lap and not to interact with it during the experiment. The caretaker sat down on a chair placed in the center of the booth. During the experiment, the caretaker listened to masking music over headphones to eliminate bias.

The three-sided test booth made of white panels had a red light at the center and green lights at the sides mounted on eye-level. A video camera recording each session was hidden above the center light. The experimenter was outside the booth in a neighboring room viewing the infant's head movements on a TV screen and controlling the start of the experiment, the onset and the end of each trial with a button box that was connected to a computer terminal.

At the beginning of the familiarization phase, the green light at the front started to blink. Once the infant looked at the light, one of the side lights started to

blink. When the infant looked towards it, the familiarization phase was initiated. The artificial language was played for three minutes from both loudspeakers. If the infant looked away from the side light for more than two seconds, familiarization with the artificial language was not interrupted, but the infant was again centered towards the light at the front panel and then to one of the side lights.[3]

After the familiarization phase, infants immediately entered the test phase. Each of the trials started with a blinking of the center light, initiated by the experimenter. Once the infant was centered, it stopped blinking and one of the side lights began to blink. Once the infant oriented towards the blinking side light, the test item started to play. Computer software selected and pseudo-randomized the stimuli (i.e., not more than two consecutive items of consistent or inconsistent trials) and kept track of the direction and duration of the infant's head turns. A trial continued until the infant looked away from the blinking side light for more than two consecutive seconds, or the trial automatically stopped after 15 repetitions of a test item. When a trial stopped, the blinking of the side light stopped, and the center light started blinking again to initiate the next trial. Data regarding direction and duration was saved on a computer data file. After the end of the experiment, a children's song was played to comfort the infant.

Besides the experimental data, data of active and passive knowledge of Dutch words and gestures were obtained. For this, we used standardized word list (N-CDI, Zink & Lejaegere, 2002). Most parents filled out the list at their homes, but a few filled it out at our lab if that was more convenient for them.

2.2.2. Results and discussion

Results

All data were recoded using the software CodeThis.[4] Sums of the looking times (LT) for each trial were calculated by a Perl script.[5] Trials during which a caretaker interacted with the infant and trials which were interrupted by the experimenter before the infant had looked away for two seconds were excluded. Since the estimated average LTs were not normally distributed, we took their natural log, and used the normally distributed logged LTs as a dependent variable (see Table 3).

---

[3] In the original procedure, the sound in the familiarization phase was only played from one side together with a blinking light (Kemler Nelson et al., 1995).

[4] We thank Luca Bonatti for supplying us with this software.

[5] Thanks to Alexis Dimitriadis for writing the script.

A multi-level model with stimulus type (ABA vs. AAB) as fixed factors and participants and items as random factors was calculated (see Appendix C). In a first model, gender and N-CDI data (passive and active word knowledge and knowledge of gestures) were added as factors. However, neither gender nor any of the obtained N-CDI data added explanation to the results. Hence, these factors were discarded from the consecutive model.

Experiment	**Test items ABA**			**Test items AAB**		
	Mean	**Ln**	**(SD)**	**Mean**	**Ln**	**(SD)**
1	5377.6	8.59	(0.042)	5659.3	8.64	(0.013)*
2 Group 1	4865.9	8.49	(0.039)	4675.1	8.45	(0.038)
2 Group 2	5064.4	8.53	(0.025)	4817.4	8.48	(0.035)

Table 3: Estimated mean looking preferences in ms and their natural logs with their standard errors in brackets in Experiments 1 (N = 15 15-month-olds) and 2 (Group 1: N = 18 15-month-olds), and Group 2: (N = 23 9-month-olds). * $p < 0.05$.

First, we looked at the variation between stimuli within infants. When the variation between stimuli of a type was extremely high, the respective observations were excluded. In Experiment 1, this was the case with three infants. Two displayed an extremely high variation between /sVxVsV/ items and one beween /sVsVxV/ items ($e_{(ij)}$ = 1.11 as opposed to $e_{(ij)}$ = 0.47 in the remaining infants). These infants had an exceptionally long LT (*SD* > 7100 ms; *log* > 8.88) and were, hence, probably sticky-fixating. Sticky fixation occurs when infants have difficulty interrupting gaze to a visual attractor (Hood, 1995). Since it is unrelated to the verbal stimulus, it is generally used as an exclusion criterion. In the following experiment in this study, we applied the same exclusion criterion. For the sticky-fixating infants, a dummy variable was kept in the model.

A Wald-test (Quené & van den Bergh, 2004) revealed a main effect of stimulus type (ABA vs. AAB). Infants listened significantly longer to /sVsVxV/ than to /sVxVsV/ items ($\chi^2 = 5.6$, $df = 1$, $p = .018$) with a "medium" effect size (Cohen's $d = 0.55$) (Cohen, 1977) (see Figure 1). Ten infants had longer estimated mean LTs for /sVsVxV/ items, and six for /sVxVsV/ items. As we do not have valid observations of two infants of /sVxVsV/, of one infant of //sVsVxV/, and of one infant of any item, we cannot make a reliable statement for four infants.

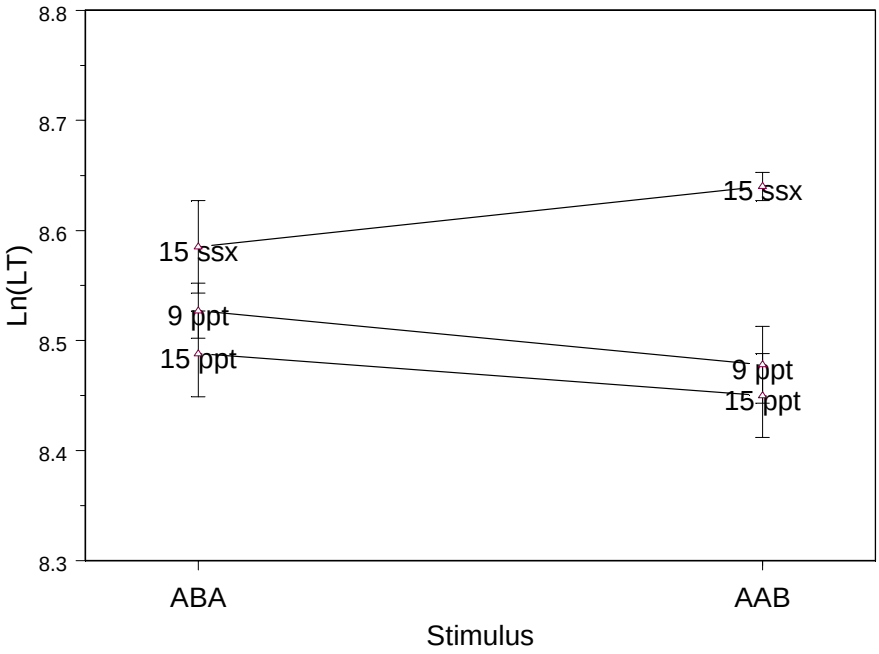

Figure 1: Logged estimated mean looking times and their standard errors for the 15-month olds in Experiment 1 (ssx-sxs) and both age groups of infants (9 months / 15 months) in Experiment 2 (ppt-ptp).

Discussion

It was predicted that infants would segment the artificial language by placing boundaries between consecutive /s/, because /sVs/ is under-represented in infant-directed continuous speech (O/E = 0.57). The significant preference for AAB (i.e., /sVsVxV/) over ABA (i.e., /sVxVsV/) stimuli in the test phase indicates that infants must have segmented the stream. As older infants commonly display novelty preferences (Hunter & Ames, 1988), and artificial language segmentation experiments usually provoke novelty preferences even at younger ages (e.g., Saffran, et al., 1996a), the results can be taken to indicate that the infants have segmented the stream into /sVxVsV/ sequences and listen longer towards /sVsVxV/ stimuli, which do not obey the segmented pattern.

In contrast to the under-representation in continuous speech, /sVs/ is over-represented (O/E = 1.55) in the lexicon. Hence, if infants had used their lexical knowledge, they would have chunked /sVs/ and had displayed a novelty

preference for /sVxVsV/ over /sVsVxV/ in the test phase. Thus, the result suggests that the source of learning phonotactic cues for speech segmentation is continuous speech rather than the lexicon.

Our hypothesis was that phonotactic knowledge revealed in Experiment 1 is a result of perceptual learning from the input. However, we also need to be concerned with the question whether the infants in our experiment have really displayed language-specific knowledge. An alternative explanation is that there is a universal auditory bias against groups of identical consonants, possibly reflected in OCP restrictions across languages (e.g., McCarthy, 1986) or even a more general cognitive bias against an AA pattern. Similarly, it might as well be that infants have a general cognitive preference for identity in edges (i.e., ABA >> AAB). Furthermore, we wanted to test whether infants attune to chunking cues from over-representations in continuous speech. This question was put under test by means of the chunking cue /pVp/ in Experiment 2.

We used an artificial language as in Experiment 1, with /s/ being replaced by /p/ and /x/ by /t/, resulting in a stream of /…pVpVtVpVpVtVpVpVtV…/. Pairs of /pVp/ are over-represented in Dutch infant-directed speech and the infant-directed lexicon. If segmentation preferences in Experiment 1 were purely due to a universal preference for chunking identical consonants, then the prediction would be that as infants preferred AAB (i.e., /sVsVxV/) items in the test phase of Experiment 1, they will also prefer AAB (i.e., /pVpVtV/) items in Experiment 2. Such a result would allow us to draw the conclusion that preferences in Experiment 1 are not due to perceptual learning, but to universal perceptual preferences. If, however, preferences in Experiment 1 were not due to universal perceptual biases, two outcomes in Experiment 2 are possible. First, perhaps infants possess knowledge that /pVp/ sequences are over-represented in their input. Then they should chunk /pVp/ in segmentation, and should, as opposed to Experiment 1, display a novelty preference for ABA (i.e., /pVtVpV/) test items in the test phase of Experiment 2. This result would suggest that infants learn both under and over-representations for speech segmentation. The other possible outcome is that infants prefer neither AAB (i.e., /pVpVtV/) nor ABA (i.e., /pVtVpV/) items. This outcome would suggest that infants in Experiment 1 did not rely exclusively on a universal perceptual segmentation strategy, but instead learned up the trough-based segmentation cue from continuous speech input.

To investigate whether infants use phonotactic chunking cues at a lexical or pre-lexical age, we tested a group of pre-lexical 9-month olds and a group of 15–month-olds. If cues for a chunk-based segmentation strategy can be acquired from continuous speech, then 9-month olds should already rely on the chunking cue /pVp/ for segmentation. However, if they can only be acquired after the onset of lexical acquisition, such that it is deduced from lexical entries, then it is predicted that only 15-month olds will show a segmentation preference.

2.3. Experiment 2

2.3.1. Method

Participants

Participants were twenty-five 9-month-old (10 girls, 15 boys, *M*age = 9 months, 14 days, age range: 9;00 – 9;29) and 20 15-month-old Dutch infants (9 girls, 10 boys, M_{age} = 15 months, 2 days, age range: 14;20 - 15;19). Two additional 9-month-old infants were tested but excluded due to crying (2/27 = 7%), and an additional three 15-month-old babies (3/23 = 13%) were tested but not included because of crying (2) and fussiness (1). The infants were selected by the same criteria as in Experiment 1.

Materials

Familiarization stimuli

For the artificial language, 6 CV syllables were selected, of which four started with /p/ (pV_1 = {pe, po}, pV_2 = {pa, pe}) and two with /t/ (tV = {ta, to}). Each syllable was assigned to a fixed slot (see Table 1). Syllables were concatenated into a speech stream without pauses (…$pV_1pV_2tVpV_1pV_2tVpV_1pV_2tV$…), which resulted into a stream such as /…pepatapopetopepetapepatopepetopepata..../. The vowels and their assignment to slots from the artificial language in Experiment 1 were maintained for two reasons. First, we wanted to keep the languages as similar as possible. Secondly, the vowels in Experiment 1 were selected carefully: They are phonetically distinguishable, have similar spoken durations in natural speech, and the *O*/*E* values of non-adjacent pairs of vowels in the artificial language were balanced and could therefore not give a cue for segmentation. An analysis of positional syllable frequencies and biphone TPs in infant-directed speech utterances in the Van de Weijer corpus (1998) revealed that these factors would not amount to cues for speech segmentation. Also, the artificial stream did not contain any real words of Dutch that might be known to infants. The only word in the stream that might be known to infants was *Opa* ‘grandfather’. Randomization and synthesis were the same as in the previous experiment.

Test stimuli

The four ABA, hence /pVtVpV/ and four AAB, hence /pVpVtV/ sequences that were chosen as test words were identical to the test items in Experiment 1 with /s/ being replaced by /p/ and /x/ by /t/. All test items are listed in Table 2.

Design, procedure & apparatus

Design, apparatus and procedure were identical to Experiment 1.

2.3.2. Results and discussion

Results

As in the previous experiment, all data were recoded and all trials during which the caretaker interacted with the infant or which were interrupted too early by the experimenter were excluded. Results were analyzed using a mixed-effect linear regression model with stimulus type (ABA vs. AAB), age (9 vs. 15 months) and experiment (/sVsVxV/ vs. /pVpVtV/) as fixed factors and participants and items as random factors. Logged LTs were the dependent variable. Based on the same exclusion criterion as in Experiment 1, two 9-month-olds and one 15-month-old were excluded. For all these infants, a dummy variable was kept in the model.

The result of a Wald-test (Quené & van den Bergh, 2004) showed that there was no main effect of stimulus type. Furthermore, there was no main effect of age, and no significant interaction of type and age (see Figure 1 and Table 3). Comparing LTs of the 15-month-olds in Experiment 1 and 2 revealed a main effect of experiment (/sVsVxV/ vs. /pVpVtV/): The LTs of the 15-month-olds in Experiment 1 were overall longer than in Experiment 2 ($\chi^2 = 31.37$, $df = 1$, $p < .001$). This effect was due to an interaction of stimulus type and experiment ($\chi^2 = 9.1$, $df = 1$, $p < .003$): 15-month-old infants listened significantly longer to /sVsVxV/ than to /sVxVsV/, /pVpVtV/ and /pVtVpV/ items. The model was calculated without N-CDI scores and gender as factors of the model, since they did not reach significance neither in any of the interactions nor as main effects in a preliminary model.

Discussion

Experiment 2 using a continuous stream of /…pVpVtVpVpVtVpVpVtV…/ was intended to, first, rule out that segmentation preferences in Experiment 1 were purely due to a general cognitive or a perceptual bias against chunking identical consonants, and, secondly, to test, whether infants would be able to use cues from over-representations in their input for a chunk-based segmentation strategy.

First, if infants had merely avoided sequences of identical consonants in Experiment 1, then such a preference would have rendered a preference for locating a boundary within /pVp/ as well as within /sVs/. Then infants should have segmented a continuous stream of /…pVpVtVpVpVtVpVpVtV…/ into sequences of /pVtVpV/. However, this was not the case. The absence of a preference for AAB or ABA stimuli in Experiment 2 with /pVp/ trials contrasts

with those of Experiment 1. Overall, infants listened longer to /sVsVxV/ than to /sVxVsV/, /pVpVtV/ and /pVtVpV/ items. This seems to indicate that in the test phase, only /sVsVxV/ items caused a novelty effect. Hence, whereas in Experiment 1, infants segmented the stream into /sVxVsV/ sequences, it seems that in Experiment 2 they did not segment the stream and were thus not attracted by either ABA nor AAB stimuli in the test phase. The different result patterns between the two experiments suggest that infants do not have a universally biased preference for putting boundaries between non-adjacent identical consonants. Therefore, segmentation performance in Experiment 1 must have been aided by language-specific knowledge acquired from distributional properties of the language input.

Second, infants did no display a novelty preference for /pVtVpV/. If they had displayed a novelty preference, this would have suggested an application of a chunk-based segmentation strategy due to knowledge that sequences of /pVp/ are over-represented in the input. Neither the 9-month-olds nor the 15-month-olds distinguished between ABA and AAB items in the test phase. This result might be surprising in light of previous studies, in which infants displayed knowledge of non-adjacent consonant dependencies in their native language at 10 months (Nazzi, et al., 2009). In fact, the over-representation of /pVp/ should be of particular salience to Dutch infants, as /pVp/ occurs frequently in their input as a result of a large part of the Dutch infant-directed lexicon containing /pVp/ sequences (e.g., *papa* "daddy", *pap* "porridge", *pop* "doll", *poep* "poo", *piepen* "to beep", *puppy* "little dog"). Hence, we can assume that Dutch infants, especially the 15-month-olds, who have initiated lexical acquisition and have access to word types with /pVp/, know of the over-representation of /pVp/. This implies that though phonotactic distributions are easy to acquire, they do not necessarily make a good cue for segmentation. As our results suggest, infants do not as readily use this knowledge as a cue for chunk-based segmentation as they use cues for trough-based segmentation.

2.4. General discussion

The current study aimed at testing three hypotheses. First, we hypothesized that phonotactic cues for speech segmentation are induced from continuous speech rather than from lexical knowledge. The hypothesis was motivated by the observation that infants would benefit from knowledge of phonotactic cues for speech segmentation in lexical acquisition. Second, we hypothesized that infants use a trough-based instead of a chunk-based segmentation strategy. This hypothesis was motivated by the following observation: Continuous speech contains fewer troughs in TP than highly probable phoneme transitions because

speech is composed of words, and phoneme pairs that occur within words tend to be over-represented. For that reason, troughs may provide a more useful cue to speech segmentation than peaks in TP. If infants apply a trough-based segmentation strategy, they need to attend to less information when processing the speech signal than if they apply a chunk-based segmentation strategy. Third, we hypothesized that infants use distributional knowledge of under-represented pairs of non-adjacent consonants for segmentation. The hypotheses were addressed in two ALL experiments assessing whether infants exploit sequential phonotactic knowledge of under- and over-represented pairs of non-adjacent phonemes in their ambient language in speech segmentation.

In this study, infants did locate boundaries within pairs of identical consonants that are under-represented in infant-directed speech (i.e., /sVs/) (Experiment 1), but did not chunk highly probable sound sequences (i.e., /pVp/) when listening to a continuous stream of an artificial language (Experiment 2). The result suggests that infants start out segmenting the continuous stream of language by locating improbable sound sequences. Furthermore, it indicates that in infancy, lexical knowledge is not the primary source from which speech segmentation cues are derived. If the (proto-)lexicon were of major relevance for deducing cues for speech segmentation, then it would have been expected that infants rely on the chunking cues from over-representations in the lexicon.

It is striking that infants do not use over-represented /pVp/ as a cue for chunking in segmentation, given its high salience with support of many lexical types entailing /pVp/ sequences in the infant-directed lexicon. With an O/E = 4.48 in infant-directed speech, which expresses that /pVp/ is about four times more likely than expected, the chunking cue in Experiment 2 should have even been more informative as a segmentation cue than /sVs/ in Experiment 1, which with an O/E = 0.57 had only about half of the expected frequency. So, if it were just a matter of distributions, then /pVp/ should give the better cue. Hence there must be a different explanation for why infants turn to under-representations. One account goes as follows: The speech signal contains more within-word than between-word phoneme-to-phoneme transitions. In general, TPs within words are high, and TPs at word boundaries are low. For the reason that the speech signal contains fewer troughs in TP, we stated the hypothesis that though-based segmentation is advantageous, because infants need to attend to less information when processing the speech stream.

There is an alternative account for why /sVs/ is a better cue for speech segmentation than /pVp/. Cross-linguistically, many languages display an under-representation of consonants with shared place, a phenomenon that has been accounted for by a universal phonotactic constraint OCP-PLACE (e.g., McCarthy, 1986). Fewer languages display patterns of consonant harmony; although it needs to be mentioned that in many languages, sequences of identical consonants (such as /sVs/ and /pVp/) have been found to escape the more general effects of OCP-PLACE (e.g., Kawahara, Ono, & Sudo, 2006; Yip, 1989). Given

the predominance of OCP-PLACE across languages and the proposal that OCP-PLACE has grounds in innate biases on perception, it could be argued that infants have not acquired the segmentation cue /sVs/ from the input language, but instead were innately biased to insert a boundary between two identical consonants. This alternative account, however, is not supported by the fact that infants behaved differently in the two experiments. If segmentation preferences in Experiment 1 were attributable to a universal preference for always inserting a word boundary between two identical consonants, then they should have done so in Experiment 2, too. If at all, it could be argued that the predisposition to insert a boundary between two identical consonants was overwritten by knowledge of the over-representation of /pVp/. This suggests that knowledge of non-adjacent consonant dependencies is acquired from the distributions in the input, underscoring the role of a statistical learning mechanism in language acquisition. Still, a cognitive bias on acquisition might have helped the infant to acquire the under-representation of /sVs/. Future studies can shed more light on this by testing whether infants use under-representations of pairs of non-identical consonants (e.g., /sVp/) for speech segmentation.

The result that infants do not use knowledge of over-represented /pVp/ as a cue for segmentation does not necessarily imply that over-representations are harder to acquire than under-representations nor does it imply that infants lack knowledge of over-represented /pVp/. Given earlier findings by Nazzi and colleagues (Nazzi, et al., 2009) who found 10-month old infants to have knowledge of an over-representation of non-adjacent labial and coronal consonant, we expect that infants know that /pVp/ is over-represented in their input. Our results only show that infants do not as readily exploit knowledge of an over-representation of a pair of non-adjacent consonants as a cue for speech segmentation as they exploit knowledge of an under-representation. Moreover, previous studies have found that infants and adults use chunking cues from an over-representation of non-adjacent pairs of vowels (i.e., vowel harmony) for speech segmentation (Suomi, et al., 1997; Van Kampen, et al., 2008; Vroomen, et al., 1998). So, how can it be explained that infants are able to use a chunking cue from vowel harmony for speech segmentation, but not a chunking cue from a very salient over-representation of a pair of identical consonants?

The reason may relate to the proposal that vowel co-occurrence patterns and consonant co-occurrence patterns might have different functions in early speech segmentation. Vowels, being carriers of prosodic information, give cues to syntax, for instance to phrase boundaries, while consonants give cues to lexical recognition (Bonatti, et al., 2005). Hence, it might be that vowel harmony is a better cue for identifying larger chunks of phrase constituents, and consonant co-occurrence probabilities are better used to identify lexical units.

The result of this study has interesting implications for models of language acquisition. First, it suggests that within-word phoneme sequences with a high

probability of occurrence are not the primary source for developing knowledge about phonotactic cues for speech segmentation. This is a challenge for models that assume that probabilistic phonotactic knowledge is purely derived from stored chunks in the mental lexicon (e.g., Jusczyk, et al., 1994; Pierrehumbert, 2003). More likely, infants acquire their knowledge about which cues lead to word boundaries from continuous speech, as previously suggested in computational studies (e.g., Adriaans & Kager, 2010; Brent & Cartwright, 1996; Cairns, et al., 1997; Perruchet & Vinter, 1998; Swingley, 2005) and ALL studies with humans (e.g., Saffran, et al., 1996a, 1996b).

Second, it indicates that in speech segmentation, infants more readily use knowledge about rarely co-occurring consonants for locating word boundaries than they use knowledge about commonly co-occurring consonants for chunking. Although infants are flooded with information about the high probability of certain consonant sequences in their input, there is little use for them in segmentation.

It might be argued that, although we could not find evidence for a use of chunking cues from the infant-directed lexicon, cues for trough-based segmentation are deduced from initial lexical representations. This, however, seems improbable, as continuous speech provides a more extensive source for extracting patterns of under-representations. So, if infants' speech segmentation performance receives stronger facilitation from cues for trough-based than for chunk-based segmentation, it is not clear why they should first acquire lexical knowledge. It seems more efficient to acquire low probability phonotactics bottom-up from speech and directly use them as a bootstrap to lexical acquisition.

An issue that remains open for future research is how non-adjacent dependencies are learned. The correlation between infants' behavior and distributions in the input suggests statistical learning. However, as infants must abstract over intervening vowels when acquiring them, statistical learning by itself is not sufficient for developing such knowledge. It will need to be investigated whether abstraction over intervening segments in non-adjacent dependency learning can only occur when the non-adjacent units are identical or similar, or if it is also possible when they are different. There is evidence to suggest that this abstraction over different non-adjacent units will be more difficult. Non-adjacent dependencies between phonemes often involve identity or similarity between co-occurring segments (e.g., Archangeli & Pulleyblank, 2007; McCarthy, 1986). Furthermore, previous work with adults and neonates found that identity between units is highly salient to listeners. Attending to co-occurrences of identical units has been suggested to be a perceptual primitive (e.g., Endress, et al., 2007; Gervain, et al., 2008). In a similar vein, ALL studies almost always only find successful learning of non-adjacent dependencies, if the non-adjacent units are identical or similar (e.g., Newport & Aslin, 2004; Onnis, et al., 2005; Perruchet & Pacton, 2006). If infants need identity between

segments to perceive segmentation cues from remote dependencies, then our results cannot be explained by only statistical learning itself. Future work will have to assess whether this is the case to further support the interpretation that two mechanisms are at play in acquiring non-adjacent dependencies in speech segmentation: a statistical learning mechanism for attending to language-specific patterns (e.g., Saffran, et al., 1996a, 1996b), and a second mechanism, such as a rule-learning mechanism (e.g., Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Peña, et al., 2002) for attending to language-general patterns such as identity between co-occurring units in speech.

2.5. Conclusion

In sum, the results of the current study suggest that phonotactics bootstrap the lexicon, and not vice versa. Infants use the continuous stream of speech that surrounds them to induce phonotactic knowledge of under-represented phoneme sequences. This knowledge aids them in detecting word boundaries, such that they can segment the speech stream into proto-words to take the first step in lexical acquisition.

2.6. References

Adriaans, F., & Kager, R. (2010). Adding Generalization to Statistical Learning: The Induction of Phonotactics from Continuous Speech. *Journal of Memory and Language, 62*, 311-331.

Archangeli, D., & Pulleyblank, D. (2007). Harmony. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 353-378). Cambridge, England: Cambridge University Press.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science, 9*, 321-324.

Baer-Henney, D., & Vijver van de, R. (submitted). Substance and Amount of Exposure in the Acquisition of Alternations.

Boll-Avetisyan, N. A. T., Chen, A., & Kager, R. W. J. (submitted). OCP-Place for speech processing: innate or acquired knowledge? The case of Mandarin Chinese.

Boll-Avetisyan, N. A. T., & Kager, R. W. J. (submitted). OCP-PLACE in speech segmentation.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations. The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science, 16*, 451-459.

Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences, 3*, 294-301.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93-125.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111-153.

Chambers, K. H., Onishi, K. E., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*, B69-B77.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221-268.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). *The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes.* Paper presented at the Proceedings of The Fourth International Conference on Spoken Language Processing (ICSLP'96) 3.

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition, 105*, 577-614.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, Serial No. 242, 59*(5), i-185.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*, 477-499.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics, 54*, 287-295.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory 22*, 179-228.

Gambell, T., & Yang, C. (2005). *Mechanisms and Constraints in Word Segmentation.*Unpublished manuscript.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105*, 14222-14227.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*, 183-206.

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics, 47*, 359-382.

Hood, B. M. (1995). Shifts of visual attention in the human infant: a neuroscientific approach. In L. Lipsett & C. Rovee-Collier (Eds.), *Advances in Infancy Research* (Vol. 9). Norwood, NJ: Ablex.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 69-95). Westport, CT, US: Ablex Publishing.

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language, 44*, 548-567.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339-345.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V., & Jusczyk, A. M. (1993). Infants‘ sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402-420.

Jusczyk, P. W., Hohne, E., & Baumann, A. (1999). Infants’ sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics, 61*, 1465-1467.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word-segmentation in English-learning infants. *Cognitive Psychology 39*, 159-207.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants’ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630-645.

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Journal of Laboratory Phonology, 1*, 207-224.

Kager, R., & Shatzman, K. (under revision). *Similarity Avoidance in Speech Processing*.Unpublished manuscript.

Kawahara, S., Ono, H., & Sudo, K. (2006). Consonant Cooccurrence Restrictions in Yamato Japanese. *Japanese/Korean Linguistics, 14*, 27-38.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The Head-Turn Preference Procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111-116.

Kerkhoff, A., de Bree, E., de Klerk, M., & Wijnen, F. (2011). Implicit learning and dyslexia: Non-adjacent dependency learning in infants at familial risk of dyslexia. In N. Danis, M. K. & H. Sung (Eds.), *BUCLD 35: Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 358-370). Sommerville (Mass): Cascadilla Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule Learning by Seven-Month-Old Infants. *Science, 283*, 77–80.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91-121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

McCarthy, J. J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry, 17*, 207-263.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology 25*(1), 83-127.

Morgan, J. L., & Demuth, K. (1996). *Signal to Syntax. Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, NJ: Erlbaum.

Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *Journal of the Acoustical Society of America, 126*, 1440-1144.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology 48*, 127-162.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*, 225-237.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604-607.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233-238.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246-263.

Pierrehumbert, J. B. (1993). *Dissimilarity in the Arabic verbal roots.* Paper presented at the Proceedings of the North East Linguistics Society.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge MA: MIT press.

Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition, 116*, 361-367.

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication, 43*, 103-121.

Rose, S., & Walker, R. (2004). A Typology of Consonant Agreement as Correspondence. *Language, 80*, 475-531.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*, 606-621.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*, 484-494.

Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86-132.

Thiessen, E. D., & Saffran, J. R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. *Developmental Psychology, 39*, 706-716.

Van de Weijer, J. (1998). *Language-Input for Word Discovery*: P.h.D. Thesis. Max-Planck Series in Psycholinguistics 9.

Van Kampen, A., Parmaksiz, G., van de Vijver, R., & Höhle, B. (2008). Metrical and statistical cues for word segmentation: Vowel harmony and word stress as cues to word boundaries by 6- and 9-month old Turkish learners. In A. Gavarró & M. J. Freitas (Eds.), *Language Acquisition and Development* (pp. 313-324). Newcastle: Cambridge Scholars Publishing.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133-149.

Weissenborn, J., & Höhle, B. (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Philadelphia, PA: John Benjamins.

Yip, M. (1989). Feature Geometry and Co-occurrence Restrictions. *Phonology, 6*, 349-374.

Zink, I., & Lejaegere, M. (2002). *N-CDI Woordenlijst [N-CDI word list]*. Leusden: Acco.

APPENDIX A

	p	b	m	f	v	w	t	d	n	s	ʃ	x	z	ʒ	l	r	j	h	ɣ	k	g	ŋ	C2
p	**4,48**	0,26		0,39	0,39	0,35	**1,24**	0,78	0,54	0,85	0,80	0,29	0,99		1,88	0,71	0,71	0,65	0,34	1,87		0,90	
b	0,18	1,27	0,85	0,34	1,05	0,91	1,12	0,98	2,70	0,50	0,37	0,07	0,76		0,57	0,55	0,37	0,40	1,28	0,63	1,00	0,59	
m	0,25	0,27	0,59	0,05	0,36	0,44	1,40	0,59	0,79	0,99	0,13	0,56	0,39		1,07	2,48	1,53	0,38	0,57	0,62		0,12	
f	0,67	0,19	0,23	1,28	0,65	0,25	1,16	0,38	1,57	0,42	0,08	0,15	1,05	3,89	1,96	1,00	3,26	0,17	1,32	0,94		0,43	
v	0,10	0,82	0,18	1,46	0,39	0,77	0,12	1,04	1,82	1,07		0,25	1,22		1,60	2,39	0,51	0,67	0,37	0,33		0,86	
w	0,07	0,32	0,48	0,34	0,25	0,70	1,56	0,72	0,53	1,37		1,20	0,96		2,70	1,41	0,32	0,96	0,98	0,51	0,54	0,33	
t	**1,09**	0,76	0,66	1,22	1,40	0,90	0,85	0,71	1,03	1,47	0,13	1,73	0,65	3,68	0,93	0,92	0,19	0,91	1,59	1,24	0,37	0,68	
d	0,41	1,60	0,78	0,24	0,61	0,85	1,38	0,79	1,01	1,01	0,13	1,19	1,67	0,97	0,36	1,46	0,66	1,13	0,38	0,59	0,67	1,94	
n	0,59	0,86	0,96	0,49	0,78	0,94	2,01	1,69	0,69	0,40	0,02	2,77	0,82	0,75	0,48	0,72	0,77	0,57	0,86	0,73		0,23	
s	2,63	0,50	0,93	1,27	3,17	0,29	1,27	0,60	1,41	**0,57**	0,07	**0,54**	0,31	2,34	0,94	0,41	0,23	0,50	0,54	1,27		0,67	
ʃ	0,38	1,58	2,52	0,88	1,55	1,36	0,36	2,45	0,90	0,46		1,28	2,04		1,02	0,19	0,53	4,19		1,33		0,86	
x	0,76	0,81	0,66	1,33	1,96	2,51	1,58	1,08	1,58	**0,57**		0,19	1,75		0,87	0,13	1,34	0,91	1,96	0,47		0,95	
z	0,21	0,74	0,88	0,18	1,14	2,04	1,06	1,18	1,18	0,37		2,85	1,80		1,41	0,10	1,39	1,54	1,20	0,90	9,76	0,74	
ʒ			1,80		1,22		0,21		0,56	3,41			0,88		0,47	1,20		1,22		3,72			
l	2,54	0,89	1,34	2,49	1,23	0,89	0,55	0,54	0,56	0,95	1,99	0,87	0,82	1,93	0,16	0,45	0,53	0,53	2,63	3,34	0,53	2,66	
r	1,30	0,81	1,04	1,60	2,02	0,57	0,90	0,76	0,96	1,32	0,48	1,58	0,44	4,11	0,62	0,11	0,57	0,41	2,81	1,96	1,71	3,50	
j	0,43	2,23	1,45	0,27	0,98	2,43	0,20	2,69	0,39	2,78	9,98	0,46	1,90	0,74	0,49	0,17	2,57	2,14	0,60	0,49	0,82	0,34	
h	3,69	1,92	0,59	4,54	0,91	0,49	0,49	0,82	0,29	0,32	0,03	0,28	0,59		2,32	1,23	3,20	2,68	0,30	0,22		1,09	
ɣ	0,24	1,14	0,72	0,19	2,72	1,76	0,68	1,95	1,66	0,27		0,20	1,27		2,22	0,70	0,59	1,69	0,63	0,48		1,11	
k	0,59	0,51	3,85	0,53	0,69	0,51	0,21	0,40	1,43	0,61	0,03	0,16	0,26	0,45	0,69	1,44	0,09	0,50	0,08	1,99	0,25	0,17	
g		1,85	0,70						1,76		4,74	0,79				0,33		3,32		0,56			
ŋ	0,46	1,09	0,90	0,54	0,98	1,65	0,65	1,35	0,68	0,36		0,08	1,18		0,48	3,37	0,39	2,13		0,39		0,52	
C1																							

The table represents O/E ratios of C_1C_2 in CvC strings in the van de Weijer corpus (1998) after eliminating all word boundaries resulting in a corpus of continuous speech. Missing values are not in the list. Pairs of identical Cs are shaded in gray. Pairs of Cs occurring in the artificial languages in Experiment 1 and 2 are in bold. Values including C_2 = /j/ and C_1 = ʃ/ need to be taken with caution, as the infant's name is *Josje* /joʃje/. Possibly, /mm/ is under-represented because the corpus only entails infant-directed speech by father.

APPENDIX B

	p	**b**	**m**	**f**	**v**	**w**	**t**	**d**	**n**	**s**	**ʃ**	**x**	**z**	**ʒ**	**l**	**r**	**j**	**h**	**ɣ**	**k**	**g**	**ŋ**	**C2**
p	**2.49**		0.26	1.13	0.3	0.34	**1.03**	0.6	0.32	1.05	3.94	0.66	2.01		1.17	1.68	1.97	1.06	0.42	1.7		1.53	
b	0.28	2.5	3.04	0.4	2.37	2.69	1.1	2.27	0.36	0.61	0.93	0.16	1.67		0.93	1.72	1	4.55	1.78	1.38	2.75	0.72	
m	0.66	0.57	0.51		0.22	0.76	0.72	3.37	0.62	1.66	2.41	0.57	0.99		0.82	1.39	2.68	0.78	1.23	1.38		0.37	
f	0.97	1.27	0.28	0.92		2.24	1.5	1.3	1.1	0.14	2.14		3.28		0.87	1.73	0.92		2.04	0.99		0.55	
v	0.78	0.31	1.16	2.65	0.24	0.27		1.24	0.65	0.6		0.26	1.05			3.62	0.22	0.83	0.65	0.19		1.32	
w	0.13	0.98	1.1	1.42		0.43	0.58	0.75	0.56	0.57		2.49	2.53		1.1	2	3.53	1.34	3.93	0.69	4.85	2.13	
t	**0.46**	0.85	0.64	0.85	1.08	0.56	0.22	0.33	2.5	0.32	0.36	1.68	0.09	3.37	0.37	0.98	0.08	0.29	0.68	0.58		0.51	
d	1.5	1.65	2.37		0.18	0.41	0.33	0.48	0.47	0.25	1.59	2.98	1.01		0.54	3.12	2.2	0.64	1.13	1.02	2.33	2.35	
n	0.73	0.14	1	0.39	0.32	0.96	6.05	0.83	0.09	0.4		0.69	0.23		0.13	0.31	1.08		1.02	0.25		0.18	
s	1.37	2.69	0.75	2.92	0.52	1.77	0.65	1.03	0.53	**1.55**		**1.51**	0.57		0.87	0.56	1.93		0.72	2.5		1.46	
ʃ	2.23		3.9	6.33			0.3		1.5	0.25					1.19	0.33				1.36		1.9	
x	2.23	7.01	1.17	1.27	3.8	5.54	0.58	3.75	0.41	**0.63**		0.3	6.92		0.98	0.61	1.51	13.3	1.12	1.14		0.61	
z	1.41		1.55		2.86	1.62	0.96	0.94	0.68	0.83		3.12	0.4		0.72	0.94	0.33	1.26	1.73	1.79	13.6	1	
ʒ			5.46						0.84	1.37					0.83	1.85				1.9			
l	1.15	0.97	0.65	1.64	1.08	0.57	0.55	0.55	0.17	3.9	1.99	0.88	0.92	1.7	0.04	0.24	0.62		1.09	1.58		1.59	
r	1.47	1.63	1.82	3.01	2.24	2.22	0.89	1.29	0.43	0.82	0.3	2.54	1.09	5.71	0.39	0.18	2.99		2.6	3.25	1.78	2.35	
j	0.47	0.47	1.04	0.67		1.23	0.19	0.71	0.37	3.94	10.9	0.26	1.6		0.19	0.77			0.5	0.29	4.59	1.01	
h	1.51	0.14	0.31	0.6	0.64	0.24	0.15	0.84	0.21	0.14	0.23	0.23	0.59		6.06	0.41	0.4		0.29	0.17		0.84	
ɣ	0.77	1.01	0.45		6.27	1.78	0.77	2.06	1.18	0.28					1.51	1.37	1.45	2.75	1.08	0.47		1.32	
k	2.12	1.04	2.01	1.51	1.34	0.91	0.22	1.24	1	0.47		0.39	0.89		0.94	2.17		0.94		1.4		0.75	
g									4.21														
C1																							

The table represents O/E ratios of C_1C_2 in CvC strings within words in the van de Weijer corpus (1998). Missing values are not in the list. Pairs of identical Cs are shaded in gray. Pairs of Cs occurring in the artificial languages in Experiment 1 and 2 are in bold.

APPENDIX C

Multilevel Model

Looking times are nested both within individuals and within items, which means that there is variation within individuals and within items. Hence, a multilevel model is appropriate for the analysis of the results. Failing to take the different variance components into account will result in an underestimation of the variance and hence the testing statistics will be too optimistic (the null-hypothesis is rejected although the data in fact do not support this conclusion). In order to test the hypothesis we define several dummy variables, which are turned on if a response is observed in the respective condition and are turned off otherwise. Let Y(ij) be the response on item *i* (i = 1, 2, …, I) of individual *j* (j = 1, 2, …, J), and 9m_ptp_AAB$_{(ij)}$, 15m_ptp_ABA$_{(ij)}$, 15m_ptp_AAB$_{(ij)}$, 15m_sVxVsV_ABA$_{(ij)}$, and 15m_sVxVsV_AAB$_{(ij)}$ be dummy variables indicating which age or participants (9m / 15m), language (pVtVpV / sVxVsV) and stimulus type (ABA / AAB) the response refers to. dummySD > 7100$_{(ijk)}$ refers to all infants that are excluded due to an exceptionally high inter stimulus variation ($e_{(ij)}$ = 1.11 on average as compared to $e_{(ij)}$ = 0.44 in all other infants). The interaction between type (ABA / AAB) and language (pVtVpV / sVxVsV) can be estimated by defining combinations of the respective dummies. Analogue to analysis of variance a saturated model with both main and interaction effects can be written as:

$$
\begin{aligned}
Y(ij) = {} & \beta_0 * \text{Constant} + \beta_1 * \text{9m_ptp_AAB}_{(ij)} + \\
& \beta_2 * \text{15m_ptp_ABA}_{(ij)} + \beta_3 * \text{15m_ptp_AAB}_{(ij)} + \\
& \beta_4 * \text{15m_sxs_ABA}_{(ij)} + \beta_5 * \text{15m_sxs_AAB}_{(ij)} + \\
& \beta_6 * \text{dummySD>7100}_{(ij)} + [e_{(ij)} + u_{i0} + v_{0j}].
\end{aligned}
$$

The model above consists of two parts: a fixed and a random part (between square brackets). In the fixed part, Constant (i.e., the intercept) represents the mean of 9-month-olds responses on ABA in the pVtVpV-language, and the other effects represent deviations from this average. So, the looking time by 9-month-olds to AAB items in the pVtVpV-language is (Constant + β_1), and the average of 15-month-olds to ABA items in the pVtVpV-language is (Constant + β_2), etc.

In the random part three residual scores are defined: $e_{(ij)}$, u_{i0} and v_{0j}. The last term (u_{0j}) represents the deviation of the average listening time of individual participant *j* from the grand mean, u_{0i} represents the deviation of item *i* from the grand mean, and finally $e_{(ij)}$ indicates the residual score of individual *j* and item *i*. These residual scores are assumed to be normally distributed with an expected score of E(e_{ij}), E(u_{io}) and E(v_{oj}) = 0.0 and a variance of $S^2e_{(ij)}$, $S^2u_{(io)}$ and S^2v_{oj} respectively.

CHAPTER 3

OCP-PLACE IN SPEECH SEGMENTATION[1]

3.1 Introduction

Numerous languages show evidence for phonotactic restrictions that refer to the featural similarity of pairs of neighboring segments. Such restrictions typically hold within morphemes or words and can be broadly divided into two classes: those requiring segments to share some feature (i.e., harmony constraints), and those requiring segments to be featurally non-identical (i.e., disharmony constraints). Phonotactic restrictions of the former class typically require phonemic segments to be similar within words. Examples can be seen in vowel harmony (for an overview, see Archangeli & Pulleyblank, 2007) and consonant harmony (for an overview, see Rose & Walker, 2004). It has been suggested that harmony patterns are a result of historical sound change caused by listeners failing to compensate for coarticulation in perception (e.g., Ohala, 1994). These patterns could also be a consequence of natural human preferences to group similar sounds (e.g., Thorpe & Trehub, 1989); the law of similarity has been proposed to be one of the gestalt principles of perception (Wertheimer, 1923). The preference for grouping similar sounds may be helpful when processing speech, as, for example, in speech segmentation. Indeed, it has been demonstrated that infants and adults locate word boundaries between two disharmonic vowels, at least if their language is categorically restricted by vowel harmony (Turkish, but not French: Kabak, Maniwa, & Kazanina, 2010; Finnish: Suomi, McQueen, & Cutler, 1997; Turkish: Van Kampen, Parmaksiz, van de Vijver, & Höhle, 2008; Finnish, but neither French nor Dutch: Vroomen, Tuomainen, & de Gelder, 1998).

Languages display disharmony patterns as well, i.e., when certain phonemes within words *may not* be similar. Hence, the question arises whether the latter

---

[1] This study, co-authored by René Kager, has been submitted for publication. The study was funded by an NWO grant (277-70-001) awarded to René Kager. The authors contributed to this study as follows: hypothesis and design: Natalie Boll-Avetisyan and René Kager; testing, analysis and report: Natalie Boll-Avetisyan.

might also provide cues for speech processing. A well-attested example for phonotactic disharmony is OCP-PLACE, a constraint against consonants with the shared feature [place] within words. As opposed to vowel or consonant harmony, OCP-PLACE creates featural disagreement between phonemes within words. Hence, OCP-Place apparently defeats the presumed cognitive bias for grouping similar sounds. The effect of OCP-PLACE spreads over intervening vowels, constraining the co-occurrence of consonants that are phonetically non-adjacent (although consonants are represented as adjacent at a more abstract level on the "consonant tier"; McCarthy, 1986). Strong OCP-PLACE effects are found, for example, in Semitic languages such as Arabic (e.g., Frisch, Pierrehumbert, & Broe, 2004; Greenberg, 1950; McCarthy, 1985) and Hebrew (e.g., Berent & Shimron, 1997; McCarthy, 1985). Furthermore, many genetically and geographically unrelated languages display a less radical, yet systematic effect of OCP-PLACE (e.g., English: Berkley, 1994; Muna: Coetzee & Pater, 2008; Niger-Congo languages: Pozdniakov & Segerer, 2007; Dutch: Shatzman & Kager, 2007). In the latter type of languages, words containing pairs of consonants with shared place features do occur, such as /pVm/ in English *spam* or /mVb/ in *mob*. However, they occur less often than expected if non-adjacent consonants co-occurred at random.

Studies of the literature and our own studies of lexical databases of languages revealed that it is difficult to find a language that is not restricted by OCP-PLACE. This observation has influenced many researchers to consider OCP-PLACE to be an innate linguistic primitive (e.g., McCarthy, 1986; Pozdniakov & Segerer, 2007). Most phonologists assume that OCP is a restriction that holds at the level of underlying forms (for a surface-based implementation of OCP-Place in an Optimality Theory framework, see Gafos, 2003). Underlyingly, phonological features are assumed to be represented on different tiers, i.e., a vowel and a consonant tier. These tier representations are further subdivided into, for instance, a labial tier or a coronal tier. The constraint OCP-PLACE disfavors repetitions of features on a tier. If a form contains such two adjacent phonemes with a shared feature, OCP-PLACE makes sure that their representations are merged, that is, doubly linked to one feature specification rather than that there are two separate specifications (e.g., Goldsmith, 1979; Leben, 1980).

Over the past few years, a number of experimental studies have shown that listeners have unconscious knowledge of OCP-PLACE. For example, in studies with native listeners of Hebrew (Berent & Shimron, 1997), Arabic (Frisch & Zawaydeh, 2001), and English (Coetzee, 2008), nonwords that violate OCP-PLACE are judged to be less well-formed than nonwords that do not. Furthermore, OCP-PLACE affects performance in lexical decision tasks: nonwords that violate OCP-PLACE are rejected faster than nonwords composed of consonants that do not share [place] by native listeners of Hebrew (Berent,

Everett, & Shimron, 2001) and Dutch (Shatzman & Kager, 2007). Also, OCP-PLACE biases phoneme identification such that forms containing sequences that violate OCP-PLACE tend to be perceived as sequences of non-harmonic consonants by native listeners of English (Coetzee, 2005). These studies suggest a role for OCP-PLACE in lexical processing.

Autosegmental phonologists like Goldsmith or Leben do not attempt to account for a role of OCP in perception, being merely concerned with the representational merger. Boersma (2000) elaborates this idea in a speech perception context and proposes OCP to be a perceptual merger. He assumes that OCP affects perception in the process of translating a perceptual form into an underlying form. If a listener perceives a surface form that has two adjacent phonemes with shared place, then OCP-PLACE—disfavoring repetitions of features in underlying forms—has the perceptual effect of merging these phonemes into a higher-order unit in their underlying representation (Boersma, 2000:23). In some way, Boersma's view of OCP as a perceptual merger comes close to the assumption of the gestalt principle of similarity to group similar elements.

Frisch (2004) opens yet another functional perspective on OCP. He assumes that identity avoidance creates a functional advantage in speech processing. In perception, speech sounds are serially encoded. Once a phoneme is encoded, its co-encoded features are immediately inhibited to speed up the encoding of the following phoneme. Frisch proposes that the second of two consonants with shared features has a high chance of not being correctly encoded because of an inhibition of relevant features. This is useful for speeding up the lexical recognition process as words typically do not contain pairs of consonants with shared place. Frisch explains the typological recurrence of OCP through the functional bias: as words containing sequences of homorganic consonants are more difficult to process, they will—in the course of language change—be replaced by words that do not contain OCP-violating sequences.

In sum, functionalist theories assume that listeners do not fully and individually perceive both consonants in a pair of homorganic consonants. Frisch assumes that the perception of the second of two homorganic consonants is distorted, whereas Boersma assumes that homorganic consonants are merged into one percept.

In the current study, we want to look into a different role that OCP-PLACE may play in speech processing. Potentially, the requirement for consonants to disagree in [place] within a word may provide a useful cue for speech segmentation. For listeners whose native language is restricted by OCP-PLACE, if one encounters a sequence of two consonants with shared [place] in continuous speech, it is most likely that it occurs at a word boundary (e.g., /pʏm/ in *happy man*), as it is unlikely to reside within words. The possibility that the low probability of occurrence of pairs of consonants with shared place is used as a cue to segmentation is supported by previous segmentation studies that

have found that troughs in probability to signal word boundaries (e.g., Adriaans & Kager, 2010; Boll-Avetisyan et al., submitted).

Using OCP for segmentation requires that humans perceptually *separate* two consecutive consonants with shared [place]. If we take the theories of OCP in perception described above at face value, then it should be difficult for OCP-PLACE to serve as a cue for speech segmentation. Perceptual segregation can only work, if both homorganic consonants are faithfully perceived as individual elements. Frisch's assumption about an inhibition of the second of two homorganic consonants is problematic for the following reason: as for words, continuous speech also needs to be encoded serially. Hence, for example, in /pVm/ across the word boundary in *happy man*, the word onset /m/ would not be properly encoded due to inhibition of the labial feature after activating the preceding word offset /p/. So, Frisch's theory predicts that word recognition of *man* after *happy* would be slowed down. This, however, would be harmful to word recognition, as it would slow down the processing of a word onset; and word onsets are usually most important for lexical access. In any case, the prediction is not confirmed by the results of a word-spotting experiment carried out in our lab (Kager & Shatzman, under revision). Here, listeners were actually faster at spotting labial-initial words if they were preceded by a labial rather than when they were preceded by a coronal consonant.

The functionalist theories can be used to explain results of studies showing that listeners have unconscious knowledge of OCP-PLACE and use it for lexical processing. However, with respect to the potential role for OCP in speech segmentation, strong functionalist theories seem to be problematic: if OCP-PLACE is used as a cue for speech segmentation, then its use would have to be triggered by the lexicon. Frisch's theory of an inhibited encoding of the second of two similar consonants can only be correct if inhibition of the place feature stops at the end of a word or a morpheme. Yet it is not clear how the inhibitory process might stop at word boundaries when listening to continuous speech.

Boersma's assumption of a merger that creates a single percept of two adjacent homorganic consonants is problematic, if they are part of two separate words. Actually, Boersma (2000) discusses a case in which feature repetition occurs at a word boundary, the specific example being /t:/ in *it told*. For such a case, he proposes that merger can be overcome and repetitions of features might well be recognized as residing at a word boundary. However, in his formalization, this requires the presence of additional cues for lexical segregation. In interaction with these additional cues, the listener can recognize the two adjacent consonants as belonging to two separate words. Minimally, this leaves unexplained how this is to account for cases in which homorganic consonants are non-adjacent (as in *happy man*) (2000:6).

These studies suggest that the effect of OCP can only be overcome if lexical knowledge is drawn upon, such that a listener recognizes two homorganic

consonants (such as /pʏm/ in *happy man*) as belonging to two separate words. However, theoretically and empirically, this can be countered: previous studies on segmentation have emphasized that sub-lexical knowledge can be used as a cue to word boundaries independently of lexical knowledge; and phonotactic cues have proven to be a classical example of sub-lexical cues to segmentation (McQueen, 1998). Moreover, phonotactic knowledge of low probability phoneme co-occurrences in speech offers a strong segmentation cue (Boll-Avetisyan, Kager, et al., submitted). Such results make it plausible that OCP-PLACE, which has the effect that pairs of consonants with shared [place] are low in probability in speech, can potentially serve as a phonotactic cue for speech segmentation independent of the lexicon.

The current study aims to show that OCP-PLACE serves as a cue for speech segmentation independently of the presence of lexical cues to word boundaries. If this can be shown, then the functionalist theories on OCP as a universal bias on speech perception have to be put in question. One would have to rethink OCP-PLACE as a constraint influencing speech segmentation without involving lexical organization. That is, any segmentation effects attributed to OCP-PLACE reflect pure knowledge of an organization of phoneme co-occurrences in continuous speech.

With that said, the question arises whether listeners know that instantiations of two non-adjacent consonants with shared [place] indicate a word boundary and use this as a cue for segmenting continuous speech. Our hypothesis is that listeners whose native language is restricted by OCP-PLACE locate a word boundary between consonants that share the feature [place]. We hypothesize that the use of this constraint as a segmentation cue reflects knowledge of phoneme distributions in the native language. Consequently, we carried out four experiments testing Dutch participants on their use of OCP-PLACE in speech segmentation. Dutch, as will be described in Section 2, is gradiently restricted by OCP-PLACE. For our experiments, we used the artificial language learning (henceforth, ALL) paradigm. In artificial languages, information can be reduced to a minimum such that the impact of segmentation cues can be tested in isolation (i.e., lexical or phonological cues to word boundaries can be maximally excluded). In this task, participants are first familiarized with a continuous stream of an artificial language, a highly reduced miniature language. Subsequently, segmentation performance is assessed in a test phase by means of a two-alternatives-forced-choice task. Here, participants judged which of two items had occurred in the artificial language they were familiarized with. Performance in ALL tasks is affected by native language phonological knowledge (Finn & Hudson Kam, 2008; Onnis, Monaghan, Richmond, & Chater, 2005; Vroomen, et al., 1998). Hence, it is likely that when listeners use OCP-PLACE for speech segmentation in their native language, they will transfer it to segmenting an artificial language. A difference between our study and previous studies is that in our artificial languages no statistical segmentation

cues were present, such as differences in TPs between words and part-words. The only cue for segmentation was OCP-PLACE.

In the next section, non-adjacent consonant distributions in Dutch will be analyzed. In the subsequent sections, four experiments will be presented. The first experiment assesses whether Dutch listeners use OCP-PLACE as a segmentation cue in an artificial language (Experiment 1). Three follow-up experiments control for the possibility that effects of Experiment 1 are a) due to idiosyncrasies of the material, b) an effect of the participants' intuitions about the well-formedness of Dutch word forms, and c) not due to OCP-PLACE, but rather a general cognitive preference for similarity between segments in edges or avoidance of similarity in adjacency.

3.2. OCP-PLACE in Dutch

3.2.1. Method

In order to test our hypothesis, we used Dutch. The Dutch lexicon has been found to be gradiently restricted by OCP-PLACE (Kager & Shatzman, 2007). Furthermore, Dutch listeners possess unconscious knowledge about this restriction, as revealed in experiments with lexical decision tasks (Shatzman & Kager, 2007). We calculated the probabilities of C_1vC_2 sequences in which $C_1$and C_2 share [place]. All CvC sequences, independent of whether they occurred word-initially, -medially, or –finally, were included. We used two different data-bases: First, we extracted a lexicon of 8,305 Dutch monomorphemic stems from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). Second, we used CGN, a corpus of phonetically transcribed spoken Dutch (Goddijn & Binnenpoorte, 2003), in which we calculated the probabilities of C_1vC_2 sequences, including sequences both within words and across word boundaries. As a measure, we used the Observed/Expected (O/E) ratio (Pierrehumbert, 1993). This measure compares the observed counts of consonant pairs to the counts expected if consonants combined at random (Frisch, et al., 2004). O stands for the number of C_1vC_2 sequences in the corpus (or lexicon). It is divided by E, computed as the probability that C_1 occurs in the initial position of CvC, multiplied by the probability that C_2 occurs in the final position of CvC, which is multiplied by the total number of CvC sequence tokens:

$O/E = N(C_1vC_2) / p(C_1) * p(C_2) * N(CvC)$

$N(C_1vC_2)$: the number of C_1vC_2 sequences; N(CvC): the number of CvC sequences

$p(C_1)$, $p(C_2)$: the probability of C_1 and C_2 calculated as proportions of consonants

O/E > 1: sequence is over-represented; O/E < 1: sequence is under-represented

3.2.2. Results and discussion

The results of our calculations show that in Dutch, CvC sequences in which both Cs share the feature [place] are under-represented both in the lexicon as well as in continuous speech. This holds in particular for sequences of labials (the phonemes /p, b, f, v, w, m/; hereafter, P) and dorsals (the phonemes /k, g, x, ŋ/; hereafter, K). The distribution of coronals (the phonemes /t, d, s, z, ʃ, ʒ, r, l, n/; hereafter, T) is somewhat less restricted (see Table 1).

Data source	PvP	TvT	KvK
CELEX types	0.45	0.77	0.58
CGN continuous	0.58	1.24	0.69

Table 1: O/E values for consonant pairs sharing place of articulation (P = labial, K = dorsal, T = coronal) in CvC sequences. Calculations are based on counts in CELEX lemma types (CELEX types) and an unsegmented corpus of spoken Dutch (CGN continuous).

The distributions of Cs with shared [place] of articulation in CvC justify the assumption that OCP-PLACE holds gradiently in the lexicon. More specifically, Dutch is restricted by the gradient phonotactic constraints OCP-LABIAL (CELEX: O/E = 0.45, CGN: O/E = 0.58) and OCP-DORSAL (CELEX: O/E = 0.58, CGN: O/E = 0.69), and possibly also OCP-CORONAL, if phonotactics were learned from the lexicon (CELEX: O/E = 0.77), but not when phonotactics are acquired from continuous speech (CGN: O/E = 1.24). Since OCP-LABIAL shows the most significant under-representation, we tested our assumptions by means of this constraint.

For Experiment 1, we created an artificial language in which coronal-initial Cv syllables (hereafter, Tv) were always followed by two labial-initial Cv syllables (hereafter, Pv). The syllables were concatenated into a speech stream without pauses (…$Pv_1Pv_2TvPv_1Pv_2TvPv_1Pv_2Tv$…). We predicted that if Dutch participants have knowledge that PvP sequences are under-represented in their native language, and are able to apply this knowledge when segmenting speech, they should insert a boundary between Pv_1Pv_2. Hence, in a test phase, PvTvPv words should be preferred over PvPvTv and TvPvPv words. Furthermore, there

should not be a preference of PVPVTV over TVPVPV or vice versa, as both of them are segmentations that violate OCP-LABIAL and, hence, should not be segmented from the stream.

3.3. Experiment 1

3.3.1. Method

Participants

Fifty-four students, (12 male, 42 female, M_{age} = 23,67 years, age range: 18-68 years) from Utrecht University participated in the experiment. All participants reported normal hearing and Dutch to be their only native language. They were compensated for their efforts.

Materials

Familiarization stimuli

The artificial language consists of nine syllables that are assigned to one of three slots (see Table 2). Three PV_1 syllables were always followed by one of the three PV_2 syllables, which were followed by one of the three TV syllables. Syllables were concatenated into a speech stream with no pauses (e.g., …*pamatumomatubibetu*…). TPs between both adjacent and non-adjacent syllables were always 0.33, so that the language itself did not contain a distributional cue for segmentation.

The syllable inventory (PV_1 = {/po, be, ma/}, TV = {/tu, do, ne/} and PV_2 = {/pa, bi, mo/}) was selected on the basis of controls for lexical, low-level phonotactics, and other phonological factors to assure that segmentation would be only cued by OCP-LABIAL, independent of the intervening vowels. To do so, we controlled for potential cues for segmentation that may resemble cues from OCP-LABIAL. The consonants /p/, /b/, and /m/ were selected to represent P, and /t/, /d/ and /n/ to represent T. By selecting three consonants of different manner (voiceless and voiced plosive, and nasal) for each consonant class (P and T), we avoided any confounding effects of manner on segmentation. We did not include any fricatives (e.g., /f/, /v/, /z/, /s/), as voiced and voiceless realizations of fricatives are highly variant (and hence, perceptually confusable) in Dutch.

Natural segment durations (voiced plosives being shorter than unvoiced plosives etc.; Waals, 1999) were maintained, but the syllables were distributed such that

their duration was controlled on average per position. Sequences of perceptually similar syllables were avoided (e.g., /mu-nu/, /bi-pi/). Furthermore, sequences that are real words (e.g., /bebi/ 'baby') or affixes (e.g., the prefix /na-/) that might affect segmentation were avoided. Moreover, we controlled for the probability of identical vowel sequences between parsings.

Experiment	**A**	**B**	**A**
1	**Pv_1**	**Tv**	**Pv_2**
	/po/	/tu/	/pa/
	/be/	/do/	/bi/
	/ma/	/ne/	/mo/
2	**Pv_1**	**Tv**	**Pv_2**
	/pe/	/tu/	/po/
	/bo/	/de/	/be/
	/mi/	/na/	/ma/
4	**Tv_1**	**Pv**	**Tv_2**
	/ta/	/po/	/tu/
	/di/	/be/	/do/
	/no/	/ma/	/ne/

Table 2: Syllable inventory of the artificial languages used in Experiments 1, 2 and 4.

The vowels were selected such that the language was controlled for cues from positional syllable frequencies which might give cues to word boundaries (e.g., a certain syllable might be infrequent in word-initial but frequent in word-medial position). Hence, based on word types in the CELEX lexical database (Baayen, et al., 1995), we selected syllables for the sets (Tv, Pv_1, Pv_2) such that their frequency would be evenly distributed in utterance-initial, -medial, and -final position (see Appendix A).[2] If it is parsed into tri-syllabic units, each of the speech streams can be parsed into 27 PvTvPv, 27 PvPvTv, and 27 TvPvPv

[2] We thank Frans Adriaans for supplying us with the software PhonotacTools for calculating lexical statistics.

words. The artificial speech stream was pseudo-randomly organized in such a way that no tri-syllabic string occurred twice in succession.

The language was synthesized with a male voice "nl2" based on Dutch biphones, provided by FLUENCY using the MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996). The synthesizer produced the stream with a monotone pitch (F0 = 100 Hz) with an average syllable duration of 232 ms.[3] In order to avoid that the endpoints (utterance boundaries) gave a cue for segmentation, intensity faded in for the first five seconds and out for the last five seconds.

Test stimuli

We synthesized three types of stimuli: PvTvPv, PvPvTv, and TvPvPv items for three different test phases. Test-phase 1 contrasted PvTvPv (e.g., *ponebi*) and PvPvTv (e.g., *pamado*) items, Test-phase 2 PvTvPv (e.g., *benemo*) and TvPvPv (e.g., *domoma*) items, and Test-phase 3 PvPvTv (e.g., *mobedo*) and TvPvPv (e.g., *nebibe*) items.

The test items were balanced for phonological and lexical factors that potentially affect decisions: First, the two types of items used in the test phase within one Test-phase were balanced for the number of nasals, voiced and unvoiced plosives with their counterparts. That is, the PvTvPv item *ponebi* would receive the PvPvTv item *pamado* as a counterpart, as both start with an unvoiced plosive, followed by a nasal and end with a syllable that starts with a voiced plosive (see Appendix B). Second, we controlled for the lexical factors cohort density and lexical neighborhood density. Cohort density was defined as the sum of the logged frequencies of the set of words in CELEX starting with the same three phonemes, whereas lexical neighborhood density was defined as the sum of the logged frequencies of the set of words in CELEX that differed from the test item by a single segmental change, e.g., by deletion, addition, permutation, alteration (Luce, 1986).

The best controls for cohort density were guaranteed using twelve test items per comparison. So, Test-phase 1, for example, used twelve PvTvPv and twelve PvPvTv items. The cohort densities for the test items can be viewed in Appendix C. Lexical neighborhood density zero for all test items.

The test phase consisted of forty-eight trials in total, of which thirty-six were real test pairs (e.g., PvTvPv-PvPvTv), and twelve filler pairs with two items from same class (e.g., PvTvPv-PvTvPv). Each item occurred four times in a different test pair, three times as a target, and once in a filler pair. The test pairs

[3] We thank Arthur Dirksen from *Fluency* for a script to generate the stream.

were combined at random with the prerequisites that two items only made a pair once, and exactly half of the test pairs started with one type of stimulus.

Apparatus and procedure

The procedure was borrowed from Peña et al. (2002). Each participant was tested individually in a sound-attenuated booth. Stimuli were presented over headphones. To acquaint participants with the two-alternatives-forced-choice task of the test phase, they first entered a pre-familiarization phase, in which they were asked to find the syllable /so/ in a pair of CV syllables. Subsequently, participants entered the training phase. They were informed that they would listen to an artificial language for the next ten minutes and were instructed to listen carefully, because they would later be tested on their knowledge about the words of this language. After the training phase, participants were tested in a two-alternative forced choice task, in which they had to indicate via a mouse-click on a button on the screen whether the first or the second of two items was more likely to be a word from the language they just heard. Test pairs were presented with an inter-stimulus interval of 500 ms. Participants were randomly distributed to one of the three test phases (Test-phase 1: N = 18; Test-phase 2: N = 18; Test-phase 3: N = 18).

3.3.2. Results and discussion

Results

The filler pairs were excluded from the analysis. Three outliers, i.e., one participant per each Test-phase, were excluded due to deviating from the mean by 2 SD. Participants preferred PVTVPV over PVPVTV items in Test-phase 1 (56% preference for PVTVPV, $p < 0.001$ binomial test, two-sided) and PVTVPV over TVPVPV items in Test-phase 2 (56% preference for PVTVPV, $p < 0.001$ binomial test, two-sided). However, participants showed no preference in Test-phase 3 (54% preference for PVPVTV, p = n.s. binomial test, two-sided).

Discussion

Our prediction was that Dutch listeners—when listening to an artificial language in which two PVs are followed by a TV (i.e., …PVPVTVPVTVPVPVTV…)—would transfer knowledge of an under-representation of non-adjacent labials in their native language, and hence would locate word boundaries within PVP sequences. As predicted, PVTVPV items, which satisfy OCP-LABIAL, were preferred over PVPVTV and TVPVPV items. The stimuli were controlled for low-level phonotactics and lexical statistics. Thus, segmentation preferences cannot be a result of distributional knowledge other than OCP-LABIAL. The null-result in Test-phase 3 is reasonable in light of the assumption that OCP-LABIAL, as a

speech segmentation cue, has guided the perception of the artificial language. If participants have segmented the stream into …PVTVPV-PVTVPV-PVTVPV…, then it should be difficult for them to display a preference during a test phase that only offered them to choose from PVPVTV and TVPVPV items, which they have not segmented.

All in all, the result is in line with our hypothesis that OCP-LABIAL is used as a cue to detect word boundaries in the speech signal. However, at this point, there are several alternative explanations that need to be ruled out, which will be addressed in the following three experiments. First, it remains possible that the results found in Experiment 1 are due to idiosyncrasies of the stimulus material, in particular the order of specific syllables in the artificial language. Classically, ALL studies on statistical learning would include a control experiment in which statistical cues are re-distributed in such a way that part-words become words, and words become part-words (e.g., Peña et al., 2002). The artificial language in Experiment 1 does not give any statistical cues for boundaries. Any cues for boundaries could only be based on distributional properties of the native language, transferred to the segmentation of the artificial language. By controlling for low-level phonotactics and lexical statistics, we already severely reduced the possibility that segmentation preferences were caused by unknown properties of the syllables and their order. Still, to ultimately exclude this confounding factor we ran Experiment 2. In this experiment, we changed syllables and their arrangement to see whether we would find the same effects as in Experiment 1.

Second, as OCP-LABIAL affects consonant distributions in the Dutch lexicon, it might be argued that the preferences in Experiment 1 were not due to segmentation of the artificial language during the familiarization phase, but were merely a result of well-formedness intuitions that were applied during the test phases. To further assess whether knowledge of OCP-LABIAL was used for segmentation or whether it merely reflected knowledge about the well-formedness of the test items in Experiment 1, we ran Experiment 3. We used the stimuli of the test phases in Experiment 1 in a well-formedness judgment task without a familiarization phase. Predictions for Experiment 3 will be explained below.

Third, a question that is still open is whether segmentation performance in Experiment 1 was driven by the participants' distributional knowledge about an under-representation of PVP in Dutch. Another possibility is that the preference to segregate consonants with shared [place] is a perceptual universal. Beyond this, listeners might display a more general cognitive preference for similarity at edges (ABA > AAB, ABB, regardless of values of A, B).[4] For example, Onnis

[4] Thanks to Jacques Mehler for pointing out the possibility that listeners prefer similarity in edges.

and colleagues (2005) showed that English speakers could only learn an X_1-C-X_2 pattern if X_1 and X_2 shared manner of articulation, i.e., both start with plosives or both with continuants. To rule out these alternative accounts, we conducted Experiment 4.

For Experiment 2, which will be described in the following section, we predict the same results as in Experiment 1: that is, a segmentation preference for PvTvPv over PvPvTv and TvPvPv.

3.4. Experiment 2

3.4.1. Method

Participants

Thirty-nine students, (3 male, 36 female, M_{age} = 23,03 years, age range: 18-50 years) from the Utrecht University participated in the experiment. They reported normal hearing and Dutch to be their only native language. They were compensated for their efforts.

Materials

Familiarization stimuli

The artificial language used here was a minimal variant of the language in Experiment 1. The crucial difference between the language in Experiment 1 and the language used here is that the syllables in word-initial position Pv_1 were assigned to word-final position Pv_2 (see Table 2). A new set of syllables was selected for T and Pv_2 (Pv_1 = {/pe, bo, mi/}, Tv = {/tu, do, na/}, and Pv_2 = {/po, be, ma/}) in order to meet the same controls that we applied to the language used in Experiment 1 (see Appendix A). The predictions for this experiment were evidently the same as in Experiment 1, that is, participants should prefer segmentations driven by OCP-LABIAL.

Test stimuli

In this experiment, we had two instead of three different test phases: Test-phase 1 (PvTvPv vs. PvPvTv) and 2 (PvTvPv vs. TvPvPv). We dropped PvPvTv vs. TvPvPv, as this comparison is not essential for testing our predictions, and results for this comparison in Experiment 1 were non-significant. For each test phase, we selected twelve test items of each type that met the same criteria as in Experiment 1 (see Appendix C and D). Again, each test trial was repeated thrice

in a test trial pair and occurred once in a filler pair with a test item of the same type.

Apparatus and procedure

The procedure was the same as in Experiment 1. The thirty-nine participants were randomly distributed to either Test-phase 1 (N=20) or Test-phase 2 (N=19).

3.4.2. Results and discussion

Two outliers, of which one was tested on Test-phase 1 and one on Test-phase 2, were excluded, because they deviated from the mean by more than 2 SD. As in Experiment 1, participants preferred PvTvPv over PvPvTv items in Test-phase 1 (56% preference for PvTvPv, $p < 0.005$, binomial test, two-sided) and PvTvPv over TvPvPv in Test-phase 2 (54% preference for PvTvPv, $p < 0.005$, binomial test, two-sided). In Experiment 2, it was tested whether the result in Experiment 1 may have been caused by idiosyncratic properties of the specific syllables and their order. The result suggests that this alternative explanation can be disregarded.

Next, in Experiment 3, we used the stimuli of the test phases in Experiment 1 in a well-formedness judgment task without a familiarization phase. Previous studies have shown that listeners use knowledge of OCP-PLACE for phonotactic well-formedness judgments (e.g., Hebrew: Berent & Shimron, 1997; Arabic: Frisch & Zawaydeh, 2001, English: Coetzee, 2008). For that reason, we need to exclude the possibility that the effects found in Experiment 1 and 2 were due to preferences that only came into play in the test phase, but did not affect segmentation during the familiarization with the artificial language.

In the discussion of Experiment 1, we argued that the null-result in Test-phase 3 is reasonable under the assumption that OCP-LABIAL has been used to segment the artificial language into …PvTvPv-PvTvPv-PvTvPv…: if participants have segmented sequences of PvTvPv, then it should be difficult for them to display a preference during a test phase that only offered them a forced choice between non-segmented PvPvTv and TvPvPv items, both OCP violators. If the forced choice task calls upon verbatim recall of items that were actually perceived during listening to the artificial language, then participants should have problems in the case of a PvPvTv versus TvPvPv comparison. In a well-formedness judgment task, however, there may be additional phonotactic knowledge that comes into play. In Dutch, Ps are likely to occur at word beginnings (O/E = 2.87 in CELEX lemma types, calculated as $E_{\#C1} = p(\#C_1) * N_{\#X}$), giving rise to the possibility that well-formedness judgments are affected

by this distributional knowledge. Kager and Shatzman (2007) assume a constraint ALIGN-LABIAL requiring labials to align with the left edge of a word, and found Dutch listeners to be influenced by knowledge of ALIGN-LABIAL in a lexical decision task.

For this reason, we expect the following results for Experiment 3: if the judgments in the test phases of Experiment 1 and 2 were independent of familiarization with the artificial language, then participants in Experiment 3 should display the same preferences as in Experiment 1. However, if the locus of the effect found in Experiment 1 was in segmentation, then the results in the following experiment should be as follows: in Test-phase 1 and 2, listeners should have a well-formedness preference for PVTVPV over PVPVTV and TVPVPV nonwords due to phonotactic knowledge of OCP-PLACE. In Test-phase 3 using PVPVTV and TVPVPV nonwords, we expect Dutch listeners to judge PVPVTV to be more well-formed than TVPVPV due to phonotactic knowledge of an over-representation of P in word-initial position.

3.5. Experiment 3

3.5.1. Method

Participants

Thirty-nine native speakers of Dutch (22 women and 17 men, M_{age}: 27,38 years, age range: 18-52 years) with normal hearing, all of whom had not been tested in Experiment 1 or 2, participated in the experiment. One additional participant who had made more than two mistakes in the pre-test was excluded.

Materials, apparatus and procedure

The stimuli were the same as used in the test phase in Experiment 1. Thirteen participants were tested in a sound-attenuated booth. Twenty-six participants were tested in a quiet room with a laptop. The set-up of the experiment was identical to that of Experiment 1 with the only difference being that there was no familiarization phase. The instructions were to indicate via a mouse-click on a button on the screen whether the first or the second of two items would be a better word of Dutch. Thirteen participants were tested on Test-phase 1, fourteen on Test-phase 2, and twelve on Test-phase 3.

3.5.2. Results and discussion

Fillers were excluded from the analysis. Results showed that participants had a preference for PvTvPv over PvPvTv (58% preference for PvTvPv, $p < 0.001$, binomial test, two-sided) and over TvPvPv items (55% preference for PvTvPv, $p < 0.05$, binomial test, two-sided). Moreover, in Test-phase 3, participants displayed a significant preference for PvPvTv over TvPvPv (58% preference for PvPvTv, $p < 0.001$, binomial test, two-sided).

As expected, Dutch listeners did not only use knowledge of OCP-PLACE when judging about the well-formedness of nonwords in Test-phase 1 and 2. Moreover, they used knowledge of, the initial predominance of labials, i.e., a constraint ALIGN-LABIAL, to which they preferred nonwords starting in Ps over nonwords starting in Ts in Test-phase 3.

The results of Test-phase 3 indicate that the dispreferences for OCP-violating nonwords in Experiment 1 and 2 cannot be simply accounted for by well-formedness intuitions that only emerged during the test phase. Rather, the preferences found in the test phase in Experiment 1 must be a result of exposure to and segmentation of the artificial language, driven by OCP-LABIAL. The results support our line of reasoning that participants have segmented the artificial language into …PvTvPv-PvTvPv-PvTvPv…, for which it was impossible for them to display a preference for PvPvTv or TvPvPv in Test-phase 3.

The last experiment addresses the question whether the result that participants segmented PvTvPv in Experiment 1 and 2 is caused by a general perceptual preference for similarity of consonants in word edges. We created an artificial language similar to the language in Experiment 1, the only difference being that one P was replaced by T (…TvTvPvTvTvPvTvTvP…). If the result in Experiment 1 is due to a general ABA preference, then participants should segment TvPvTv words from the speech stream in Experiment 4. Our hypothesis is that listeners do not insert boundaries between similar consonants unless this is driven by knowledge of the native language's distributions in the lexicon. In Dutch, distributional properties give less reason to use OCP-CORONAL (TvT: O/E = 0.77) for segmentation than OCP-LABIAL (PvP: O/E = 0.45). So, Dutch listeners should not insert boundaries between two Ts.

A question to be posed is whether Dutch listeners will have a preference for segmenting the stream. There are two reasons for which we expect them to have a preference for PvTvTv: First, in this task, Dutch listeners should put the only Pv-syllable of the stream in initial position, reflecting positional phonotactic knowledge of the over-representation of P in word-initial position, i.e., ALIGN-LABIAL. Second, with respect to the preferences of distributing the Tv-syllables in the stream, the lexical statistics of Dutch by themselves do not give a cue.

However, as already mentioned in the introduction, humans in general like to group similar sounds (e.g., Thorpe & Trehub, 1989; Wertheimer, 1923). As the distribution of TVTV is neither specified by statistical cues in the stream nor by distributional knowledge of the native language's lexicon, we predict that participants prefer grouping TVTV.

3.6. Experiment 4

3.6.1. Method

Participants

Participants were forty-three students (6 male, 37 female, M_{age} = 20.05 years, age range: 18-31 years) without hearing difficulties and Dutch as a native language, none of whom participated in any other experiment of this study.

Materials

The artificial language was similar to that of Experiment 1, with the only difference being that P_1={pa, bi, mo} was replaced by T (TV_2={ta, di, no}). Syllables and their order met the same controls as the languages created for Experiment 1 and 2 (see Table 2). The stream was controlled for positional syllable frequency (see Appendix A). As in the previous experiments, each 12 TVPVTV, TVTVPV and PVTVTV test items were selected that matched in cohort density and manner of articulation (see Appendix C and E). No test items had any lexical neighbors. As in Experiment 1, forty-eight test pairs were constructed with each test item being repeated four times: three times as a target item and once as a filler item.

Apparatus and procedure

Apparatus and procedure were identical to Experiment 1. After familiarization, participants were randomly distributed to one out of three different test phases. Fifteen participants were tested in Test-phase 1 on pairs of TVPVTV-TVTVPV, fourteen participants were tested in Test-phase 2 on pairs of TVPVTV-PVTVTV, and fourteen participants were tested in Test-phase 3 on pairs of PVTVTV-TVTVPV.

3.6.2. Results and discussion

Results

Fillers were excluded from the analysis. In Test-phase 1, participants had a preference for neither TvPvTv nor TvTvPv items (52% preference for TvTvPv, $p > 0.05$, binomial test, two-sided). In Test-phase 2 and 3, participants significantly preferred PvTvTv over TvPvTv items (64%, $p < 0.001$ binomial test, two-sided) and over TvTvPv items (57%, $p < 0.01$ binomial test, two-sided).

Discussion

No preference for TvPvTv items was found. If TvPvTv had been the preferred segmentation, the result in Experiment might have been attributed to a general cognitive preference for identical elements at edges. This explanation can now be rejected, strengthening the interpretation that the effects found in Experiment 1 are due to an influence of language-specific knowledge of OCP-LABIAL on segmentation. The current experiment showed an unexpected effect, as listeners preferred a PvTvTv segmentation that assigns P to an initial position. Evidently, this is not an OCP-LABIAL effect. Two possible explanations arise, one based on language-specific distributional factors, the other on a general bias for grouping similar sounds together. First, the preference for PvTvTv may be due to participants' use of the constraint ALIGN-LABIAL, reflecting knowledge of the over-representation of Ps at word beginnings in the Dutch lexicon. Additionally, as the distribution of pairs of Ts is not severely restricted in Dutch, a general bias for grouping similar sounds together may be responsible. In any case, the result of Experiment 3 supports the conclusion that the results in Experiment 1 are due to knowledge of OCP-LABIAL.

3.7. General discussion

The hypothesis stated in this paper was that listeners whose native language displays an under-representation of non-adjacent consonants with shared [place] will use their knowledge of this phonotactic restriction in speech segmentation. The hypothesis was put to the test by examining the role of OCP-PLACE in speech segmentation in Dutch listeners. In line with our prediction, Dutch listeners preferred inserting word boundaries between two consecutive labials in an artificial language, suggesting their use of OCP-LABIAL for speech segmentation. Control experiments confirmed that the use of OCP-LABIAL

affected segmentation. It is neither the result of phonotactic well-formedness intuitions that determined decisions in the test phase (Experiment 3), nor an artifact of idiosyncrasies of the artificial language (Experiment 2).

The question was raised whether segmentation preferences in Experiment 1 could have been due to a general preference for segmenting similar consonants or to a cognitive preference for aligning similar units with edges. However, in Experiment 4, participants did not segregate two consecutive coronals. They would have done so if the use of OCP-PLACE were a preference independent of distributions in the lexicon. The under-representation of TvT sequences in the Dutch lexicon is far less severe than the under-representation of PvP, which renders OCP-CORONAL largely ineffective as a cue for speech segmentation.

The results of the current study can be interpreted as evidence against current functionalist theories of OCP. Functionalist theories assume that if homorganic consonants occur in adjacency, then it is not possible to fully perceive both of them. Frisch (2004) assumes that features that are activated when processing a phoneme are consecutively inhibited, leading to difficulties in encoding the second of two homorganic consonants. Boersma (2000) assumes that two homorganic consonants are merged into a single percept. The problem with both of these functionalist accounts is that they can only explain effects of OCP on word processing, but fail to account for effects of OCP on speech segmentation.

In order to use OCP as a speech segmentation cue, it is necessary that both consonants in a homorganic pair are perceived accurately. Moreover, it is crucial that they are identified as belonging to two separate words. Hence, they need to be segregated in perception. In both Frisch's and Boersma's functionalist accounts, OCP-violating sequences can only be identified as belonging to separate words if segmentation processes draw on lexical information. In all segmentation experiments of the current study, cues for segmentation from lexical distributions were rigidly controlled for. This shows that OCP-PLACE is useful for finding word boundaries by itself without top-down feedback from the lexicon.

The results of Experiment 4 further suggest that OCP effects on segmentation cannot be attributed to a general cognitive bias without involving language-specific knowledge. The fact that OCP-LABIAL is used as a cue for segmentation while OCP-CORONAL is not makes it most likely that the use of the cue is language-specific and acquired from the input language. It is unlikely that there is a functional bias exclusively targeting co-occurrences of labials but not co-occurrences of coronals.

The status of OCP as a language-inspecific functional bias is furthermore unlikely in light of the results of another study carried out in our lab. Here, monolingual Mandarin Chinese listeners, whose native language does not display an under-representation of labial-labial, do not use OCP-LABIAL as a cue for segmentation, whereas advanced Mandarin Chinese learners of Dutch do so

(Boll-Avetisyan, Chen, & Kager, submitted). This suggests that listeners' ability to use OCP-PLACE as a segmentation cue depends on their prior exposure to sound distributions. Hence, we can conclude that OCP-PLACE is a language-specific cue for speech segmentation. Its effect coincides with phonotactic distributions in the native language's lexicon, yet segmentation by OCP is not dependent on top-down cues from lexical knowledge.

The finding of the current study that native listeners possess knowledge of non-adjacent consonant probabilities may have important implications for language acquisition. The artificial languages we used were controlled for low-level phonotactic and lexical factors. Therefore, it is highly unlikely that knowledge of the probability of specific bi- or triphones (e.g., /pa:/, /a:m/, or /pa:m/) or the probability of similar real words of Dutch (e.g., /pa:r/ *paar* 'few') have influenced segmentation preferences. Instead, the results suggest that listeners have a representation of non-adjacent dependencies that abstract over intervening vowels (e.g., /pVm/, /bVp/, etc.).

The fact that native listeners have knowledge of non-adjacent dependencies involving dissimilarity is interesting: It has been suggested that non-adjacent dependencies between phonological units are difficult to acquire due to their weak statistical support in the input (e.g., Pierrehumbert, 2003). Previous ALL studies have suggested that non-adjacent dependencies can only be learned if they affect units that are similar to some degree. Non-adjacent dependencies between phonemes (i.e., between consonants or vowels) have been found to be learnable (e.g., Newport & Aslin, 2004), while non-adjacent dependencies between syllables have been found to be learned only if sequences begin with consonants that share manner of articulation features (e.g., Newport & Aslin, 2004; Onnis, et al., 2005; Perruchet & Pacton, 2006).

However, in the case of OCP-PLACE, the phonotactic requirement seems to be rather the opposite, i.e. that words do *not* contain sequences of similar phonemes. If similarity is needed for non-adjacent dependency learning, then how can phonotactic knowledge of OCP-PLACE be acquired?

A clue to this may come from the following: listeners do not only have knowledge about which phoneme sequences are likely to co-occur (as, for example, in the case of vowel harmony). Results from studies with infants at our lab indicate that humans also need to build up representations of improbable phoneme sequences. The study tested the potential use of constraints on a non-adjacent dependency between identical consonants in infant speech segmentation. It was only found that infants can use knowledge of under-representations of pairs of identical consonants for speech segmentation, while no evidence that knowledge of over-representations of identical consonants are a useful cue for speech segmentation (Boll-Avetisyan, Kager, et al., submitted).

It is possible that not only when acquiring knowledge of an over-representation, but also when acquiring knowledge about an under-representation of phoneme co-occurrences, learners rely on an innate or universal bias for similarity grouping (e.g., Thorpe & Trehub, 1989; Wertheimer, 1923). That is, under-representations that involve featurally identical segment pairs will naturally stand out, and hence would be easier to learn than under-representations involving featurally non-identical pairs. It will be interesting for future studies to further investigate the role of a bias for attending to similarity in the acquisition of OCP-PLACE. For example, it will be interesting to test whether non-adjacent dependencies involving pairs of non-similar consonants can be used for segmentation as well, and also whether such dependencies are indeed less learnable than those involving pairs of similar consonants. In sum, the current study provides evidence against a functionalist theory of OCP that assumes natural inhibitions against perceiving consecutive homorganic consonants. However, the study leaves open the possibility that an innate cognitive bias for grouping similar elements may nevertheless play a role in the acquisition of OCP.

3.8. Conclusion

We conclude that OCP-PLACE gives cues to word boundaries in speech segmentation. The effect of this constraint depends on an under-representation of segment sequences in the lexicon of the native language and is, hence, language-specific. The use of OCP-PLACE as a speech segmentation cue is independent of the lexicon, and cannot be explained by a functional bias on perception.

3.9. References

Adriaans, F., & Kager, R. (2010). Adding Generalization to Statistical Learning: The Induction of Phonotactics from Continuous Speech. *Journal of Memory and Language, 62*, 311-331.

Archangeli, D., & Pulleyblank, D. (2007). Harmony. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 353-378). Cambridge, England: Cambridge University Press.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philedelphia: Linguistic Data Consortium, University of Pennsylvania.

Berent, I., Everett, D. L., & Shimron, J. (2001). Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology, 42*, 1-60.

Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition, 64*, 39-72.

Berkley, D. (1994). The OCP and gradient data. *Studies in the Linguistic Sciences, 24*, 59-72.

Boersma, P. (2000). The OCP in the perception grammar. *Rutgers Optimality Archive, 435*, 1-52.

Boll-Avetisyan, N. A. T., Chen, A., & Kager, R. W. J. (submitted). OCP-Place for speech processing: innate or acquired knowledge? The case of Mandarin Chinese.

Boll-Avetisyan, N. A. T., Kager, R. W. J., de Bree, E., Kerkhoff, A., van den Bergh, H., & den Boer, S. (submitted). Does the lexicon bootstrap phonotactics, or vice versa? .

Coetzee, A. (2005). The OCP in the perception of English. In S. Frota, M. Vigario & F. M. J. (Eds.), *Prosodies* (pp. 223-245). New York: Mouton de Gruyter.

Coetzee, A. (2008). Grammar is both categorical and gradient. In S. Parker (Ed.), *Phonological Argumentation*. London: Equinox.

Coetzee, A., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory, 26*, 289-337.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). *The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes.* Paper presented at the Proceedings of The Fourth International Conference on Spoken Language Processing (ICSLP'96) 3.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*, 477-499.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. M. Kirchner & D. Steriade (Eds.), *Phonetically based Phonology* (pp. 346-371). Cambridge: Cambridge University Press.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory 22*, 179-228.

Frisch, S. A., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language Acquisition, 77*, 91-106.

Gafos, A. I. (2003). Greenberg's Asymmetry in Arabic: A Consequence of Stems in Paradigms. *Language 79*(2), 317-355.

Goddijn, S., & Binnenpoorte, D. (2003). *Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus.* Paper presented at the 15th ICPhS, Barcelona.

Goldsmith, J. (1979). *Autosegmental Phonology*. New York: Garland.

Greenberg, J. (1950). The patterning of root morphemes in Semitic. *Word, 5*, 162-181.

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Journal of Laboratory Phonology, 1*, 207-224.

Kager, R., & Shatzman, K. (2007). Phonological constraints in speech processing. In B. Los & M. v. Koppen (Eds.), *Linguistics in the Netherlands 2007* (pp. 100-111). Amsterdam: John Benjamins.

Kager, R., & Shatzman, K. (under revision). *Similarity Avoidance in Speech Processing*.Unpublished manuscript.

Leben, W. (1980). *Suprasegmental phonology*. New York: Garland Press.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Doctoral dissertation, Indian University, Bloomington, IN.

McCarthy, J. J. (1985). *Formal Problems in Semitic Phonology and Morphology*. New York: Garland.

McCarthy, J. J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry, 17*, 207-263.

McQueen, J. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language, 39*(1), 21-46.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology 48*, 127-162.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*, 225-237.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233-238.

Pierrehumbert, J. B. (1993). *Dissimilarity in the Arabic verbal roots.* Paper presented at the Proceedings of the North East Linguistics Society.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge MA: MIT press.

Pozdniakov, K., & Segerer, G. (2007). Similar Place Avoidance: A statistical universal. *Linguistic Typology, 11*, 307-348.

Rose, S., & Walker, R. (2004). A Typology of Consonant Agreement as Correspondence. *Language, 80*, 475-531.

Shatzman, K., & Kager, R. (2007). A role for phonotactic constraints in speech perception. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1409-1412). Dudweiler: Pirrot.

Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.

Thorpe, L. A., & Trehub, S. E. (1989). Duration Illusion and Auditory Grouping in Infancy. *Developmental Psychology, 25*(1), 122-127.

Van Kampen, A., Parmaksiz, G., van de Vijver, R., & Höhle, B. (2008). Metrical and statistical cues for word segmentation: Vowel harmony and word stress as cues to word boundaries by 6- and 9-month old Turkish learners. In A. Gavarró & M. J. Freitas (Eds.), *Language Acquisition and Development* (pp. 313-324). Newcastle: Cambridge Scholars Publishing.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133-149.

Waals, J. (1999). *An experimental view of the Dutch Syllable. Doctoral dissertation* (Vol. 18). Utrecht: LOT.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psycologische Forschung, 4*, 301-350.

Appendix A

Syllable Frequency in Experiment 1, 2 & 4 in % per within word position. Counts are based on word types in the Dutch CELEX lexical database.

Syllable Frequencies	Syllable	% ini	% med	% fin
Experiment 1	**PV_1**	0.50	0.43	0.07
	TV	0.60	0.37	0.03
	PV_2	0.45	0.52	0.03
Experiment 2	**PV_1**	0.48	0.48	0.04
	TV	0.61	0.37	0.02
	PV_2	0.50	0.43	0.07
Experiment 4	**TV_1**	0,48	0.48	0.04
	PV	0.61	0.37	0.02
	TV_2	0.50	0.43	0.07

Appendix B

Test items in Experiment 1.

F = voiceless bilabial, V = voiced bilabial, N = nasal bilabial.

Test-phase 1			
PPT	**Manner**	**PTP**	**Manner**
papotu	FvFvFv	potupa	FvFvFv
papone	FvFvNv	potumo	FvFvNv
papodo	FvFvVv	potubi	FvFvVv
pamatu	FvNvFv	ponepa	FvNvFv
pamane	FvNvNv	ponemo	FvNvNv
pamado	FvNvVv	ponebi	FvNvVv
pabetu	FvVvFv	podopa	FvVvFv
pabene	FvVvNv	podomo	FvVvNv
pabedo	FvVvVv	podobi	FvVvVv
mobetu	NvVvFv	madopa	NvVvFv
mobene	NvVvNv	madomo	NvVvNv
mobedo	NvVvVv	madobi	NvVvVv

Test-phase 2			
PTP	**Manner**	**TPP**	**Manner**
potupa	FvFvFv	tupapo	FvFvFv
potumo	FvFvNv	tupama	FvFvNv
potubi	FvFvVv	tupabe	FvFvVv
ponepa	FvNvFv	tumopo	FvNvFv
ponemo	FvNvNv	tumoma	FvNvNv
ponebi	FvNvVv	tumobe	FvNvVv
betupa	VvFvFv	dopapo	VvFvFv
betumo	VvFvNv	dopama	VvFvNv
betubi	VvFvVv	dopabe	VvFvVv
benepa	VvNvFv	domopo	VvNvFv
benemo	VvNvNv	domoma	VvNvNv
benebi	VvNvVv	domobe	VvNvVv

Test-phase 3			
PPT	**Manner**	**TPP**	**Manner**
papotu	FvFvFv	tupapo	FvFvFv
Papone	FvFvNv	tupama	FvFvNv
papodo	FvFvVv	tupabe	FvFvVv
pabetu	FvVvFv	tubipo	FvVvFv
pabene	FvVvNv	tubima	FvVvNv
pabedo	FvVvVv	tubibe	FvVvVv
mopotu	NvFvFv	nepapo	NvFvFv
mopone	NvFvNv	nepama	NvFvNv
mopodo	NvFvVv	nepabe	NvFvVv
mobetu	NvVvFv	nebipo	NvVvFv
mobene	NvVvNv	nebima	NvVvNv
mobedo	NvVvVv	nebibe	NvVvVv

Appendix C

Mean Cohort Densities with a cut-off-point of 3 phonemes of the items used in the different test phases of Experiment 1-4.

Exp.	Test-phase	Cohort density 1	Cohort density 2
1, 3	Test-phase 1	PvTvPv 16.89 (12.57)	PvPvTv 16.28 (20.30)
	Test-phase 2	PvTvPv 29.09 (15.57)	TvPvPv 28.24 (17.02)
	Test-phase 3	PvPvTv 16.28 (20.30)	TvPvPv 16.13 (15.75)
2	Test-phase 1	PvTvPv 22.00 (13.35)	PvPvTv 14.66 (14.13)
	Test-phase 2	PvTvPv 37.89 (30.21)	TvPvPv 36.36 (25.29)
4	Test-phase 1	TvPvTv 39.36 (37.57)	TvTvPv 56.33 (55.87)
	Test-phase 2	TvPvTv 39.36 (37.57)	PvTvTv 49.28 (41.00)
	Test-phase 3	TvTvPv 56.33 (55.87)	PvTvTv 49.28 (41.00)

Appendix D:
Test items in Experiment 2.

F = voiceless bilabial, V = voiced bilabial, N = nasal bilabial.

Test-phase 1				**Test-phase 2**			
PvTvPv		**TvPvPv**		**PvTvPv**		**PvPvTv**	
petube	FvFvVv	tupubi	FvFvVv	petupo	FvFvFv	popitu	FvFvVv
petuma	FvFvNv	tupumo	FvFvNv	petube	FvFvVv	popide	FvFvVv
pedepo	FvVvFv	tubope	FvVvFv	petuma	FvFvNv	popina	FvFvNv
pedema	FvVvNv	tubomo	FvVvNv	bodepo	VvVvFv	bebatu	VvVvFv
penabe	FvNvVv	tumibi	FvNvVv	bodebe	VvVvVv	bebade	VvVvVv
botuma	VvFvNv	depumo	VvFvNv	bodema	VvVvNv	bebana	VvVvNv
bodepo	VvVvFv	debope	VvVvFv	bonapo	VvNvFv	bemotu	VvNvFv
bodema	VvVvNv	debomo	VvVvNv	bonabe	VvNvVv	bemode	VvNvVv
bonapo	VvNvFv	demipe	VvNvFv	bonama	VvNvNv	bemona	VvNvNv
mitube	NvFvVv	napubi	NvFvVv	midepo	NvVvFv	mabatu	NvVvFv
midepo	NvVvFv	nabope	NvVvFv	midebe	NvVvVv	mabade	NvVvVv
minapo	NvNvFv	namipe	NvNvFv	midema	NvVvNv	mabana	NvVvNv

Appendix E

Test items in Experiment 4.

F = voiceless bilabial, V = voiced bilabial, N = nasal bilabial.

TvTvPv		**TvPvTv**		**PvTvTv**	
tudipo	FvVvFv	tabetu	FvVvFv	podota	FvVvFv
tudibe	FvVvVv	tabedo	FvVvVv	pododi	FvVvVv
tudima	FvVvNv	tabene	FvVvNv	podono	FvVvNv
tunopo	FvNvFv	tamatu	FvNvFv	poneta	FvNvFv
tunobe	FvNvVv	tamado	FvNvVv	ponedi	FvNvVv
tunoma	FvNvNv	tamane	FvNvNv	poneno	FvNvNv
dotapo	VvFvFv	dipotu	VvFvFv	betuta	VvFvFv
dotabe	VvFvVv	dipodo	VvFvVv	betudi	VvFvVv
dotama	VvFvNv	dipone	VvFvNv	betuno	VvFvNv
netapo	NvFvFv	nopotu	NvFvFv	matuta	NvFvFv
netabe	NvFvVv	nopodo	NvFvVv	matudi	NvFvVv
netama	NvFvNv	nopone	NvFvNv	matuno	NvFvNv

CHAPTER 4

OCP-PLACE FOR SPEECH PROCESSING: A BIAS ON PERCEPTION OR ACQUISITION? THE CASE OF MANDARIN CHINESE[1]

4.1 Introduction

In any language, possible words are restricted by phonotactic restrictions that define which phonemes may or may not co-occur. It has often been observed that certain phonotactic restrictions can be found across many genetically and geographically unrelated languages. Cross-linguistic attestations of certain restrictions have led to the assumption of markedness constraints, grounded in universal biases. Two possible biases have been proposed: On the one hand, cross-linguistically frequent patterns may reflect functional biases on speech perception and production (i.e., channel bias). On the other hand, it may reflect cognitive biases on language acquisition (i.e., analytic bias) by which the learner innately focuses on certain aspects of the input. A third, albeit unlikely, possibility is that their typological frequency is purely coincidental (for an overview, see Moreton, 2008).

An elegant approach to study the effect of biases on production, perception and acquisition is to test listeners or learners whose native language lacks direct evidence for a particular cross-linguistically well-attested markedness constraint. Using this methodology, it has been shown, for example, that listeners have well-formedness intuitions about phonotactic patterns that are unattested in their native (L1) or second language (L2), which can only be accounted for by

[1] This study, submitted for publication, was carried out in collaboration with Ao Chen and René Kager, Utrecht University. The study was funded by an NWO grant (277-70-001) awarded to René Kager. The authors contributed to this study as follows: Hypothesis: Natalie Boll-Avetisyan and René Kager; design: Natalie Boll-Avetisyan, René Kager, and Ao Chen; testing: Ao Chen, analysis and report: Natalie Boll-Avetisyan.

intuitions about phonological markedness (e.g., Dong, 2008; Moreton, 2002). Furthermore, it was found that markedness constraints influence the perception of clusters unattested in the native language (e.g., Berent, Steriade, Lennertz, & Vakninc, 2006). To our knowledge, only one study (Moreton, 2008) has tried to disentangle whether the hidden knowledge of markedness constraints governing the sonority contour of syllable onsets had its basis in a functional bias on perception, or alternatively in a UG-based cognitive bias on acquisition. This study compared the acquirability of two phonotactic patterns in an artificial language learning paradigm. The result was that native speakers of English acquired a harmony pattern of vowel height in non-adjacent vowels and of voice in non-adjacent consonants in CVCV nonwords. However, they did not acquire a vowel-voice pattern on adjacent VC in CVCV items. As vowel harmony is the typologically well-attested pattern, the result was interpreted as evidence for an analytic bias on acquisition (Moreton, 2008).

No study, to our knowledge, has yet tested whether markedness constraints influence speech processing, such as speech segmentation, independently of whether they affect the L1 of a listener. The current study will address this while separating two questions: (1) Are markedness constraints due to perceptual biases and affect processing independent of whether they constrain the L1? (2) If not, can they be used for processing by listeners who have acquired an L2 that is restricted by such a markedness constraint?

Among the cross-linguistically well-attested phonotactic constraints, we are particularly interested in constraints that affect non-adjacent phonemes, because such constraints have been claimed to be hard to acquire, as will be discussed below. Constraints on non-adjacent dependencies classically affect sequences of phonemes that share features. Many phonotactic restrictions require phonemic segments to be similar within words, as, for example, vowel harmony (for an overview, see Archangeli & Pulleyblank, 2007) and consonant harmony (for an overview, see Rose & Walker, 2004). On the one hand, a possible account for the typological frequency of harmony constraints is that they are due to an innate markedness constraint. On the other hand, they might be a result of a general cognitive bias for a preference for grouping similar sounds (e.g., Thorpe & Trehub, 1989). Another functional account is that phonologization occurs as a result of listeners failing to compensate for coarticulation in perception (e.g., Ohala, 1994).

Languages display disharmony patterns as well, i.e., when certain phonemes within words *may not* be similar. A well-attested example is OCP-PLACE, a constraint that will be the focus of the current study. As opposed to vowel or consonant harmony, OCP-PLACE favors featural disagreement between phonemes within words or morphemes by restricting co-occurrences of consonants with the shared feature [place]. The effect of OCP-PLACE crosses intervening vowels, constraining the co-occurrence of consonants that are non-

adjacent (in formal accounts, this follows from consonants being represented as adjacent on the "consonantal tier," where there may not be two adjacent consonants with shared place (e.g., McCarthy, 1986). Strong OCP-PLACE effects are, for example, found in Semitic languages such as Arabic (e.g., Frisch, Pierrehumbert, & Broe, 2004; Greenberg, 1950; McCarthy, 1985) and Hebrew (e.g., Berent & Shimron, 1997; McCarthy, 1985). Furthermore, and interestingly in light of the current study, many genetically and geographically unrelated languages display a less radical, yet systematic under-representation of consonant co-occurrences with shared [place] (e.g., Muna: Coetzee & Pater, 2008; Niger-Congo languages: Pozdniakov & Segerer, 2007; Dutch: Shatzman & Kager, 2007). In the latter type of languages, words containing pairs of consonants with shared place features do occur, such as /pVm/ in English *spam* or /mVb/ in *mob*. However, they occur less often than expected if non-adjacent consonants co-occurred at random.

Several studies have found native listeners to have unconscious knowledge of OCP-PLACE. Nonwords that violate OCP-PLACE are judged to be less well-formed than nonwords that do not in studies with native listeners of Hebrew (Berent & Shimron, 1997), Arabic (Frisch & Zawaydeh, 2001) and English (Coetzee, 2008). Furthermore, OCP-PLACE affects speech processing. In lexical decision tasks, nonwords that violate OCP-PLACE are rejected faster than nonwords composed of consonants that do not share place by native listeners of Hebrew (Berent, Everett, & Shimron, 2001) and Dutch (Shatzman & Kager, 2007). Also, OCP-PLACE influences perception: forms that violate OCP-PLACE are perceptually avoided by native listeners of English (Coetzee, 2005). Moreover, recent studies (Boll-Avetisyan & Kager, submitted; Kager & Shatzman, under revision) have demonstrated the use of OCP-PLACE as a cue for speech segmentation. In a word-spotting task, Kager and Shatzman (under revision) showed that native listeners of Dutch were faster at finding words with initial labial (=P) (e.g., *boek* 'book') when embedded in a nonsense string that started in a P (e.g., *piboek*) than embedded in a string starting with T (e.g., *tiboek*). Furthermore, Boll-Avetisyan & Kager (submitted) familiarized native listeners of Dutch with an artificial speech stream without pauses in which coronal-initial Cv syllables (hereafter, Tv) were always followed by two labial-initial Cv syllables (hereafter, Pv), resulting in …$PV_1PV_2TvPV_1PV_2TvPV_1PV_2$….

In a subsequent forced choice task, participants had to indicate which of two items—which in fact had both occurred in the artificial language—were part of the language they had been familiarized with. As a result, PVTVPV items were preferred over PVPVTV and TVPVPV items. This indicates that participants must have inserted a boundary between PV_1PV_2, revealing their knowledge of OCP-PLACE. As yet, all studies on the effects of OCP-PLACE in speech processing have tested speakers of languages that display an underrepresentation of consonant pairs with shared place. This raises the question as to whether speakers of languages without OCP-PLACE effects in the lexicon would also use the constraint, for example, as a cue for speech segmentation.

The use of vowel harmony as a speech segmentation cue is language-specific, as prior studies have shown. It affects speech segmentation in infant and adult native listeners of Finnish and Turkish, whose native language is restricted by vowel harmony (Kabak, Maniwa, & Kazanina, 2010; Suomi, McQueen, & Cutler, 1997; Van Kampen, Parmaksiz, van de Vijver, & Höhle, 2008; Vroomen, Tuomainen, & de Gelder, 1998). However, in a cross-linguistic comparison, only native listeners of Finnish, but not native listeners of French or Dutch experienced facilitation of statistical learning and segmentation of the artificial language from vowel harmony cues (Vroomen, et al., 1998). If vowel harmony were due to a general perceptual preference for grouping similar sounds, then the French and Dutch listeners should also use that cue for speech segmentation. This, in addition to the results of Moreton's study (2008), suggests that vowel harmony patterns are *not* due to a channel bias.

In light of these findings, it will be interesting to find out whether the use of OCP-PLACE as a segmentation cue is language-independent. Studies of the literature and our own studies of lexical databases of languages revealed that it is difficult to find a language that is not restricted by OCP-PLACE. This observation has influenced many researchers to consider OCP-PLACE to be an innate linguistic primitive (e.g., McCarthy, 1986; Pozdniakov & Segerer, 2007), which would relate to analytic biases on acquisition. Fewer studies have found functional accounts for the typological frequency of OCP-PLACE. Boersma (1998) argues that OCP cannot be a linguistic primitive, as its effects across languages are too divergent to be captured by one and the same principle. In his view, the listener uses OCP to prepare lexical access after retrieving a perceptual form. The constraint, however, interacts in a language-specific and hence acquired ranking with other constraints (e.g., Boersma, 1998, 2000; Frisch, 2004).

Frisch (2004) proposes that identity avoidance creates a functional advantage in speech processing. In perception, speech sounds are serially encoded. Once a phoneme is encoded, its feature nodes are immediately inhibited to speed up the encoding of the following phoneme. Thus, Frisch assumes that the second of two consonants with shared features will not be correctly encoded because of an inhibition of the repeated feature which is important for its recognition. This theory may well explain effects of OCP-PLACE on word processing. However, it cannot explain how OCP-PLACE can be used as a cue for speech segmentation (see Boll-Avetisyan & Kager, submitted). Just as with words, continuous speech also needs to be encoded serially. Hence, for instance, in /pvm/ across a word boundary in *happy man*, the word onset /m/ would not be properly encoded due to inhibition of the labial feature after activating the preceding word offset /p/. This predicts that word recognition of *man* after *happy* would be slowed down. This, however, is not confirmed by the results of the word-spotting experiment by (Kager & Shatzman, under revision). Here, listeners were actually faster at

spotting P-initial words if they were preceded by a P than when they were preceded by a coronal (= T) consonant. These results suggests that listeners use knowledge of OCP-PLACE as a cue to detect word boundaries in the signal. Hence, in light of the segmentation results, Frisch's theory of an inhibited encoding of the second of two similar consonants can only be right if inhibition of the place feature stops at the end of a word or a morpheme. If that were the case, then it would be difficult to account for the results of Boll-Avetisyan & Kager's (submitted) study. Here, OCP-PLACE was used as a cue for segmenting an artificial language that was free of lexical information. All in all, these studies suggest that OCP-PLACE is a good cue for speech segmentation that is independent of lexical access and cannot be explained by a functional bias on perception.

If the theory that a use of OCP-PLACE in speech segmentation arises due to channel bias is wrong, then two options remain. Either, it is due to analytic bias on acquisition, or its frequent attestation is coincidental. Results from learning studies suggest that there might be a special role for similarity in non-adjacent dependency learning. Constraints on non-adjacent dependencies are difficult to acquire from distributions in the input, as the learner must abstract over the intervening phonemes that have different features. It has been claimed that non-adjacent dependencies can only be learned if the units are similar at least to some degree (e.g., Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005; Perruchet & Pacton, 2006). In fact, in most studies that demonstrated non-adjacent dependency learning, the non-adjacent dependency held between elements that were similar to the extent that they were either both vowels or both consonants, or even shared articulation or place features (Baer-Henney & Vijver van de, submitted; Boll-Avetisyan et al., submitted; Bonatti, Peña, Nespor, & Mehler, 2005; Newport & Aslin, 2004; Onnis, et al., 2005; Pons & Toro, 2010). Such results may be interpreted as evidence for an analytic bias on acquiring non-adjacent dependencies between similar elements.

Together, the results from ALL studies suggest that constraints on non-adjacent dependencies should be able to be acquired from input distributions, especially if the dependency holds for phonemes that belong to the same major class. It needs to be kept in mind that OCP-PLACE does not require similarity but "similarity avoidance" within words (Frisch, et al., 2004). Hence, if learners acquire OCP-PLACE from the input, they need to learn that consonants with shared place are unlikely to co-occur. In principle this should be possible, as it has been shown that infants already display language-specific knowledge of phoneme co-occurrences that are low in probability (Boll-Avetisyan, et al., submitted).

We are interested in whether OCP-PLACE restrictions need to be acquired from the input,[2] or whether knowledge of OCP-PLACE is pre-given (by channel bias). In order to shed light on this issue, two kinds of experiments are required. First, it needs to be tested whether native speakers of a language that does not display OCP-PLACE effects in the lexicon use OCP-PLACE for speech processing. Second, it needs to be tested whether *learners* acquire OCP-PLACE.

Previous studies have found highly proficient L2 learners to use knowledge of L2 phonotactic probabilities for speech processing, if the phonotactic structures are not illicit but merely occur with different probabilities in the L1 (Majerus, Poncelet, Van der Linden, & Weekes, 2008; Trapman & Kager, 2009), and even if they are L1 illegal sequences (Altenberg, 2005; Boll-Avetisyan, 2011; Lentz & Kager, submitted). Such results suggest that phonotactics can be acquired not only in an L1, but also in an L2. Furthermore, highly proficient L2 learners have been shown to use phonotactic knowledge of adjacent consonant clusters as a cue for speech segmentation (Lentz & Kager, in prep., under revision; Weber & Cutler, 2006).

To our knowledge, no study has yet looked at whether phonotactic constraints on non-adjacent dependencies can be acquired in an L2. In light of the results reported above, we would predict that L2 learners can acquire an underrepresentation of co-occurrences of non-adjacent consonants with shared place. If they can acquire this knowledge, then they might also use it as a speech segmentation cue.

Furthermore, only few studies on L2 phonotactic perception have addressed the issue of the learning conditions in which phonotactic knowledge can be developed. Some studies found L2 proficiency to correlate with L2 phonotactic perception (e.g., Boll-Avetisyan, 2011; Trapman & Kager, 2009). A measure of L2 proficiency, however, cannot be seen as a factor that explains phonotactic acquisition. It rather gives a general indication of how good an L2 learner is at producing or comprehending various linguistic aspects of an L2. We assume that the acquisition of L2 phonology, as it is acquired sub-consciously, should strongly depend on the amount of exposure to the spoken language. L2 proficiency, however, can be determined by many other factors that do not relate to exposure at all, such as motivation, attention, and formal education. Boll-Avetisyan (2011) as well as Trapman and Kager (2009) have only tested L2 learners who lived and had acquired the L2 in the country where the language is spoken. There are indications by other studies that the amount of input from native speakers, defined by the Length of Residence (LoR) in the L2-speaking country, correlates with L2 phonotactic perception (e.g., Altenberg, 2005; Flege

[2] It will be discussed in the general discussion, whether OCP-PLACE is possibly acquired through statistical learning, and possibly supported by an analytic bias.

& Liu, 2001). It will be interesting to test whether the amount of input spoken by native speakers independently of general L2 proficiency plays a crucial role.

The current study addresses the following research questions:

(1) Is the use of OCP-Place in speech processing due to channel bias on perception, or due to language-specific knowledge? We hypothesize that OCP-Place effects on speech segmentation are not due to channel bias but are acquired from the input, possibly by analytic bias. Hence, listeners should only use OCP-Place in actual speech processing, e.g., in speech segmentation, if their native language displays effects of OCP-Place in the lexicon.

(2) Can OCP-PLACE be acquired and used in speech processing in an L2? Our hypothesis is that OCP-PLACE can be acquired and used for speech processing in an L2 that restricts its lexicon by this constraint.

(3) What are the factors that determine a successful L2 acquisition of OCP-PLACE? Phonotactic cues for segmentation should be acquired as a function of experience with listening to continuous speech. Hence, a large amount of exposure to input spoken by native speakers should be a pre-requisite for their acquisition. We hypothesize that not only general L2 proficiency by itself but the amount of exposure to the L2 spoken by native speakers leads to a successful use of OCP-PLACE in L2 speech processing.

To answer the research questions, a new language is brought into the scope of OCP-PLACE: Mandarin Chinese. As opposed to many other languages, Mandarin Chinese is not restricted by OCP-PLACE, as we will show in the following section by a corpus analysis based on a lexical database of Mandarin Chinese. To answer the research questions regarding segmentation, we ran a series of artificial language learning experiments with native speakers of Mandarin Chinese, and Mandarin Chinese second language learners of Dutch.

4.2. Corpus study: OCP-PLACE in Mandarin Chinese

4.2.1. Method

Measurement: Observed/Expected ratio

As a measurement, we used the Observed/Expected (O/E) ratio (Pierrehumbert, 1993). This measure compares the observed counts of consonant pairs to the counts that would be expected if consonants combined at random (Frisch, et al., 2004). O stands for the number of C_1vC_2 sequences in a corpus or lexicon. It is divided by E, computed as the probability that C_1 occurs in the initial position of CvC, multiplied by the probability that C_2 occurs in the final position of CvC, which is multiplied by the total number of CvC sequence tokens:

$O/E = N(C_1vC_2) / p(C_1) * p(C_2) * N(CvC)$

$N(C_1vC_2)$: the number of C_1vC_2 sequences; N(CvC): the number of CvC sequences

$p(C_1)$, $p(C_2)$: the probability of C_1 and C_2 calculated as proportions of consonants

O/E > 1: sequence is over-represented; O/E < 1: sequence is under-represented

Observed/Expected ratio for C_1vC_2 sequences in Mandarin Chinese

We calculated the probability of C_1vC_2 sequences in which both Cs shared place in the CADCC Lexical Database, a Mandarin Chinese lexicon containing phonological transcriptions of telephone conversations (Li & Xu, 2001).[3] Each word type is denoted with its frequency of occurrence, amounting to 266,642 word tokens. We based our calculations on the consonant categories proposed by Duanmu (2000), who classifies four of the 19 Mandarin consonants as labials {/p^h, p, f, m/}, 11 as coronals {/t, t^h, ts, ts^h, s, n, l, tʂ, $tʂ^h$, ʂ, ʐ/} and 4 as dorsals {/k, k^h, x, ŋ/}. Consonant distributions are examined without distinguishing between different tones, i.e., all the words in the corpus are considered for analysis.

Calculations were based on both word types and word tokens. Word type and token frequencies of phonotactic distributions may vary significantly. In the literature, it has been reported that type frequency counts better reflect the phonotactic knowledge used in online speech processing than token frequency

[3] Thanks to Frans Adriaans for providing us with the software PhonotacTools, which we used for our calculations.

(e.g., Pierrehumbert, 2001). However, for completeness, we also calculated how consonants are distributed in Mandarin Chinese word tokens. It has often been noted that the effect of OCP-PLACE is morpheme-based. However, it cannot be in Mandarin Chinese, as here, no morpheme extends a simple CV(N) syllable, in which N is a nasal /n/ or /ŋ/, but not /m/. Hence, Mandarin Chinese morphemes are too short to carry effects of OCP; in particular, they are too short to carry effects of OCP-LABIAL.

The counts of CVC in word types were enriched by two further comparisons. That is, we calculated the O/E ratio for co-occurrences of labials when excluding [w], since in the Beijing area, [w] is sometimes realized as [v], for which [v] has been considered a free variant allophone of [w]. Mandarin Chinese uses syllable reduplication as a morphological process for word formation. As we are interested in OCP-PLACE as a constraint independent of the effect of reduplication, we furthermore calculated the O/E ratio over a lexicon excluding reduplicated forms (901 occurrences were taken out, with 265741 words remaining in the corpus). Bisyllabic words consisting of two identical syllables with different tones were not considered as reduplicated forms. However, if the second syllable is realized with neutral tone, the bisyllabic words are treated as reduplicated forms.

4.2.2. Results and discussion

Results

As can be read from Table 1, the O/E value of consonant co-occurrences with shared [place] in CVC sequences is always around 1.00, indicating that pairs of consonants with shared [place] are neither over- nor underrepresented.

Basis of counts	Type of words	PvP	KvK	TvT
Word types	General	1.00	1.43	1.11
	Excl. reduplications	0.91	1.23	1.23
	Identical Cs	1.42	2.31	1.35
Word tokens	General	1.42	0.70	0.94
	Identical Cs	2.87	0.58	1.11

Table 1: O/E for all CVC with C_1 and C_2 sharing [place]

Among word types, this holds in particular for PVP (O/E = 1.00). Co-occurrences of dorsals (K) and coronals (T) are even slightly overrepresented (O/E for KVK = 1.43; O/E for TVT =1.11). A calculation over word tokens similarly suggests no role for OCP-PLACE in Mandarin Chinese. Sequences of KVK turn out to be slightly underrepresented (O/E = 0.70). However, distributions of Ts are neutral (O/E = 0.94), and PVP sequences even turn out to be slightly overrepresented (O/E = 1.42).

Among word types, the O/E value for PVP sequences remains 1.00 with or without [w]. When excluding reduplications, again, pairs of consonants with shared [place] turn out to neither be over- nor underrepresented (PVP: O/E = 0.91, KVK: O/E = 1.23; TVT: O/E = 1.23).

Discussion

Consonant distributions in Mandarin Chinese do not justify the assumption of OCP-PLACE in this language. If at all, one would have to propose a harmonic (anti-OCP) effect for pairs of Ks[4] and possibly Ts, as they are somewhat over- instead of underrepresented.[5]

The O/E ratios for Cs in CVC sequences with shared [place] in Mandarin Chinese differ strongly from the O/E ratios in the many other languages that have been described in previous studies (e.g., Coetzee & Pater, 2008; Frisch, et al., 2004; Kager & Shatzman, under revision). In Dutch, which we selected to be the L2 of the learners tested in our study in Experiments 2 and 3, the O/E ratio of PVP in word types in the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995) is 0.44, for TVT 0.75, and for KVK 0.50. In word tokens, the Dutch values are as follows: PVP = 0.29, TVT = 0.77, and KVK = 0.50 (values are taken from Kager & Shatzman, under revision). Such values can clearly be interpreted as a presence of an effect of OCP-PLACE in Dutch.

All in all, Mandarin Chinese provides a completely different picture with respect to the distributions of consonants with shared [place] as many other languages. Hence, the Mandarin Chinese data shows that OCP-PLACE is not a constraint that universally affects the lexica of all natural languages. This might be that it cannot affect the Mandarin Chinese lexicon, because its morphemes are too short.

The absence of OCP-PLACE effects in the Mandarin Chinese lexicon presents an ideal case for testing the assumption that the use of OCP-PLACE in speech

[4] The over-representation of KVK may be due to the limited diversity of dorsal consonants and that only nasals are allowed in coda position, i.e., /n/ and /ŋ/.

[5] However, counts of KVK in word tokens are under-represented, suggesting a potential role for OCP-DORSAL.

perception is a reflection of universal, innate preferences by the listener. To address this, we carried out an artificial language segmentation experiment similar to the one by Boll-Avetisyan & Kager (submitted). If OCP-PLACE is directly due to universal perceptual preferences, we would expect native speakers of Mandarin Chinese to be sensitive to OCP-PLACE violations in online processing tasks in spite of evidence for OCP-PLACE in the Mandarin lexicon. If this is the case, then they should segment the artificial language such that a stream of …$PV_1PV_2TvPV_1PV_2TvPV_1PV_2Tv$... into …$PV_1$—$PV_2TvPV_1$—$PV_2TvPV_1$—$PV_2Tv$…. In a test phase, when asked which words they think are part of the artificial language, PVTVPV words should be preferred over PVPVTV (and TVPVPV) words. If, however, as we hypothesize, OCP-PLACE is not a perceptual universal, but arises from cognitive biases on acquisition, then speakers of a language that does not display OCP-PLACE effects in the lexicon should not use it for speech segmentation. Hence, we predict that monolingual speakers of Mandarin Chinese will not show any OCP-based segmentation preference.

4.3. Experiment 1

4.3.1. Method

Participants

Eighteen monolingual native speakers of Mandarin Chinese were tested in Beijing. Three of them had to be excluded because of recent exposure to English and another one was excluded for not having understood the experiment instructions. Among the 14 included participants (6 male, 8 female, M_{age} = 44 years (*SD* = 10.53), age range: 24-53 years), five had had occasional courses in English during high school, but by the time of the experiment, none of them had been exposed to English for at least the past four years. They never used English in their daily life, and none of them considered him- or herself to be able to communicate in English. None of the participants had lived in countries other than China. Two participants were raised in an area where a Chinese dialect different from Mandarin was spoken, but both of them were exposed substantially to Mandarin from birth and consider Mandarin as a native language. All other participants were truly monolingual native speakers of Mandarin, raised in Beijing without exposure to other dialects. All but two participants had a college degree from a university in China. All participants reported normal hearing. All were compensated for their participation.

Material

The artificial language

As in Boll-Avetisyan & Kager (submitted), we selected nine syllables that were assigned to one of three slots and concatenated to a continuous stream of speech without pauses (e.g., …bubidebumudepepeponamimute…). All syllables started in either P {/p, b, m/} or T {/t, d, n/}. Three P_{V_1} syllables were always followed by one of the three P_{V_2} syllables, which were followed by one of the three T$_V$ syllables. TPs between both adjacent and non-adjacent syllables were always 0.33. Hence, the language itself did not contain a distributional cue for segmentation.

The artificial language was further controlled for factors that could potentially affect segmentation preferences other than OCP-LABIAL. First, the fundamental frequency (F0) of the artificial language was kept flat to control for prosodic cues such as stress, tone, or intonation. As syllables, we selected P_{V_1} = {/pe, bu, mi/}, T$_V$ = {/te, de, na/}, and P_{V_2} = {/po, bi, mu/}. The selection was based on controls for lexical statistics of Mandarin Chinese, aiming to remove any effects of segmental distribution of the native language in segmentation. Since native listeners of Mandarin Chinese perceive flat prosody as a high level tone (tone 1), controls for lexical statistics were based on words composed of tone 1 syllables. In order to minimize interference from lexical knowledge, all consonants were combined with vowels amounting to biphones with the lowest biphone frequency and a low cohort density with a cut-off point of 2 phonemes (see Appendix A).

The artificial language stream was synthesized using the text-to-speech software MBROLA (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996). Due to the fact that at present no biphone set of Mandarin Chinese is available for synthesis, we had to use the voice of a different language. For the purpose of our study, it was crucial that participants could discriminate between the labial plosives /p/ and /b/ well. We had to select the voice of a language that would contrast /p/ and /b/ by the feature [aspiration] (e.g., Turkish, German) and not by the by the feature [voice], which cannot be perceived by Mandarin listeners, but is used in many other languages (e.g., French, Dutch).

We selected a Turkish male voice "tr1." The synthesizer produced the stream with a monotone pitch (F0 = 100 Hz) with an average syllable duration of 232 ms.[6] Segment durations were based on average natural durations in Mandarin Chinese. The syllables were concatenated to a ten minutes long stream without

[6] We thank Arthur Dirksen from "Fluency" for a script to generate the stream.

pauses, consisting of 2700 syllables in total. Intensity of the stream faded in for the first and out for the last five seconds, so that the endpoints (utterance boundaries) did not give a cue for segmentation. Two native speakers of Mandarin Chinese listened to the artificial language and reported to be able to discriminate between voiced and unvoiced consonants. Furthermore, they heard no familiar words in the stream. None of the syllables occurred to them with a particular stress or tone other than tone 1.

Test stimuli

We synthesized two types of stimuli: PvTvPv items (e.g. *penamu*) and PvPvTv items (e.g. *pobude*). All 27 PvPvTv and all 27 PvTvPv items were included and each were repeated twice. Items were presented in 54 test pairs. Each pair contrasted one PvTvPv with a PvPvTv item, and each time mapped with a different word from the alternative category.

In order to make sure that the artificial language synthesized using Turkish voices does not interfere with the perception of the individual segments, 14 native speakers of Mandarin Chinese[7] were first tested on their perception of the test items in a non-word repetition task. The repetition task had the purpose of being able to exclude the possibility that items were not correctly perceived due to the non-native Turkish synthesis of the material. Participants received the 54 PvTvPv and PvPvTv items over headphones. After hearing an item, they were required to write down what they heard, either in Hanyu Pinyin or in Latin alphabet, depending on their preference. When they finished writing, they could continue with the next item by clicking the mouse on the button "next" on the screen. The test confirmed that participants did not have severe or systematic difficulties with perception even though the stimuli were synthesized with a Turkish voice.[8]

Apparatus and procedure

Each participant was tested individually in a quiet room. The experiment was carried out using a laptop. Stimuli were received over headphones. The procedure, originally borrowed from Peña et al. (2002), was as in Boll-Avetisyan & Kager (submitted). Participants first entered a pre-training phase, in

[7] The 14 participants were the same that were tested in Experiment 3 of this study.

[8] Random occasional confusions in both directions occurred between /t/ and /d/ (two participants: /t/ > /d/, two participants: /d/ > /t/), and /b/ and /p/, respectively (two participants /b/ > /p/; another two participants: /p/ > /b/). Five participants sometimes perceived /n/ as /m/, which renders a PvPvPv perception instead of PvPvTv or PvTvPv sequences. Two participants occasionally misperceived nonword-initial /p/ as /t/ (i.e. PvTvPv > TvTvPv, and PvPvTv > TvPvTv).

which they were required to find the syllable /so/ in a pair of CV syllables. The purpose of the pre-training phase was to acquaint participants with the two-alternatives-forced-choice task. Subsequently, participants entered the familiarization phase. They were informed that they would listen to an artificial language for the next ten minutes and were instructed to listen carefully, because they would later be tested on their knowledge about the words of this language. After the familiarization phase, participants were tested in a two-alternative forced choice task, in which they had to indicate, via a mouse-click on a button on screen, whether the first or the second of two items was more likely to be a word from the language they had just heard. Test pairs were presented with an inter-stimulus interval of 500 ms and an inter-trial interval of 2000 ms. After the experiment, the participants filled in a short language-history questionnaire.

4.3.2. Results and discussion

Results

As a result, participants did not significantly prefer one type of item over the other (53% preference for PVTVPV over PVPVTV, binomial test: $p > .5$). As displayed in Figure 1, the preferences of the individual participants clustered around chance, indicating that decisions were random.

Discussion

For the artificial language learning task, it was predicted that Mandarin Chinese participants should use OCP-PLACE as a cue to set boundaries in the artificial language, if the typological frequency of OCP-PLACE were due to a universal preference for an avoidance of consonants with shared place of articulation. However, contrary to native listeners of Dutch, the native listeners of Mandarins did not prefer a segmentation that would keep consonants with shared [place] apart.

Although it is difficult to draw conclusions from null-results, it is striking that the Mandarin Chinese listeners' non-use of OCP-PLACE for segmenting the artificial language coincides with the distribution of consonants in the Mandarin Chinese lexicon, in which pairs of consonants with shared [place] are, unlike Dutch, not underrepresented. This may suggest that a dis-preference for pairs of consonants with shared [place] is not merely due to a cognitive or perceptual preference for identity avoidance. It rather suggests that OCP-PLACE is acquired from consonant distributions in the input language, possibly by cognitive bias.

This hypothesis will now directly be tested with native speakers of Mandarin Chinese who are advanced L2 learners of Dutch in Experiment 2. Recall that in Dutch word types, /PVP/ is underrepresented (O/E = 0.45, see Kager &

Shatzman, 2007), and native listeners of Dutch use this underrepresentation as a segmentation cue (Boll-Avetisyan, submitted; Kager & Shatzman, under revision). Furthermore, highly proficient L2 learners have been found to have knowledge of L2 probabilistic phonotactics (e.g., Boll-Avetisyan, 2011; Majerus, et al., 2008; Trapman & Kager, 2009). Hence, for Experiment 2, we predict that native speakers of Mandarin Chinese who are advanced learners of Dutch should have knowledge of the underrepresentation of /PVP/ in Dutch and insert word boundaries between two labials, given they experience much exposure to Dutch spoken by native speakers.

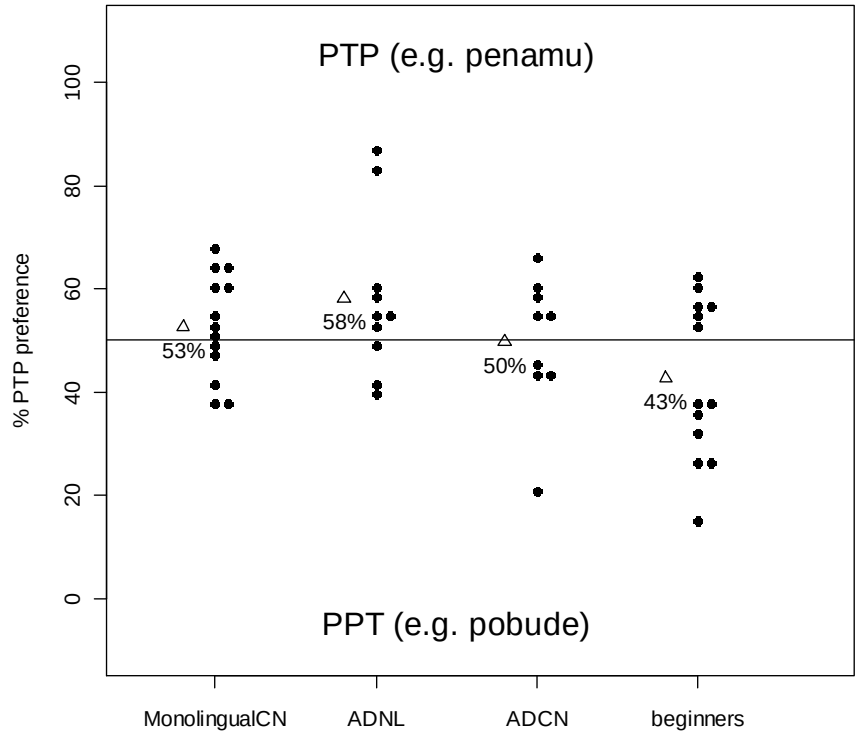

Figure 1: Results of the monolingual native speakers of Mandarin Chinese (MonolingualCN) in Experiment 1, the advanced L2 learners who had learned Dutch in the Netherlands (ADNL) and in China (ADCN) in Experiment 2, and the beginning L2 learners who had started to learn Dutch in the Netherlands (beginners). Each dot represents the average preference for PVTVPV in the test phase. Triangles indicate the group means.

4.4. Experiment 2

4.4.1. Method

Participants

Twenty-two advanced learners of Dutch with Mandarin Chinese as their native language participated in the experiment. All participants reported to have normal hearing. A reasonable sum was paid for their participation. Participants were sub-divided into two groups: Participants in Group 1 were advanced learners of Dutch who lived and had learned Dutch in the Netherlands. Participants in Group 2 were advanced learners of Dutch, who lived and had learned Dutch in China.

For Group 1, we tested 12 participants. Of these, two participants were excluded: one had not understood the instructions, and another turned out to have acquired Dutch in childhood. Among the remaining 10 participants in Group 1 (N = 10, 2 male, 8 female, M_{age} = 36 years (*SD* = 9.37), age range: 26-52 years, $M_{Age_of(Ao)Arrival}$ = 25 (*SD* = 6.73), AoArrival range: 16-36; $M_{Length_of_Residence\ (LoR)}$: 11years (*SD* = 5.46), LoR range: 4-19), all reported to use Dutch frequently in daily life. Their educational level was a bachelor's degree or higher. All participants had knowledge of English. All except one participant reported that they "often" or "always" choose Dutch television or radio stations. All participants were working or studying in the Netherlands. Two of them had obtained a university degree in the Netherlands that was partly taught in Dutch. Furthermore, they spoke Dutch frequently in daily life: "always" in shops, and at least "sometimes" at work or school. Participants in Group 1 had only taken random courses of Dutch after coming to the Netherlands. None has systematically been formally educated in Dutch as a foreign language. They had learned Dutch for the need of living or studying in the Netherlands.

Group 2 (N = 10, 5 male, 5 female, M_{age} = 21 years (*SD* = 1.08) age range: 20-24 years, $M_{AoAquisition}$ = 18 (*SD* = 0.85), AoAquisition range: 17-20) had learned Dutch in China. Nine were studying Dutch as a major in the last year of their study program in Beijing, and had learned Dutch for three years as part of their study program. One had finished a B.A. in the Dutch language, and was employed as a Dutch lecturer at a university in Beijing. As to formal education of Dutch, they all had been taking intensive courses on Dutch pronunciation, listening, reading, writing, and translation taught by professional Dutch teachers as part of their study program. Two had visited a Dutch-speaking country once, and one had visited the Netherlands twice. However, none of the three had spent more than one year continuously in a Dutch-speaking country ($M_{AoArrival}$ = 21 years (*SD* = 1.73), AoArrival range: 20-23 years, M_{LoR} = 0.25 year (*SD* = 0.83),

LoR range: 0-10 months). These three participants were tested during their stays in the Netherlands. The remaining seven were tested in Beijing.

None of them was exposed much to native spoken Dutch during their daily life in China. All had learned English through formal education in China, but none had ever spent time in an English-speaking country. All participants reported that they "always" hear and communicate in Dutch during courses at the university. The courses were mostly given by teachers who were native speakers of Mandarin Chinese. Hence, the quality of their exposure to Dutch in the course cannot be compared to the quality of input by native speakers. Different from the participants in Group 1, most participants in Group 2 did not expose themselves to Dutch through television or radio, with the exception of two participants. Furthermore, these two participants reported that they "often" speak Dutch at work and in shops, while the other participants did not speak Dutch more often than "sometimes." Hence, we can safely conclude that participants of Group 2 did not have much spoken input of native Dutch, while participants of Group 1 did.

Material

The material was the same as those for monolingual Mandarin listeners.

Apparatus and procedure

Apparatus and procedure were as in Experiment 1. Furthermore, a C-test was used to assess the language proficiency of the participants. A C-test is similar to a cloze test with the difference being that a few initial letters are always given, and the participants have to complete the words by filling the gap. Performance in this procedure has been found to correlate with multiple domains of language proficiency (e.g., Eckes, 2010). Here, five short texts with each 20 gaps—i.e., 100 gaps in total—were given. After the two tasks, they filled in a short language-history questionnaire.

4.4.2. Results and discussion

Results of C-test

Seven participants of Group 1 and 9 participants of Group 2 participated in the C-test. Any type of error, no matter whether a gap is not filled at all or whether the error is only a typo, counts as one error point. As a result, participants of group 1 and of group 2 did not differ in terms of their proficiency (Group 1: M_{score} = 38.30 (*SD* = 28.50), Group 2: M_{score} = 39.00 (*SD* = 14.34).

Results of the artificial language learning task

One outlier, whose mean diverged from the standard mean by more than 2 *SD*, was excluded from Group 2. Participants of Group 1, who had learned Dutch in the Netherlands, significantly preferred PVTVPV over PVPVTV items in the test phase (58%, binomial test: $p < 0.001$, two-sided, see Figure 1). Group 2, who had learned Dutch in China did not show a preference for PVPVTV or PVTVPV items (preferences at 50%, binomial test: $p > 0.5$, see Figure 1). The interaction between the two groups did not reach significance in a linear mixed model.

Discussion

The results of Experiment 2 are in line with the prediction: Proficient L2 learners of Dutch insert word boundaries between /PVP/, but only, if they have been exposed to a sufficient amount of input spoken by native speakers of their L2. Hence, the preference for using OCP-PLACE as a cue for speech segmentation must have been acquired with their exposure to spoken Dutch. The fact that segmentation performance did not differ from chance in Group 2 suggests that proficient learners who had acquired Dutch in China did not use OCP-PLACE as a segmentation cue. More evidence for this would have been given by a significant interaction between the two groups, which we might have yielded, had we been able to test more participants. However, this was impossible: in the whole of China there are just 10 Chinese students who enroll in the Dutch study program, which is only offered every four years.

In Experiment 3, we will investigate whether daily exposure to the L2 is enough to acquire the use of OCP-LABIAL as a phonotactic cue for speech segmentation, independent of L2 proficiency, and also independent of age of acquisition. If this were the case, then Dutch-like segmentation preferences should be found with Chinese L2 learners of Dutch who are low in proficiency but live in the Netherlands and are frequently exposed to Dutch. This is unlikely to be the case, given results of a previous study that found no distributional phonotactic knowledge in beginning L2 learners of Dutch who lived in the Netherlands (Trapman & Kager, 2009). However, in this particular study, beginning learners of Dutch had acquired some categorical phonotactic knowledge. Hence, it might still be the case that OCP-LABIAL is salient enough as a phonotactic cue to be picked up from the ambient L2 without proper L2 knowledge by Mandarin Chinese beginning learners of Dutch living in the Netherlands. For Experiment 3, we selected native listeners of Chinese that lived in the Netherlands. Their L2 knowledge of Dutch was minimal. However, due to their residence in the Netherlands they were exposed to native spoken Dutch more often than Group 2 of Experiment 2.

4.5. Experiment 3

4.5.1. Method

Participants

Fifteen native listeners of Mandarin Chinese who had resided in the Netherlands for a short period of time and with limited knowledge of Dutch were tested. Two of them had to be excluded because of equipment failure. Among the 13 remaining participants (N = 13, 11 female, 2 male, M_{age} = 26 years (SD = 3.82), age range: 24-38 years; $M_{AoArrival}$ = 24.23 years (SD = 4.00), AoArrrival range: 22-37 years; M_{LoR} = 1.75 years (SD = 1.34), LoR range: 0.6-4.6 years), all were enrolled in an international graduate study program at the Leiden University, for which English, and not Dutch, was primarily used. Three participants reported that they had knowledge of Dutch. Their self-rated proficiency of Dutch was 1 on average (range: 1-2) on a scale from 1 (none) to 7 (high). All participants reported that they heard Dutch on a daily basis during their courses at the university with their Dutch classmates and in general in their daily life. Eight reported that they heard Dutch from the radio or TV. In total, they were exposed somewhat frequently to the Dutch language; however, they were less exposed to Dutch than the participants of group 1 in Experiment 2.

All participants had learned English as part of their formal education in China, and considered themselves as proficient in English ($M_{self\text{-}rating}$ = 5 (SD = 0.82) range 3.5-6). All used English as the language for their studies in the Netherlands. However, they were hardly exposed to English spoken by native speakers. Three had learned French in school, but none considered him-/herself as proficient enough to communicate in French ($M_{self\text{-}rating}$ = 1.33 (SD = 0.57), range 1-2). None of them had lived in other countries besides China and the Netherlands. None reported hearing problems. A reasonable sum was paid for their participation.

Material

The material was the same as in Experiment 1.

Apparatus and procedure

Apparatus and procedure were the same as in Experiment 1, except that only three participants filled in the questionnaire right after the experiment, and the rest were later interrogated over the phone. Furthermore, no C-test was used, as proficiency was always low.

4.5.2. Results and discussion

Results

As displayed in Figure 1, the participants significantly dispreferred PvTvPv words (43%, binomial test: $p < 0.001$, two-sided). In a linear mixed model, their performance interacted significantly with segmentation preferences by Group 1 in Experiment 2 ($p < 0.05$), but not by Group 2 ($p > 0.2$). Furthermore, an interaction between the beginning learners and the monolingual participants in Experiment 1 was marginally significant (F (25) = 4.02 p = 0.56).

Discussion

The preference for PvPvTv items suggests that beginning learners of Dutch segmented the artificial language by inserting boundaries between /TvP/ rather than /PvP/. Hence, they did segment the language, but not by means of OCP-PLACE. The preference for a PvPvTv segmentation is difficult to explain, as /PvP/ is not overrepresented in Mandarin Chinese word types, and, furthermore, participants in Experiment 1 and in Group 2 of Experiment 2, i.e., the participants with residence in China, did not show this preference, too.

There are some explanations that might account for these results. The purpose of speech segmentation is to segment the speech stream into its smallest lexical units, i.e. into words and the morphemes that compose words (e.g. Hay, 2003), 2000). In Mandarin Chinese, morphemes, the smallest lexical units, are monosyllables shaped Cv(N), in which N is a nasal /n/ or /ŋ/, but not /m/. In light of this, it might not make much sense for a native listener of Mandarin to create segmentation hypotheses about units that extend beyond a syllable. Hence, it is likely that participants of Experiment 1 and Group 2 of Experiment 2, i.e., the groups who had never had much exposure to languages with lexical units of a larger size, may have tried to segment the stream into monosyllables. The lack of the difference between monolinguals and highly proficient L2 learners of Dutch with residence in China may be due to the lack of experience with the actual need to segment Dutch speech produced by native speakers of Dutch.

The beginning learners were by far less proficient in Dutch than the learners who resided in China. However, still they had more exposure to Dutch spoken by native speakers. Maybe the exposure was enough to change their segmentation hypotheses from monosyllabic to larger strings.

As English shows effects of OCP-PLACE (e.g., Berkley, 1994), in theory, the beginning learners could have acquired OCP-PLACE from their daily English conversations. The fact that they did not again suggests that—in line with the

results of Experiment 2—exposure to a language spoken by native speakers is crucial for the acquisition of phonotactic constraints for speech segmentation.

If it is the case that the beginning learners had started to create segmentation hypotheses for units larger than monosyllables, how can their preference for grouping PVP be accounted for? Two explanations are possible. First, /PVP/ is overrepresented in Mandarin Chinese, if probabilities are calculated over word tokens. Hence, it might be that probabilistic phonotactic cues for speech segmentation are not based on calculations over phonotactic patterns in word types but word tokens. Second, it might be that participants have a preference for words that group identical consonants, since sequences of identical consonants are overrepresented in Mandarin Chinese. This preference may have triggered an overall PVPVTV segmentation.

The results of Experiment 3 suggest that the beginning learners with residence in the Netherlands were beginning to acquire some L2 speech segmentation strategies. However, the experiment does not offer any evidence that daily exposure to a language is sufficient for acquiring a specific phonotactic cue for speech segmentation. For acquiring knowledge that OCP-PLACE can be used as a cue for detecting word boundaries in the signal, some degree of general L2 proficiency is also required.

4.6. General discussion

In the current study, we dealt with the question whether OCP-effects across languages can be accounted for by channel bias, i.e., functional biases on perception and production. The issue was approached in three experiments testing whether OCP-PLACE would serve as a cue for speech segmentation for native listeners of Mandarin Chinese, a language that is not restricted by this typologically well-attested phonotactic constraint. If OCP-PLACE were due to channel bias, then monolingual listeners of Mandarin Chinese should nevertheless use the segmentation cue. On the contrary, we hypothesized that native speakers' knowledge of OCP-PLACE is acquired from the input, possibly reflecting a cognitive bias on acquisition. As acquisition interfaces with the input a learner receives, language-specific effects are to be expected. Hence, we assumed that only Mandarin Chinese listeners who had acquired an L2 that displays effects of OCP-PLACE would use the cue. An additional hypothesis was that speech segmentation cues were acquired from experience with segmenting speech. Hence, we hypothesized that only L2 learners who had had extensive exposure to an L2 that displayed OCP-PLACE effects would use the constraint for speech segmentation.

4.6.1. OCP-PLACE is not a result of channel bias

Previous literature has found effects of OCP-PLACE across geographically and genetically unrelated languages. For that reason, OCP-PLACE has been proposed to be a linguistic universal (Pozdniakov & Segerer, 2007). The first finding of the current study is that there is at least one language, namely Mandarin Chinese that does not constrain its lexicon by gradient OCP-PLACE. A calculation of consonant co-occurrence distributions in Mandarin Chinese yielded the results that co-occurrences of consonants with the shared feature [place] are not restricted. These results makes sense in light of previous studies which had argued that the effect of OCP-PLACE is morpheme-based (e.g., McCarthy, 1986). In Mandarin Chinese, generally, each syllable forms a morpheme (Duanmu, 2000). Besides this, Mandarin Chinese has a very simple syllable structure. Consonant clusters are absent, and the only possible syllable codas are nasals [n] or [ŋ]. Hence, in Mandarin Chinese, not many morphemes consist of two consonants, for which reason OCP-PLACE may never take an effect.

Still, if the typological frequency of OCP-PLACE were due to a channel bias, then it could nonetheless be predicted that OCP-PLACE would universally affect processing when segmenting an artificial language that contains no other cues for segmentation than OCP-PLACE. However, although previous studies have shown that native Dutch listeners can use OCP-PLACE as a cue for online speech segmentation (Boll-Avetisyan & Kager, submitted; Kager & Shatzman, under revision), the current study finds no evidence that Mandarin Chinese listeners use this cue. Hence, the assumption that processing effects of OCP-PLACE are directly due to channel bias is not supported by our study.

The fact that monolingual speakers of Mandarin Chinese were not biased to use OCP-PLACE for speech segmentation invites a comparison with studies that found the perception of L1 unattested consonant cluster to be influenced by typological tendencies of markedness (Berent, et al., 2006). The crucial difference between these studies and our study may be as follows. In the studies by Moreton or Berent and colleagues, listeners were tested on consonant clusters (e.g., /#bn/, /#bd/ and /#lb/ in Berent, et al., 2006; and /#bw/ or /#tl/ in Moreton, 2002) that they would never encounter in their native language. Possibly, the influence from an innate bias (may it be channel bias or analytic bias) has not died out here, because the clusters are heavily under-represented within words and continuous speech. From this it follows that in the case of perception of L1 unattested clusters, there is no need from the input language to create pressure against the markedness constraints. In contrast, the monolingual Mandarin Chinese listeners in our study do encounter sequences of /PvP/ within words, and hence, also in speech—just as often as they would be expected. Thus, consonant distributions in Mandarin Chinese may give counter-pressure to OCP-PLACE. For that reason, it would be counter-productive if Chinese listeners

allowed for an innate bias to insert word boundaries between two /P/. So, possibly the innate bias is overwritten by the distributions.

Then again, in the case of our study, the monolingual Mandarin listeners did not receive any cue for speech segmentation. In terms of distributional information, it was completely flat. Hence, if OCP-PLACE were due to a general channel bias on perception, then it should have affected the perception of the monolingual listeners in this task. Since this has not been the case, we conclude that OCP-PLACE is acquired from the input, and will only affect perception, if it has been acquired.

4.6.2. OCP-PLACE as a result of analytic bias?

To examine our hypothesis that OCP-PLACE effects on processing arise from an acquisition of input distributions, we tested whether Mandarin Chinese L2 learners of Dutch would use OCP-PLACE for segmenting an artificial language. The hypothesis was confirmed to the extent that Mandarin Chinese participants who were highly proficient L2 learners of Dutch with large exposure to Dutch input spoken by native speakers used OCP-PLACE as a cue to word boundaries in artificial language segmentation. Prior to our study, only Weber and Cutler (2006) have shown that L2 learners can use phonotactics of the target language for speech segmentation. However, their study only tested the use of knowledge of specific L2 biphones. In the current study, we found that a phonotactic constraint of a certain degree of complexity can be acquired in an L2, and that this L2 knowledge influences speech segmentation preferences. L2 learners used a typologically well-attested markedness constraint that affects pairs of non-adjacent phonemes, and, moreover, refers to consonants that are part of a natural class.

A question that remains open at this point is whether L2 learners' acquisition process was guided by an innate analytic bias. We cannot offer direct evidence for this assumption. However, OCP-PLACE is a typologically well-attested constraint, and usually the existence of cross-linguistically recurrent patterns is attributed to either channel bias or analytic bias. As our study provides evidence against a channel bias on effects of OCP-PLACE on perception, its typological frequency may speak for an effect of analytic bias on acquisition. Moreover, the fact that it constrains non-adjacent consonants, which requires the learner to have certain abstraction skills, and the fact that it refers to consonants that are part of the natural class of labials may speak against the view that knowledge of OCP-PLACE is acquired purely by distributional learning.

We would like to speculate what could be the triggering cognitive force that aids the acquisition of OCP-PLACE. In the introduction, we discussed that non-adjacent dependencies have been claimed to be learned only when they involve similarity between the non-adjacent units (e.g., Newport & Aslin, 2004; Onnis,

et al., 2005; Perruchet & Pacton, 2006). How can OCP-PLACE, which formally requires the opposite—namely similarity avoidance between non-adjacent consonants—be acquired if a pre-requisite for non-adjacent dependency learning is similarity? Boll-Avetisyan et al. (submitted) demonstrated that an underrepresentation but not an overrepresentation of two identical non-adjacent consonants can be acquired as a segmentation cue by infants. It will be interesting if future studies investigate whether the learner when acquiring OCP-PLACE builds up knowledge about under-represented but similar consonants, or about the over-representation of dissimilar consonants within words.

Interestingly, the highly proficient L2 learners who had learned Dutch in the Netherlands have transferred L2 knowledge of OCP-PLACE when processing an artificial language, but not when processing their Dutch L2. The artificial language was synthesized using a Turkish voice, and hence was neither similar to their L1 nor to their L2. Thus, the artificial language did not give any cues to the L2 learners that they should have applied Dutch L2 phonotactic segmentation strategies. What exactly can one conclude from the discovery that participants had used OCP-PLACE for segmenting an independent L3? On the one hand, the learners might have been aware of the fact that they were recruited for participation in the experiment because of their knowledge of Dutch, and hence activated their L2 Dutch phonology. Then the results would be due to language mode (Grosjean, 2001). On the other hand, if there is an analytic bias on OCP-PLACE, then it is not surprising that listeners prefer to use it for segmenting a "neutral" artificial language once it has been acquired, either in an L1 or an L2.

4.6.3. The role of exposure

The results revealed that only proficient L2 learners who had acquired Dutch in the Netherlands—and not non-proficient learners or proficient learners who had acquired Dutch in China—use the constraint in processing. Hence, its acquisition seems to be connected to specific learning conditions. Studies up to now had mostly indicated that L2 phonotactic knowledge was correlated with general L2 proficiency (e.g., Boll-Avetisyan, 2011; Trapman & Kager, 2009). Others (e.g., Altenberg, 2005; Flege & Liu, 2001) found that exposure to native spoken input (e.g., defined by LoR) is a factor that accounts for an influence of L2 phonotactic knowledge on L2 perception. Our study reveals that essentially it is large-scale exposure to the L2 spoken by native speakers that drives L2 phonotactic learning.

4.7. Conclusion

To sum up, our study indicates that knowledge of OCP-PLACE does not reflect functional biases on perception. If it is not channel bias that causes OCP effects in perception, then the constraint OCP-PLACE must be learned. Given the fact that OCP-PLACE reflects a cross-linguistically well-attested pattern, it is likely that there is support from analytic bias on its acquisition. OCP-PLACE influences speech processing in listeners who have acquired an L1 or L2 that gives reason for the assumption of the relevance of that constraint. The pre-requisite for the use of OCP-PLACE in an L2 is two-fold: the learner has to be proficient and has to have been exposed to a sufficient amount of input spoken by native speakers.

4.8. References

Altenberg, E. P. (2005). The Judgment, Perception, and production of consonant clusters in a second language. *International Review of Applied Linguistics in Language Teaching, 43*(1), 53-80.

Archangeli, D., & Pulleyblank, D. (2007). Harmony. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 353-378). Cambridge, England: Cambridge University Press.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philedelphia: Linguistic Data Consortium, University of Pennsylvania.

Baer-Henney, D., & Vijver van de, R. (submitted). Substance and Amount of Exposure in the Acquisition of Alternations.

Berent, I., Everett, D. L., & Shimron, J. (2001). Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology, 42*, 1-60.

Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition, 64*, 39-72.

Berent, I., Steriade, D., Lennertz, T., & Vakninc, V. (2006). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104*(3), 591-630.

Berkley, D. (1994). The OCP and Gradient Data. *Studies in the Linguistic Sciences, 24*, 59-72.

Boersma, P. (1998). *The OCP in functional phonology*.Unpublished manuscript.

Boersma, P. (2000). The OCP in the perception grammar. *Rutgers Optimality Archive, 435*, 1-52.

Boll-Avetisyan, N. A. T. (2011). Second Language Probabilistic Phonotactics and Structural Knowledge in Short-Term Memory Recognition. In N. Danis, K. Mesh & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (Vol. 1, pp. 60-72). Somerville, MA: Cascadilla Press.

Boll-Avetisyan, N. A. T., & Kager, R. W. J. (submitted). OCP-PLACE in speech segmentation.

Boll-Avetisyan, N. A. T., Kager, R. W. J., de Bree, E., Kerkhoff, A., van den Bergh, H., & den Boer, S. (submitted). Does the lexicon bootstrap phonotactics, or vice versa? .

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations. The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science, 16*, 451-459.

Coetzee, A. (2005). The OCP in the perception of English. In S. Frota, M. Vigario & F. M. J. (Eds.), *Prosodies* (pp. 223-245). New York: Mouton de Gruyter.

Coetzee, A. (2008). Grammar is both categorical and gradient. In S. Parker (Ed.), *Phonological Argumentation*. London: Equinox.

Coetzee, A., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory, 26*, 289-337.

Dong, X. (2008). *Hidden Phonotactic Knowledge of L2 Listeners.* Utrecht University, Utrecht.

Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). *The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes.* Paper presented at the Proceedings of The Fourth International Conference on Spoken Language Processing (ICSLP'96) 3.

Eckes, T. (2010). Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In R. Grotjahn (Ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-test: Contributions from current research* (pp. 125-192). Frankfurt am Main: Lang.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a Second language. *Studies in Second Language Acquisition, 23*, 527-552.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. M. Kirchner & D. Steriade (Eds.), *Phonetically based Phonology* (pp. 346-371). Cambridge: Cambridge University Press.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory 22*, 179-228.

Frisch, S. A., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language Acquisition, 77*, 91-106.

Greenberg, J. (1950). The patterning of root morphemes in Semitic. *Word, 5*, 162-181.

Grosjean, F. (2001). The bilingual's language modes. In P. Hallé, J. Segui, U. Frauenfelder & J. L. Nicol (Eds.), *One Mind, Two Languages* (pp. 1-22). Oxford: Blackwell.

Hay, J. (2003). *Causes and Consequences of Word Structure*. New York and London: Routledge.

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Journal of Laboratory Phonology, 1*, 207-224.

Kager, R., & Shatzman, K. (2007). Phonological constraints in speech processing. In B. Los & M. v. Koppen (Eds.), *Linguistics in the Netherlands 2007* (pp. 100-111). Amsterdam: John Benjamins.

Kager, R., & Shatzman, K. (under revision). *Similarity Avoidance in Speech Processing*.Unpublished manuscript.

Lentz, T. O., & Kager, R. W. J. (in prep.). *Conflicting phonotactics in segmentation*.Unpublished manuscript.

Lentz, T. O., & Kager, R. W. J. (submitted). *Categorical and Gradient Phonotactics Are Represented Differently: Evidence from Word Recognition*.Unpublished manuscript.

Lentz, T. O., & Kager, R. W. J. (under revision). Phonotactic cues initiate lexical look-up: how both illegality and wellformedness infuence speech segmentation.

Li, A., & Xu, B. (2001). *A spontaneous conversation corpus CADCC*. Korea: Oriental COCOCSDA.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107*, 395-419.

McCarthy, J. J. (1985). *Formal Problems in Semitic Phonology and Morphology*. New York: Garland.

McCarthy, J. J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry, 17*, 207-263.

Moreton, E. (2002). Structural constraints in the perception of English stopsonorant clusters. *Cognition, 84*, 55-71.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology 25*(1), 83-127.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology 48*, 127-162.

Ohala, J. (1994). *Towards a universal, phonetically-based theory of vowel harmony.* Paper presented at the Proceedings of the International Conference on Spoken Language Processing (ICSLP).

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*, 225-237.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604-607.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233-238.

Pierrehumbert, J. B. (1993). *Dissimilarity in the Arabic verbal roots.* Paper presented at the Proceedings of the North East Linguistics Society.

Pierrehumbert, J. B. (2001). Stochastic phonology. *Glot International, 5*(6), 195-207.

Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition, 116*, 361-367.

Pozdniakov, K., & Segerer, G. (2007). Similar Place Avoidance: A statistical universal. *Linguistic Typology, 11*, 307-348.

Rose, S., & Walker, R. (2004). A Typology of Consonant Agreement as Correspondence. *Language, 80*, 475-531.

Shatzman, K., & Kager, R. (2007). A role for phonotactic constraints in speech perception. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1409-1412). Dudweiler: Pirrot.

Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.

Thorpe, L. A., & Trehub, S. E. (1989). Duration Illusion and Auditory Grouping in Infancy. *Developmental Psychology, 25*(1), 122-127.

Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition, 16*, 178-221.

Van Kampen, A., Parmaksiz, G., van de Vijver, R., & Höhle, B. (2008). Metrical and statistical cues for word segmentation: Vowel harmony and word stress as cues to word boundaries by 6- and 9-month old Turkish learners. In A. Gavarró & M. J. Freitas (Eds.), *Language Acquisition and Development* (pp. 313-324). Newcastle: Cambridge Scholars Publishing.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133-149.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America, 119*, 597-607.

Appendix A

Inventory of the syllables (all tone 1) of the artificial language and their biphone frequency within CADCC word types and word tokens as well as their cohort density (CD).

Syllable type	Specific syllable	Biphone frequency (Word types)	Biphone frequency (Word tokens)	CDs
Pv_1	pe	0 (0.00)	0 (0.00)	0.00 (0.00)
	bu	0 (0.00)	0 (0.00)	0.00 (0.00)
	mi	0 (0.00)	0 (0.00)	0.00 (0.00)
Tv	te	1 (0.70)	1 (0.01)	0.00 (0.00)
	de	1 (0.70)	2 (0.02)	0.30 (0.32)
	na	0 (0.00)	0 (0.00)	0.00 (0.00)
Pv_2	po	3 (2.10)	6 (0.05)	0.48 (0.50)
	bi	2 (1.40)	22 (0.18)	1.76 (1.85)
	mu	0 (0.00)	0 (0.00)	0.00 (0.00)

CHAPTER 5

PROBABILISTIC PHONOTACTICS IN LEXICAL ACQUISITION: THE ROLE OF SYLLABLE COMPLEXITY[1]

5.1. Introduction

Numerous psycholinguistic studies provide experimental evidence that human knowledge of phonotactics is gradient. For example, in well-formedness judgment tasks, participants rate nonwords to be more word-like when they are composed of frequently occurring phoneme sequences (e.g., Bailey & Hahn, 2001; Coleman & Pierrehumbert, 1997; Scholes, 1966). Furthermore, phonotactic distributions affect human performance in online processing tasks such as lexical decision, nonword repetition, and same-different tasks, such that nonwords composed of high frequency biphones are processed faster and more accurately than nonwords composed of low frequency biphones (e.g., Frisch, Large, & Pisoni, 2000; Pitt & McQueen, 1998; Vitevitch & Luce, 1998, 1999). Phonologists assume that phonotactic well-formedness is assessed by constraints that are part of a phonological grammar. In the traditional generative view, the phonological grammar is assumed to be strictly categorical, which means that phonotactic constraints determine whether a nonword is licit or illicit (e.g., Chomsky & Halle, 1968). Recent models of phonological grammar include gradient constraints on phonotactic well-formedness or gradience as a result of stochastic constraint interaction (e.g., Boersma & Hayes, 2001; A. Coetzee & Pater, 2008; Frisch, Pierrehumbert, & Broe, 2004; Hayes, 2000; Pierrehumbert, 2003), which can also account for probabilistic phonotactic effects on the lexicon.

---

[1] This study is published as Boll-Avetisyan, Natalie (in press): Probabilistic phonotactics in lexical acquisition: The role of syllable complexity. In Hoole, P., Bombien, L., Pouplier, M., Mooshammer, C., & Kühnert, B. (Eds): *Consonants and structural complexity*. Mouton de Gruyter, Interface Explorations Series. It was funded by an NWO grant (277-70-001) awarded to René Kager.

The standard view among phonologists is that phonotactic constraints, which define the possible set of words in a language, emerge from generalizations over items stored in the lexicon (Hayes & Wilson, 2008; Pierrehumbert, 2003). Such *unidirectional* models of the relation between the phonotactic grammar and the lexicon, however, predict that the language-learning child needs to establish a lexicon of considerable size before being able to deduce phonotactics. This is implausible, given that nine-month-old infants already show evidence of language-specific phonotactic knowledge, for example, by preferring frequent phonotactic patterns in their language to infrequent ones (Jusczyk, Luce, & Charles-Luce, 1994), and by using language-specific phonotactics for speech segmentation (Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999), at an age at which the lexicon is not likely to have been fully developed (e.g., Fenson et al., 1994). Such experimental studies are consistent with the hypothesis that the lexicon is not the only source of phonotactic knowledge, at least not in early acquisition. Still, they fail to provide conclusive evidence, as it is unknown when lexical entries initially are created and what the properties of entries in a proto-lexicon might be.

Over the past decade, studies have been carried out that demonstrate the effects of probabilistic phonotactics on lexical acquisition. In light of these studies it becomes more likely that the relation between phonotactics and the lexicon is *bidirectional.* For example, when children are taught unfamiliar word/meaning pairs, they learn CVC nonwords with high phonotactic probability biphones more rapidly than CVC nonwords with low phonotactic probability biphones (e.g., Storkel, 2001; Storkel & Rogers, 2000). Similar evidence for the role of phonotactics in lexical acquisition comes from nonword recall experiments with children and adult second language learners. Short-term memory (STM) recall performance in nonword recall tasks has often been found to be correlated with lexicon size in adults and children in their native language (L1) (for an overview see Gathercole, 2006) and foreign language (L2) (e.g., Service & Kohonen, 1995), and has, hence, been suggested to reflect the role of STM in lexical acquisition. In fact, nonword recall can be seen as the initial step in storing new words in the mental lexicon. The easier it is to hold an item in STM, the easier it is to store it in long-term memory (LTM). Interestingly, STM recall is affected by LTM representations. Hence, it has been suggested that LTM knowledge is used for reconstructing degraded memory traces during sub-vocal rehearsal in STM—a process referred to as *redintegration* (Schweickert, 1993).

As to the role of phonotactics in STM recall, a pioneer study by Gathercole and colleagues (Gathercole, Frankish, Pickering, & Peaker, 1999) showed that seven to eight-year-old children were better at recalling CVC nonwords with high rather than low phonotactic probability in serial nonword repetition tasks. Moreover, children who did particularly well at relying on cues from phonotactic probability in nonword repetition were shown to have a larger lexicon size than children who did less well in relying on the phonotactic cues.

Similar results have been found in a study with adult L2 learners (Majerus, Poncelet, Van der Linden, & Weekes, 2008).

Although the studies reviewed above may seem to offer evidence for direct effects of probabilistic phonotactics on lexical acquisition, it is important to guard against premature conclusions, given that there are two levels of processing from which probabilistic phonotactic information can be derived. One is a *sub-lexical level*, at which knowledge of sub-word units, such as phonemes and biphones and their probability of occurrence (e.g., Vitevitch & Luce, 1999) is represented. The other is the *lexical level*, at which the phonological forms of words and morphemes are represented. Probabilistic phonotactics can be deduced from lexical items by comparing phonologically similar items and their frequencies.

As mentioned before, it would be helpful for a language-learning child or an L2 learner, if (s)he could draw on phonotactic knowledge to facilitate word learning before the onset of lexical acquisition. To open up this possibility, it is necessary to distinguish between lexical and sub-lexical knowledge, as only the latter can possibly be acquired independently of the lexicon. The study by Gathercole et al. (1999) was later criticized on the grounds that when manipulating sub-lexical factors (such as biphone frequencies), lexical factors (such as lexical neighborhood density) had not been controlled for (Roodenrys & Hinton, 2002). Lexical and sub-lexical probabilities are highly correlated: words composed of high frequency biphones tend to have many lexical neighbors (e.g., Landauer & Streeter, 1973). Experimental evidence suggests that both lexical and sub-lexical factors function as independent predictors of well-formedness judgments on nonwords, even though they are highly correlated (e.g., Bailey & Hahn, 2001). Furthermore, they are known to have different effects on online word recognition. In a series of experiments using different online-processing tasks, Vitevitch and Luce (1999) found that CVC nonwords with a high phonotactic probability and high lexical neighborhood density were processed faster than CVC nonwords with low phonotactic probability and low lexical neighborhood density. For real words, the effect was reversed. The authors interpreted their findings as lexical frequency effects being dominant in word processing, with large competing cohorts slowing down processing. The processing of nonwords, lacking a lexical entry, might predominantly happen at the sub-lexical level and receive facilitation from high phonotactic probability, as frequent biphones can be decoded more quickly.

Studies following up on Gathercole et al. (1999) that have controlled for lexical factors do not give a coherent picture of whether only lexical factors or also sub-lexical factors play a role in STM nonword recall. Roodenrys and Hinton (2002) ran two experiments in order to establish the extent to which sub-lexical knowledge (probabilistic phonotactics) or lexical knowledge (lexical neighborhood density) affects STM nonword recall. In their experiments with adults, biphone frequencies did not affect recall accuracy when lexical

neighborhood density was controlled for, while lexical neighborhood density was a significant factor when biphone frequency was controlled for. They conclude that recall of CVC nonwords is primarily influenced by lexical neighborhood density, more than by frequency of the CV and VC biphones of which they are composed. However, using a more stringent control for lexical neighborhood, Thorn and Frankish (2005) found a facilitatory effect of high biphone frequency on recall accuracy.

When Storkel and colleagues separately investigated the influence of lexical and sub-lexical factors on word learning in adults (Storkel, Armbrüster, & Hogan, 2006), they found facilitatory effects of high neighborhood density on word learning, but—contrary to their previous studies—an *inhibitory* effect of high phonotactic probability. The effect of phonotactic probability, however, only occurred among items that were partially remembered, i.e., among items in which only two out of three phonemes were correctly reproduced. The authors conclude that novel words with low phonotactic probability might stand out in speech and therefore be relevant in the initial stages of word learning. Later, during the integration of novel forms into the lexicon, lexical neighborhood information might be of higher importance. In sum, these contradictory findings concerning a sub-lexical involvement in lexical acquisition are far from conclusive.

It is remarkable that the aforementioned studies have only addressed probabilistic phonotactic knowledge, which merely refers to the frequency or probability of co-occurrences of phonemes in a language. A number of researchers (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Pierrehumbert, 2003) have equated the sub-lexical level, as it is phonological by definition, with a phonological level that represents a phonological grammar. Under current conceptualizations, phonological grammar can be interpreted as a knowledge system that contains, among others, markedness constraints referring to abstractions of structure, such as syllable structure (Prince & Smolensky, 1993/2004). Markedness constraints ban marked structures, such as complex syllable onsets and codas. The notion of a phonological grammar is supported by data from typology, acquisition and processing. Typologically, languages that tolerate complex syllable margins, such as Dutch and English, also tolerate simple syllable margins; in contrast, however, there are no languages that tolerate complex syllable margins, but disallow simple syllable margins. This implication indicates that complex constituents are more marked than simple constituents. Complex constituents are restricted by the constraints *Complex-Onset, penalizing consonant clusters in syllable onsets, and *Complex-Coda, penalizing consonant clusters in syllable codas. CVC syllables, however, are relatively unmarked, being constructed of simple syllable constituents only.

It has been noted that unmarked structures are acquired earlier than marked structures (e.g., Jakobson, 1941; Smolensky, 1996). This has also been found to be the case with syllable structure. Dutch children start producing CV and CVC

syllables before CVCC and CCVC syllables. CCVCC seems to be even harder to acquire, possibly due to a conjoined constraint *Complex-Onset&*Complex-Coda (Levelt & Vijver, 1998/2004).[2] This means that words with a CCVCC structure are intrinsically more marked than words with a CVC structure, even if complex constituents are phonotactically legal, as in languages such as English or Dutch. When learners are faced with the task of learning a word, they need to make reference to their phonological grammar, which informs them about whether the form violates or satisfies markedness constraints. When a word of a simple CVC structure is acquired, markedness constraints will hardly be violated. The more complex a structure, the more relevant markedness constraints will be for the processing system. The reference to the phonological grammar when acquiring a word should furthermore have an effect on the influence of probabilistic phonotactics.

The aforementioned studies on the role of sub-lexical representations in lexical acquisition have not only neglected the fact that sub-lexical processing is informed by two different components; they have also used exclusively structurally simple test items, in particular CVC syllables. CV and VC biphones are subject to weaker phonotactic restrictions as compared to consonant clusters in syllable margins, which, being part of single syllable constituents (onsets and codas), are subject to stronger co-occurrence constraints (Selkirk, 1982). Therefore, my interpretation of the previous studies is that by only using CVC items as stimuli, these studies only had a slight chance of being able to demonstrate sub-lexical effects on STM recall or lexical acquisition.

The low probability of consonantal biphones often coincides with the fact that they violate more markedness constraints. For example, the Dutch cluster /kn/ is intrinsically less well-formed than /kr/ due to the universal Sonority Sequencing Principle, which states that consonants in clusters should differ maximally in sonority (Broselow, 1984; Selkirk, 1982). Hence, for words with consonant clusters, markedness constraints and biphone probabilities typically converge. In the case of learning a CVC word, however, the influence of structural constraints is at a minimum, since markedness constraints expressing co-occurrence restrictions between C and V, or V and C are not typologically common. Hence, it should not make much of a difference whether a CVC word is composed of high or low frequency biphones. In the case of structurally more complex items, the influence of structural constraints increases. Structurally

[2] The order of acquisition matches the frequency of occurrence of the templates in Dutch. Hence, it might be that frequency of occurrence dominates markedness in acquisition, and the order of acquisition might be different in other languages. This, however, is not likely, as markedness dominates frequency in other cases such as the acquisition of sC-clusters in Dutch, in which unmarked structures are acquired first, although they are less frequent than the marked (see Fikkert, P. (2007). Acquiring phonology. In P. d. Lacy (Ed.), *Handbook of phonological theory* (pp. 537-554). Cambridge, MA: Cambridge University Press.).

complex and low biphone frequency nonwords should be the most difficult to remember, as they violate the most markedness constraints.

There are some studies that provide evidence for the role of syllable structure in STM recall. Treiman and Danis (1988), for example, let participants repeat a series of nonwords with CVC, CCV, or VCC structures. A count of the erroneous repetitions revealed fewer errors in biphones that are part of a syllable constituent than in biphones that straddle a constituent boundary. Also, 4-year-olds make fewer errors when repeating nonwords with singleton consonants than with consonant clusters (Gathercole & Baddeley, 1989). These studies, however, did not control for probabilistic phonotactics, so it might be the case that fewer errors occurred in biphones of higher frequency—a factor that was unknown to influence recall at the time the studies were carried out.

The hypothesis of the current paper is that sub-lexical representations are two-fold, involving representations of "non-grammatical" low-level probabilistic phonotactics on the one hand and a phonological grammar with markedness constraints on the other. We hypothesize that effects of phonotactic probability are modulated by markedness constraints.

Lexical acquisition, according to this hypothesis, should be influenced by an interaction of the two sub-lexical components, such that facilitatory effects from high phonotactic probability on word learning should be aggravated by structural complexity. More precisely, the difference in learning difficulty between words containing biphones with low versus high phonotactic probability should be larger for words containing structurally complex syllable constituents (such as complex onsets or complex codas) than for nonwords containing structurally simple syllable constituents (such as singleton onsets and codas).

The hypothesis was tested in two STM experiments with adult native speakers of Dutch. Dutch allows for complex syllable onsets and syllable codas. Yet the markedness interpretation of structural complexity predicts that complex onsets and complex codas should be less well-formed than simple onsets and codas. I used a probed STM recognition task (Sternberg, 1966), which has the advantage that no articulatory motor programs are co-activated. This is a change from previous studies on probabilistic phonotactics in lexical acquisition, which mostly used production tasks (e.g., Gathercole, et al., 1999; Roodenrys & Hinton, 2002; Storkel, et al., 2006; Thorn & Frankish, 2005). A study (Storkel, 2001) that used perception-based tasks revealed facilitatory effects of high phonotactic probability to the same extent as a production-oriented task.

The prediction was that with nonword stimuli doubly manipulated for both phonotactic probability and syllabic complexity, phonotactic probability would affect recognition performance such that high biphone frequency would facilitate nonword recognition, but only or more severely so, in the case of complex syllables.

5.2. Experiment 1

5.2.1. Method

Participants

Participants were 30 native speakers of Dutch without self-reported hearing disorders. All were drawn from the Utrecht Institute of Linguistics' participant pool and compensated for participation.

Materials

All stimuli were nonwords that are phonotactically legal in Dutch. That is, all phonemes are part of the Dutch phoneme inventory and all biphones are licit sequences in the syllable positions in which they occur. The stimuli were manipulated for two factors. One was syllable structure type, a factor of four levels (CVC, CVCC, CCVC, and CCVCC). The second was biphone frequency, a factor of two levels (high versus low biphone frequency). Biphone frequency counts were based on biphone occurrences in a lexicon of 8,305 monomorphemic word types extracted from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). The low biphone frequency stimuli consisted of biphones taken from the least frequent 33% of all biphones, and the high biphone frequency stimuli were composed of biphones taken from the most frequent 33% of all biphones. Mean biphone frequencies are given in Table 1.

The stimuli were controlled for lexical neighborhood density at all levels of both factors in order to rule out lexical effects. A lexical neighbor was defined as any word differing from the nonword by the addition, deletion, or substitution of a single phoneme (e.g., Vitevitch & Luce, 1999). Lexical neighborhood density was calculated by summing the logged token frequencies of lexical neighbors for each target item.

Biphone frequency	High	Low
#CC	5252.72 (4986.55), N=46	580.17 (309.65), N=46
CV	6190.61 (3960.58), N=92	632.47 (444.91), N=92
VC	9132.51 (6510.22), N=92	624.74 (575.24), N=92
CC#	8711.96 (6777.85), N=46	1096.13 (560.07), N=46

Table 1: Averages of token biphone frequencies (#CC= consonant cluster in onset, CC# = consonant cluster in coda position) of the stimuli used in Experiments 1 and 2. Numbers in parentheses represent standard deviations.

The target items were created such that they would only minimally differ from each other, such as the low biphone probability nonwords /lum/, /lump/, and /xlump/, or the high biphone probability nonwords /vo:k/ and /vo:kt/. In this way, interference of singleton frequency effects, which are known to influence phonotactic well-formedness (e.g., Bailey & Hahn, 2001), was minimized.

The CVC, CCVC, CVCC and CCVCC filler items used in this experiment were randomly selected from a list of Dutch nonwords. Each filler item occurred only once throughout the experiment. The stimuli were spoken in a sound-proof booth by a Dutch native speaker, who was naïve of the purpose of the study. Two Dutch native speakers confirmed that the stimuli sounded natural. A list of all target items is given in Appendix A.

Procedure

Participants were tested in a probed recognition task (Sternberg, 1966), in which they were presented a series of four nonwords followed by a probe. The task was to decide whether the probe was in the series or not. Each series contained one target and three filler items. The series were designed such that every syllable type occurred once in each trial. Every series thus contained the same number of segments. An example is given in Table 2.

Series	Item 1	Item 2	Item 3	Item 4	Probe	Appeared before?
Target	/do:rk/	/tɛx/	/zwa:lm/	/gɔp/	/tɛx/	yes
Filler	/brint/	/tɛp/	/bi:rf/	/frɑp/	/lyn/	no

Table 2: Examples of both a target and a filler series used in Experiment 1 and 2a.

The experiment consisted of 184 target series. The design had two factors (biphone frequency, syllable structure) with 2x4 levels (high/low; CVC, CVCC, CCVC, CCVCC); accordingly, the 184 targets divide into 23 targets of each type. In addition, 184 filler-series, in which the probe did not match any of the prior four fillers, were included. All series and all items within the series were randomized for every participant in order to avoid item- or series-specific order effects.

Participants were tested individually in a soundproof booth facing a computer screen. The stimuli were played by computer over headphones at a comfortable volume. The stimuli were presented with a relatively long inter-stimulus-interval (ISI) of 700 ms, and the probe was presented 1400 ms after the last item of the stimulus series. This was done to add a high memory load to the task to invoke

reference to more abstract phonological representations (e.g., Werker & Logan, 1985).

Yes/No-decisions were made on a button-box. After the end of each series, an icon occurred on the screen indicating the beginning of a new trial. The dependent measure was reaction time (RT). When no decision was made after 3000 ms, the trial was stopped and counted as an error. The experiment took 90 minutes. Three breaks were included. There were three practice trials before the experiment started.

Data selection

All 30 participants were included in the analysis. Fillers, errors, no-responses, responses slower than 2500 ms, and responses faster than 200 ms were excluded from the analysis. The RT was measured from the onset of the target probe.

5.2.2. Results and discussion

Results

A linear mixed regression model with RT as dependent variable, Participants and Targets as random factors, and Biphone frequency (high/low), Syllable structure (CVC, CCVC, CVCC, CCVCC) and Biphone frequency*Syllable structure as fixed factors revealed significant differences between nonwords of different syllable structures as well as interaction effects between syllable structure and biphone frequency, but no significant differences between high and low biphone frequency nonwords (see Appendix B).

Syllable Structure	**High**	**Low**	**Δ**	**Total *M***
CVC	1117.82	1103.75	-14.70	1110.79
CVCC	1114.42	1172.82	58.40	1143.62
CCVC	1107.68	1182.36	74.68	1145.02
CCVCC	1133.08	1221.40	88.32	1177.24
Total *M*	1118.25	1170.08	51.68	1144.17

Table 3: Estimated reaction time averages and the differences (Δ) between reaction time means for high versus low biphone frequency in Experiment 1 in ms measured from the target onset.

An analysis of the estimates of the fixed effects with "low biphone frequency-CVC" as a reference point revealed a significant interaction of biphone frequency and syllable structures between the simple syllable structure CVC and the more complex syllable structures CCVC and CCVCC (see Appendix C). That is, compared to high and low biphone frequency CVC nonwords, participants were significantly slower at recalling CCVC and CCVCC nonwords of low biphone frequency than CCVC and CCVCC nonwords of high biphone frequency (see Figure 1). The averages of the estimated RT means are given in Table 3.

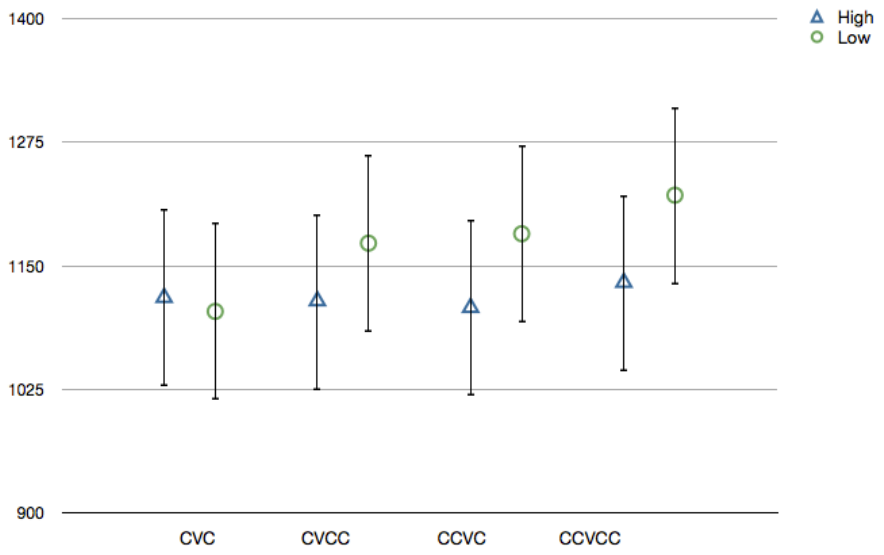

Figure 1: Averages of estimated means and 95% confidence intervals of RT in ms for each syllable structure (CVC, CVCC, CCVC and CCVCC) for each high (H) and low (L) biphone frequency nonwords.

Exploration

The results in Experiment 1 may, however, be due to another factor, which we had not controlled for: high biphone frequency words are generally spoken faster than low biphone frequency words (Kuperman, Ernestus, & Baayen, 2008).[3] In order to check whether such a difference might account for our results, we measured the duration of our target items. We did not find that our speaker had pronounced all high biphone frequency target items in a shorter time than the low biphone frequency target items. However, we observed that

[3] I am indebted to Mirjam Ernestus for pointing this out.

the spoken durations of each item type overall matched the respective reaction times. So, both durations and reaction times were longer for low rather than for high biphone frequency CCVC and CCVCC nonwords. For CVC nonwords, this effect was reversed (compare Figure 1 and Figure 2).

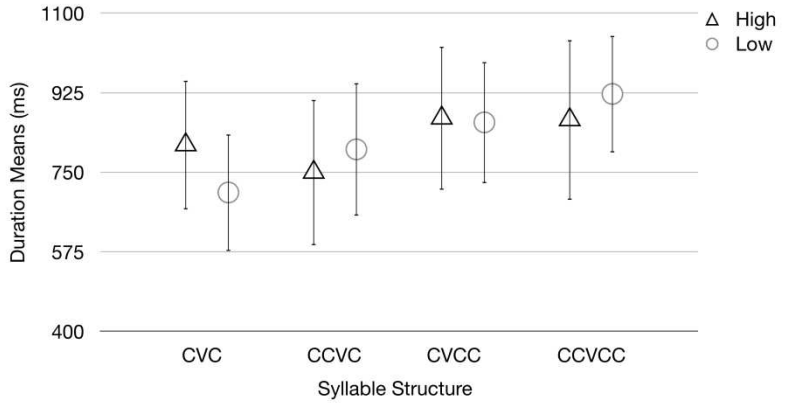

Figure 2: Duration means of the target items and the differences (Δ) between durations of high versus low biphone frequency items in Experiment 1 in ms (*SD* in brackets).

This is problematic, as speech-rate is known to affect recall latencies: the longer the duration of an item, the more difficult it is to recall (e.g., Baddeley, Thomson, & Buchanan, 1975; Cowan, Wood, Nugent, & Treisman, 1997). Hence, the faster RTs on high biphone frequency items may be due to the fact that they were shorter in duration. We added speech-rate as a co-variate to the analysis, and found the effects to remain significant.

Discussion

It was predicted that, when holding CVC items in the phonological loop, there should be little support from sub-lexical LTM representations, as CV and VC biphones are hardly restricted by structural constraints. CC biphones, on the contrary, are much more restricted. Hence, reference to sub-lexical LTM representations with representations of specific biphones making reference to the phonological grammar should be important in preventing complex CVCC, CCVC, or CCVCC items from decaying in the phonological loop. Hence, it was predicted that effects of biphone frequency on STM recognition performance would increase with increasing syllable complexity. This effect should occur

while lexical neighborhood frequency is controlled for, to make sure that effects relate to sub-lexical rather than lexical knowledge.

As displayed in Figure 1, the result is as predicted: The differences in recognition performance between high and low biphone frequency in interaction with syllable structure increased from simple to complex structures. Biphone frequency only affected STM recognition of complex syllables, but not of simple CVC items. It is in line with our prediction that effects of sub-lexical knowledge do not necessarily surface in the recognition performance of CVC items. Prior studies on the effects of sub-lexical knowledge on lexical acquisition—provided they had controlled for lexical factors—also had difficulties finding effects of phonotactic probability on memorizing CVC items. Roodenrys and Hinton (2002) did not find effects of phonotactic probability on recall. Storkel and colleagues (2006) only revealed effects of sub-lexical knowledge on partially, but not correctly memorized nonwords. Still, Thorn and Pickering (2005) found CVC recall to be facilitated by knowledge of phonotactic probability independent of lexical knowledge. Possibly, the lack of a biphone frequency effect on CVC recognition in the current study is due to the long ISIs used in the task. Roodenrys & Hinton (2002) as well as Thorn & Pickering (2005) played their lists at a rate of one item per second. They do not inform us about the mean duration of their items. However, if their duration was similar to ours, then their ISIs were significantly shorter. The longer the ISIs, the more processing of verbal material is influenced by more abstract phonological knowledge (e.g., Werker & Logan, 1985). Thus, when decreasing memory load by using shorter ISIs, low-level probabilistic knowledge might be more involved to support the maintenance of verbal information in the phonological loop.

As mentioned above, the result of our study might be confounded by the speech-rate of the stimuli used in the experiment. Even if the interaction of biphone frequency and syllable structure effect remains significant when speech-rate is added as a co-variate, we need to be careful about drawing conclusions from the results of Experiment 1, since co-variate analysis do not take into account that speech-rate might have different effects at different levels of independent factors.

In order to confirm that the interaction of phonotactic probability and syllable structure is not an artifact of speech-rate, a second experiment controlling for this possible confound was carried out. In Experiment 2, two conditions were tested. The first condition, which will be referred to as Experiment 2a, repeats Experiment 1 while controlling for speech-rate. A second condition, referred to as Experiment 2b, merely contrasted the two extremes CVC and CCVCC. This was done since in order to confirm the central hypothesis of this paper, it suffices to find an interaction between CVC and CCVCC. Considering that Experiment 1 revealed the largest difference in high and low biphone frequency between CVC and CCVCC nonwords, it was predicted that a difference between

high and low phonotactic probability was most likely to occur between CVC and CCVCC nonwords in Experiment 2.

5.3. Experiment 2

5.3.1. Method

Experiment 2 aimed at replicating the results of Experiment 1 while controlling for speech-rate. The experiment was carried out in two conditions: Condition 1 repeated Experiment 1 using exactly the same stimuli, but controlled for speech-rate. Condition 2 only used the CVC and CCVCC nonwords of Condition 1. CVCC and CCVC were excluded because the interaction in Experiment 1 did not occur between CVCC and CVC and was therefore also not expected to occur here. Furthermore, the interaction was strongest between CVC and CCVCC nonwords and our hypothesis can also be tested using two syllable structure types only.

Participants

Sixty native speakers of Dutch without self-reported hearing disorders—all drawn from the Utrecht Institute of Linguistics' participant pool, none of whom had participated in Experiment 1—participated in the experiment. They were compensated for their participation.

Materials

The stimuli were identical to those used in Experiment 1, with two differences: First, the target items were controlled for duration such that for each class of syllable structure the duration of the stimuli did not differ between high and low frequency biphone nonwords (see Figure 3). This was realized by manually adjusting the vowel durations, which produces more natural results than adjusting the duration of the nonwords as a whole, as durational variation in natural speech usually affects vowels more than consonants (e.g., Greenberg, Carvey, Hitchcock, & Chang, 2003). Manipulating vowel length should not have caused perceptual confusion since long and short vowels contrast in terms of quality (F1, F2), making them distinguishable. Finally, it was of utmost importance to maintain the naturalness of the onset and coda constituents, which are the focus of this study. Using the software "Praat" (Boersma & Weenink, 2007), manipulations were carried out on both sides: stimuli with long durations were shortened, and stimuli with short durations were lengthened. To ensure that they would not alter the vowel quality, manipulations were carried

out only in the middle of the vowel. For shortenings, a portion in the middle was cut out, and for lengthenings, the waves of the middle were copy-pasted. Two native speakers of Dutch confirmed that the stimuli sounded natural.

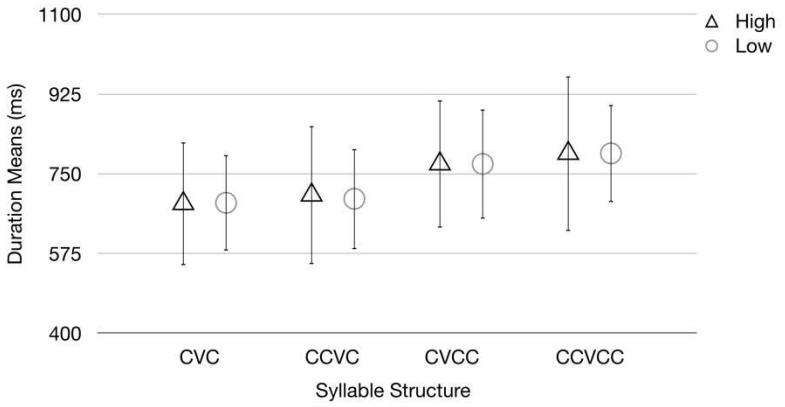

Figure 3: Duration means of the target items in Experiment 2 after controls in ms (*SD* in brackets).

Second, as a probe, we used another recording of the same type by the same speaker to avoid a situation in which participants' decisions are potentially influenced by additional cues from complete acoustic overlap of target and probe.[4]

Procedure

There were two conditions for this experiment. Thirty participants were assigned to Experiment 2a, and thirty participants were assigned to Experiment 2b.

Procedure for Experiment 2a

In Experiment 2a, the ISI was 500 ms, and the probe occurred after an interval of 1000 ms of silence following the series. By doing this, the total duration of the experiment was decreased from 90 to 75 minutes. The shortenings were done in order to increase the chance of revealing effects of low-level phonotactic probability knowledge across all syllable structures. For the rest, the procedure in Experiment 2a was identical to that in Experiment 1.

[4] Thanks to Silke Hamann for this suggestion.

Procedure for Experiment 2b

The procedure was identical to that in Experiment 2a, except for the test series containing CVC and CCVCC. This holds both for the targets stimuli and the fillers (see Table 4). Thus, each series had the same number of segments as in Experiment 2a. There were 92 target series and 92 filler series. This is only half of the target series compared to Experiment 2, from which the stimuli were borrowed, since half of the items (CVCC and CCVC) were excluded. The experiment took 35 minutes.

Series	Item 1	Item 2	Item 3	Item 4	Probe	Appeared before?
Target	/do:rk/	/tɛx/	/zwa:lm/	/gɔp/	/tɛx/	yes
Filler	/brint/	/tɛp/	/xle:ps/	/za:m/	/lyn/	no

Table 4: Examples of both a target and a filler series in Experiment 2b.

Data selection

The data of Experiment 2a and 2b were pooled.[5] All 60 participants were included into the analysis. Fillers, errors, no-responses, and responses that exceeded 2500 ms or fell below 200 ms were excluded. The RT was measured from the onset of the target probe.

5.3.2. Results and discussion

Results

A linear mixed regression model was calculated. The dependent variable was RT. Participants and targets were random factors. The fixed factors were Experimental condition (2a/2b), Biphone frequency (high/low), Syllable structure (CVC, CCVC, CVCC, CCVCC), and interactions of Biphone frequency * Syllable structure, Biphone frequency * Experimental condition,

[5] As Experiment 2a and 2b test the same predictions, the data were pooled to increase power and to minimize the number of models tested. Separate analyses of Experiment 2a and 2b did not reveal all predicted effects.

Experimental condition * Syllable structure, and Biphone frequency * Syllable structure * Experimental condition.

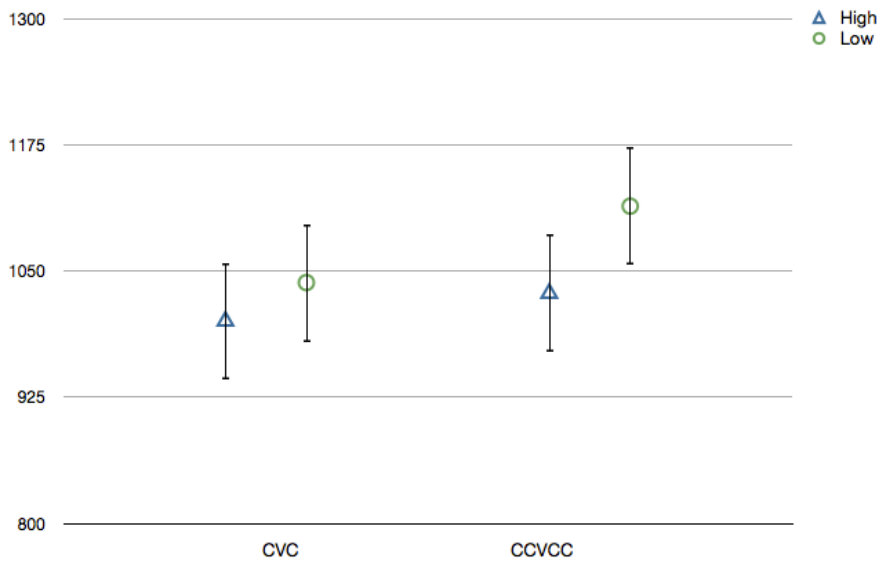

Figure 4: Averages of the estimated means and 95% confidence intervals of the reaction time in ms for CVC and CCVCC syllable structure nonwords for each high (H) and low (L) biphone frequency.

A linear mixed model revealed a significant main effect of condition (F (1, 5.798) = 5.884, $p < 0.05$) with items all over being recognized 129.78 ms faster in Experiment 2b than in 2a (SE = 53.69). This difference can be accounted for by a greater decay of attention in Experiment 2a, since the experiment session in Experiment 2b only took 35 minutes, while Experiment 2a took 75 minutes. Crucially, there were no interactions of Experimental condition * Biphone frequency, Experimental condition * Syllable structure, or a three way-interaction of Experimental condition * Biphone frequency * Syllable structure. This means that the two experimental conditions did not lead to differences in recognizing the nonwords in terms of their intrinsic properties. This allows for a pooled analysis of the data obtained in both conditions using the fixed factors biphone frequency (high/low), syllable structure (CVC, CCVC, CVCC, CCVCC), and the interaction of Biphone frequency * Syllable structure.

Syllable Structure	High	Low	Δ	Total *M*
CVC	1000.43	1038.53	38.10	1019.48
CVCC	1059.23	1137.62	78.39	1098.43
CCVC	1074.00	1118.14	44.14	1096.07
CCVCC	1027.93	1114.20	86.27	1071.07
Total *M*	1040.40	1102.12	61.73	1071.26

Table 5: Estimated reaction time averages in Experiment 2a and 2b in ms for high (High) and low (Low) biphone frequency measured from the target onset, and the differences (Δ) between them.

A linear mixed model of the fixed effects revealed a significant main effect of biphone frequency (see Appendix D), with high biphone frequency nonwords being recognized faster than low biphone frequency nonwords (see Table 5). There was also a significant main effect of syllable structure (see Appendix 3) with CCVCC nonwords being recognized more slowly than CVC nonwords (see Table 5). Finally, there was a significant interaction between biphone frequency and syllable structure (see Appendix D). More precisely, the difference in RTs for high versus low biphone frequency nonwords was significantly larger among complex CCVCC nonwords than among simple CVC items (see Figure 4).

Discussion

As predicted, participants performed significantly better in recognizing high rather than low biphone frequency nonwords. The effect of biphone frequency interacts with syllable structure. The difference in RTs between high and low biphone frequency is larger among complex CCVCC nonwords than among simple CVC nonwords (see Figure 2), indicating that the effect of probabilistic phonotactics increases with increasing syllable complexity. As opposed to Experiment 1, here, biphone frequency affected recognition latencies of CVC nonwords. On the one hand, this may be caused by the fact that items were controlled for speech rate. On the other hand, in Experiment 2 the ISIs were shorter than in Experiment 1, which may have elicited more low-level processing than the longer ISIs in Experiment 1.

5.4. General discussion

The results of the two experiments in this study indicate that, as hypothesized, sub-lexically represented knowledge affects phonological memory. Crucially, the sub-lexical representations that are used for redintegration are twofold with low-level probabilistic phonotactics on the one hand, and structural constraints as part of a phonological grammar on the other. The interaction of these two components, i.e., growing effects of phonotactic probability with increasing structural complexity, indicates that sub-lexical LTM knowledge is particularly important when rehearsing phonologically complex word forms. Less sub-lexical LTM knowledge is at play when simple CVC nonwords are rehearsed in the phonological loop. These results suggest that when processing hardly restricted CV and VC biphones, listeners make reference to low-level phonotactic probability knowledge, which, however, does not necessarily need feedback from a phonological grammar, as it is the case when structurally more restricted CC biphones are processed. With respect to Phonological Theory, this study supports the view that the effects of phonological grammar are not only categorical. In our experiments, all nonwords were made up of legal structures. They only differed in terms of the probability of biphones. Hence, the binary grammar distinction between legal and illegal cannot be the ultimate account. Furthermore, knowledge of phonological grammar seems to modulate the processing of categorically legal forms depending on their probability (e.g., Albright, 2009; Boersma & Hayes, 2001; Andries Coetzee, 2008; Hayes, 2000; Pierrehumbert, 2003).

Future studies may want to investigate whether the additive effect of structural complexity in low biphone frequency items necessarily relates to two representational components with probabilities on the one hand and markedness constraints on the other, or whether all effects may be accounted for by either a grammar or by probabilities.

The result of this study has indirect implications for theories of lexical acquisition. Factors that influence performance in STM nonword recall tasks have been suggested to similarly constrain lexical acquisition. Among these factors is, for example, the mechanism to draw onto LTM representations such as sub-lexical knowledge. LTM knowledge aids when holding a novel word form in short-term memory. Similarly, it helps to keep a novel word in the LTM storage when it has to be remembered for a long time, i.e., when it has to be acquired. Such conclusions are supported by the fact that performance in STM tasks has often been found to be correlated with lexicon size and lexical development over time (see Gathercole, 2006 for an overview).

The finding that both phonotactic probability and structural knowledge affect recognition memory thus indicates that each of these two sub-lexical components may be involved in facilitating lexical acquisition. As lexical neighborhood density was controlled for, the result must be attributed to effects

of a sub-lexical rather than the lexical level. Thus, the results are consistent with the hypothesis that the dependence between phonotactics and lexical acquisition is not only unidirectional, with the lexicon informing the phonological grammar, as is assumed by most phonologists (e.g., Hayes & Wilson, 2008; Pierrehumbert, 2003). Instead, two interacting sub-lexical knowledge components may play a role in lexical acquisition, in particular when complex word forms are remembered. This implies a bidirectional dependence.

Considering that phonotactic knowledge is represented at a sub-lexical level raises the question of how these sub-lexical representations are acquired. Most studies assume that sub-lexical representations emerge as abstractions over lexical representations. Pierrehumbert (2003), for example, assumes that at first the sub-lexical level only contains phonetically detailed representations. These phonetically detailed representations are used to create lexical representations. Later, phonotactic knowledge emerges as abstractions over the lexicon. An alternative view is that phonotactics is acquired bottom-up from speech (e.g., Adriaans & Kager, 2010; Boll-Avetisyan et al., submitted). For a large part, the source of lexical acquisition might be continuous speech rather than isolated words (e.g., Christophe, Dupoux, Bertoncini, & Mehler, 1993). The advantage of a bottom-up acquisition of phonotactics is that sub-lexical representations could facilitate lexical acquisition from the start when the first words are acquired. The current study cannot provide an ultimate answer to this question, as here effects of sub-lexical probabilistic and grammar knowledge were tested on nonwords presented in isolation. It would be interesting for future studies to test whether or not prosodic structure influences the acquisition of words from continuous speech.

We want to raise attention to the necessity of controlling for speech-rate in studies that test effects of probabilistic phonotactics on processing. The need for controlling for speech rate has also been discussed by Lipinsky and DasGupta (2005). They pointed out the relevance of the problem by demonstrating that the effects of probabilistic phonotactics on processing found by Vitevitch and Luce (1999) are hard to replicate if speech-rate is controlled (cf. Vitevitch & Luce, 2005). It is a non-trivial task to estimate the consequences of the confound for the hypothesis, since words composed of high frequent biphones are intrinsically spoken faster (Kuperman, et al., 2008). This means that two factors are correlated which are difficult to disentangle under natural conditions. However, a certain degree of naturalness may have to be sacrificed under experimental conditions if we want to ensure that predicted effects truly relate to the manipulated factor. Therefore, future studies should take this confound serious and control their test stimuli for speech-rate.

Hypothetically, the results of the current study could also be due to a mere interaction of phonotactic probability with word length determined by the total number of phonemes rather than the structural difference between CVC and CCVCC. A comparison of our results with a study by Frisch et al. (2000),

however, suggests that this should not be the case. Frisch et al. contrasted nonwords of different lengths (CVC, CVCVC, CVCVCVC, and CVCVCVCVC) of high versus low onset/rhyme probabilities in a recognition memory task, in which participants first judged the well-formedness of the items, and later had to distinguish the same items from filler items. They found that phonotactic probability affected recognition memory, while word length and an interaction of the two factors did not. Since the CCVCC items used in our study are of the same length as the CVCVC items used in Frisch et al.'s study, but differ in terms of complexity, we may assume that the factor modulating phonotactic probability is syllable complexity rather than word length.

5.5. Conclusion

The finding that structural complexity modulates phonotactic probability effects suggests that the sub-lexical level is informed by two components: probabilistic phonotactics and phonological grammar. Previous studies on the role of the sub-lexical level in nonword recognition have solely focused on one sub-lexical component, namely representations of probabilistic phonotactics. The relevance of a phonological grammar has been neglected. The results of the current study therefore suggest that the effects of sub-lexical knowledge on STM recognition might have arisen in previous studies (e.g., Roodenrys & Hinton, 2002), had they used complex stimuli rather than exclusively simple CVC stimuli. Therefore, it seems desirable that future experiments involve stimuli of increased structural complexity to test assumptions concerning the role of sub-lexical representations in lexical acquisition.

5.6. References

Adriaans, F., & Kager, R. (2010). Adding Generalization to Statistical Learning: The Induction of Phonotactics from Continuous Speech. *Journal of Memory and Language, 62*, 311-331.

Albright, A. (2009). Feature-based generalization as a source of gradient acceptability. *Phonology, 26*(1), 9-41.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philedelphia: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighbourhoods? *Journal of Memory and Language 44*, 568-591.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*(1), 45-86.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.09). Retrieved June 24, 2007, from http://www.praat.org/.

Boll-Avetisyan, N. A. T., Kager, R. W. J., de Bree, E., Kerkhoff, A., van den Bergh, H., & den Boer, S. (submitted). Does the lexicon bootstrap phonotactics, or vice versa? .

Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics in Language Teaching, 22*, 253-269.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1993). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95*(3), 1570-1580.

Coetzee, A. (2008). Grammar is both categorical and gradient. In S. Parker (Ed.), *Phonological Argumentation*. London: Equinox.

Coetzee, A., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory, 26*, 289-337.

Coleman, J., & Pierrehumbert, J. B. (1997). *Stochastic phonological grammars and acceptability*. Paper presented at the 3rd Meeting of the ACL Special Interest Group in Computational Phonology, Somerset, NJ.

Cowan, N., Wood, N. L., Nugent, L. D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposing effects of duration and complexity. *Psychological Science, 8*, 290–295.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77-91.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, Serial No. 242, 59*(5), i-185.

Fikkert, P. (2007). Acquiring phonology. In P. d. Lacy (Ed.), *Handbook of phonological theory* (pp. 537-554). Cambridge, MA: Cambridge University Press.

Frisch, S. A., Large, N., & Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*, 481-496.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory 22*, 179-228.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513-543.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84-95.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech--a syllable-centric perspective. *Journal of Phonetics, 31*, 465-485.

Hayes, B. (2000). Gradient well-formedness in Optimality Theory. In J. Dekkers, F. v. d. Leeuw & J. v. d. Weijer (Eds.), *Optimality Theory: Phonology, syntax, and acquisition*. Oxford: Oxford Univeristy Press.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39*(3), 379-440.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell. Reprint (1969). Frankfurt am Main: Suhrkamp.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630-645.

Kuperman, V., Ernestus, M., & Baayen, R. H. (2008). Frequency Distributions of Uniphones, Diphones and Triphones in Spontaneous Speech. *Journal of the Acoustical Society of America, 124*(6), 3897-3908.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.

Levelt, C., & Vijver, R. v. d. (1998/2004). Syllable types in cross-linguistic and developmental grammars. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Fixing Priorities: Constraints in Phonological Acquisition* (pp. 204-218). Cambridge, England: Cambridge University Press.

Lipinsky, J., & Gupta, P. (2005). Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language, 52*, 171-192.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107*, 395-419.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91-121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge MA: MIT press.

Pitt, M., & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347-370.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*: Rutgers University and University of Colorado at Boulder, 1993, Revised version published by Blackwell, 2004

Roodenrys, S., & Hinton, M. (2002). Sublexical or Lexical Effects on Serial Recall of Nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 29-33.

Scholes, R. (1966). *Phonotactic Grammaticality*. The Hague: Mouton.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall: short-term memory. *Memory and Cognition, 21*, 168-175.

Selkirk, E. O. (1982). The Syllable. In H. v. d. Hulst & N. Smith (Eds.), *The Structure of Phonological Representations, Part II* (pp. 337–383). Dordrecht: Foris.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155-172.

Smolensky, P. (1996). The Initial State and 'Richness of the Base' in Optimality Theory. *154*(1196). Retrieved from http://roa.rutgers.edu/files/154-1196/roa-154-smolensky-2.pdf

Sternberg, S. (1966). High speed scanning in human memory. *Science, 5*, 653-654.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321-1337.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotacic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*, 1175-1192.

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics., 14*, 407-425.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory and Cognition., 31*(4), 729-735.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 145-152.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325-329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighbourhood Activation in Spoken Word Recognition. *Journal of Memory and Language, 40*, 374-408.

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language 52*, 193-204.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Appendix A

Target Stimuli used in Experiment 1 and 2

High biphone frequency nonwords.

CVC: beːl, bɛx, deːf, deːk, foːm, foːt, xaːk, hɑs, jaːt, kɑx, laːm, meːf, mɛl, neːk, raːl, raːn, rɪn, rɔf, roːn, sɛŋ, tɛx, tɛŋ, voːk

CVCC: beːls, beːrk, bɛxt, deːft, deːks, xaːkt, hɑst, jaːrt, kɑrk, liːnt, liːts, meːft, mɛls, neːks, rɑls, rɛns, rɔxt, sɛŋk, sɛrt, tɛxt, tɛkt, tɔrm, voːkt

CCVC: blɑs, brɛx, brɛl, brɪn, broːn, dreːk, froːm, froːn, froːt, xraːk, xroːn, klɑx, klɑr, krɛx, krɔf, praːn, prɪn, slaːm, slaːr, traːl, trɛŋ, trɪn, twɪl

CCVCC: blɑst, bleːrk, brɛlt, brɪnt, dreːks, frɑls, frɛns, froːns, xraːkt, klɑrm, kliːnt, krɛxt, pliːts, prɪŋk, slaːrs, stɛŋk, stɛrt, traːls, trɛŋk, trɔxt, trɔrm, twɛkt, twɪlt

Low biphone frequency nonwords.

CVC: bɪr, dyl, hyl, kiːŋ, kux, kym, lux, lum, lʏt, mʏp, myt, ryk, sum, sur, tɪr, vʏŋ, wiːŋ, wur, zaŋ, zus, zʏp, zyx, zyl

CVCC: bɪrx, dylk, hylm, kiːŋs, kuŋs, kymp, lump, lurx, lʏmp, lʏmt, mʏps, nurn, sums, surp, tɪrf, viːŋt, vʏŋt, vʏrf, wumt, wuŋt, wurn, zaŋt, zylm

CCVC: dwiːŋ, dwuŋ, dwur, dwʏŋ, dwyw, flux, xlun, xlʏt, knux, knʏm, knʏŋ, knʏp, kwiːŋ, smyx, smyt, snum, vlum, vluŋ, wryk, zwʏŋ, zwʏp, zwyx, zwyl

CCVCC: dwiːŋt, dwumt, dwuŋt, dwurx, dwʏŋt, dwʏrf, dwywt, xlump, xluŋt, xlʏmp, knʏms, knʏŋs, knʏps, snump, snurn, snurp, vluŋt, vlʏmt, xlʏps, vlyms, vlʏps, dwyŋt, xlumt

Appendix B

Multilevel Model

Reaction times are nested both within individuals and within stimuli. Hence, a multilevel model is appropriate for the analysis of the results. Failing to take the different variance components into account will result in an underestimation of the variance and hence the testing statistics will be too optimistic (the null-hypothesis is rejected although the data in fact do not support this conclusion). In order to test the hypothesis we define several dummy variables, which are turned on if a response is observed in the respective condition and are turned off otherwise. Let Y(ij) be the response on item *i* (*i* = 1, 2, …, I) of individual *j* (*j* = 1, 2, …, J), and High biphone frequency$_{(ij)}$, CCVC$_{(ij)}$ CVCC$_{(ij)}$ and CCVCC$_{(ij)}$ be dummy variables indicating whether the item is a High biphone frequency nonword with CCVC, CVCC or CCVCC structure respectively. The interaction between biphone frequency and syllable type can be estimated by defining combinations of the biphone frequency dummy and the syllable type dummies. Analogue to analysis of variance a saturated model with both main and interaction effects can be written as:

$$
\begin{aligned}
Y(ij) = {} & \text{CONSTANT} + \beta_1 * HI_{(ij)} + \\
& \beta_2 * CCVC_{(ij)} + \beta_3 * CVCC_{(ij)} + \beta_4 * CCVCC_{(ij)} + \\
& HI_{(ij)} (\beta_5 * CCVC_{(ij)} + \beta_6 * CVCC_{(ij)} + \beta_7 * CCVCC_{(ij)}) + \\
& [e_{(ij)} + u_{i0} + u_{0j}]
\end{aligned}
$$

The model above consists of two parts: a fixed and a random part (between square brackets). In the fixed part the constant (i.e., the intercept) represents the mean of Low biphone frequency CVC items, and the other effects represent deviations from this average. So, the reaction time to High biphone frequency CVC items is (Constant + β_1), and the average of low frequency CCVC items is (Constant + β_2), etc.

In the random part three residual scores are defined: $e_{(ij)}$, u_{i0} and u_{0j}. The last term (u_{0j}) represents the deviation of the average reaction time of individual *j* from the grand mean, u_{i0} represents the deviation of item *i* from the grand mean, and finally $e_{(ij)}$ indicates the residual score of individual *j* and item *i*. We assume that these residual scores are normally distributed with an expected value of 0.0 and a variance of S^2_e, S^2_{ui} and S^2_{uj} respectively.

Appendix C

Results of Experiment 1

Estimates of Fixed Effects			
Parameter	**Estimate**	**SD**	**Significance**
Intercept (low biphone frequency CVC)	1103.75	44.58	.000
high biphone frequency	14.07	28.84	ns
CCVCC	117.65	28.93	.000
CCVC	78.61	28.91	.007
CVCC	69.07	28.87	.018
high * CCVCC	-102.39	40.79	.013
high * CCVC	-88.75	40.79	.031
high * CVCC	-72.47	40.77	.077

Estimates random Parameters			
Parameter	**Estimate**	**SD**	**Significance**
S^2_{uj} (items)	6926.49	1,019.25	< 0.001
S^2_{uj} (participants)	47,118.08	12,486.48	< 0.001
$S^2_{e(ij)}$ (residual)	72,772.29	1,473.28	< 0.001

Appendix D

Results of Experiment 2

Estimates of fixed effects			
Parameter	**Estimate**	**SD**	**Significance**
Intercept (low biphone frequency CVC)	1038.53	29.20	.000
high biphone frequency	-38.10	15.17	.012
CCVCC	75.67	15.27	.000
CCVC	79.61	20.19	.000
CVCC	99.09	21.44	.000
High biphone frequency * CCVCC	-48.17	20.60	.019
high biphone frequency * CCVC	-6.04	28.48	ns
high biphone frequency * CVCC	-40.29	29.65	ns

Estimates random Parameters			
Parameter	**Estimate**	**SD**	**Significance**
$S^2_{e(ij)}$ **(residual)**	94,929.01	1,344.73	< 0.001
S^2_{uj} **(participants)**	23,922.15	5,992.15	< 0.001
S^2_{uj} **(items)**	6,345.37	912.50	< 0.001

CHAPTER 6

SECOND LANGUAGE PROBABILISTIC PHONOTACTICS AND SYLLABLE STRUCTURE IN SHORT-TERM MEMORY RECOGNITION[1]

6.1. Introduction

Many studies informed by the working memory model (Baddeley & Hitch, 1974) have demonstrated a link between short-term memory (STM) recall performance and long-term memory (LTM) storage. The easier it is to hold an item in STM, the easier it is to store it in LTM. A large body of psycholinguistic experiments has suggested that this link influences lexical acquisition. Nonword recall can be seen as the initial step in storing new word forms in the mental lexicon. Indeed, performance in nonword recall tasks correlates with vocabulary size in adults and children down to 2 years of age in a native language (L1) (for an overview see Gathercole, 2006) and second language (L2) (e.g., Cheung, 1996; Masoura & Gathercole, 1999; Service & Kohonen, 1995; Speciale, Ellis, & Bywater, 2004).

The link between phonological STM storage and LTM knowledge, however, does not seem to be exclusively unidirectional, as LTM representations have an influence on how well an item can be held in the phonological loop. It is assumed that LTM representations are used for reconstructing degraded memory traces during sub-vocal rehearsal in STM. This process is referred to as

---

[1] A shorter version of this chapter was published as Boll-Avetisyan, N. A. T. (2011). Second Language Probabilistic Phonotactics and Structural Knowledge in Short-Term Memory Recognition. In N. Danis, K. Mesh & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (Vol. 1, pp. 60-72). Somerville, MA: Cascadilla Press.

The current version has been submitted for publication. The study was funded by an NWO grant (277-70-001) awarded to René Kager.

redintegration (Schweickert, 1993). Thus, long-term language knowledge mediates STM recall and can have a facilitatory effect on lexical acquisition. LTM effects on STM recall are manifold. For example, words are recalled with more accuracy than nonwords (e.g., Hulme, Maughan, & Brown, 1991), recall performance is higher for high frequency words than for low frequency words (e.g., Engle, Nations, & Cantor, 1990; Hulme et al., 1997), and recall performance is higher for words of a familiar language than for words of an unfamiliar language (e.g., Hulme, et al., 1991).

STM recall is not only affected by lexical knowledge but also by knowledge of the probability of sub-lexical units. For example, when stimuli are manipulated for phonotactic probability, children and adults are better at recalling nonwords composed of high frequency biphones than those consisting of low frequency biphones (Boll-Avetisyan, to appear; Coady & Aslin, 2004; Gathercole, Frankish, Pickering, & Peaker, 1999; Thorn & Frankish, 2005). This effect can also be found in adult L2 learners, who are better at recalling nonwords of high than of low biphone frequency in the L2 (Majerus, Poncelet, Van der Linden, & Weekes, 2008). Note that sub-lexical knowledge of probabilistic phonotactics is different from lexical knowledge derived by accessing the frequency of occurrence of phonologically similar words (*lexical neighbors*) in the mental lexicon, although the two are highly correlated (Landauer & Streeter, 1973; Vitevitch & Luce, 1999). Previous research has shown that lexical factors, such as lexical neighborhood density, and sub-lexical factors, such as biphone frequency, influence nonword recall independently of each other. That is, a dense lexical neighborhood facilitates recall when biphone frequency is controlled for (Roodenrys & Hinton, 2002; Thorn & Frankish, 2005, but c.f. Goh & Pisoni, who find inhibitory effects), and high biphone frequency facilitates recall when lexical neighborhood density is controlled for (e.g., Boll-Avetisyan, to appear; Majerus, Van der Linden, Mulder, Meulemans, & Peters, 2004; Thorn & Frankish, 2005).

There is some evidence that knowledge of probabilistic phonotactics influences lexical acquisition. It predicts vocabulary size in studies with monolingual and bilingual children (Gathercole, et al., 1999; Messer, Leseman, Boom, & Mayo, 2010) and general L2 proficiency in adult L2 learners (Majerus, et al., 2008). Furthermore, in offline lexical learning experiments, unfamiliar word/meaning pairs are learned more rapidly for nonwords with high probabilistic phonotactics than for those with low probabilistic phonotactics (e.g., Storkel, 2001). Given the evidence suggesting that sub-lexical knowledge is an important factor in the success of lexical acquisition, the current study takes a closer look at which sub-lexical factors influence memory of L2 nonwords.

Crucially, previous studies suggest that two different kinds of phonotactic knowledge are represented at a sub-lexical level. The first type is specific sequential knowledge of *n*-phones (i.e.,, linear strings of phonemes) and their

probability of occurrence, usually referred to as probabilistic phonotactics (e.g., Vitevitch & Luce, 1999). The second type is more abstract knowledge of structural restrictions on the (co-)occurrence of phonemes in a language; for example, constraints on the complexity of syllable structure (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Pierrehumbert, 2003). When investigating the role of sub-lexical knowledge in STM performance, most studies have looked at effects from one factor: probabilistic phonotactics. Structural phonological knowledge has rarely been addressed by research on effects of LTM knowledge on phonological STM, even though, as will be discussed below, it might have an important influence on lexical acquisition, especially in an L2.

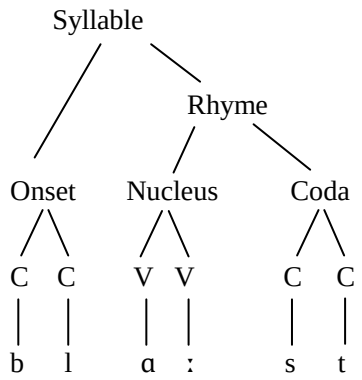

Figure 1: Syllable tree for the English CCVCC word *blast.*

The notion of structural phonological knowledge has been proposed in phonological theory, based on evidence from typology and acquisition. Many languages (e.g., Japanese, Mandarin Chinese) ban complex syllable structures with consonant clusters in syllable margins, but there are no languages which tolerate complex syllables, but disallow simple syllable structures. This mirrors patterns in language acquisition: children usually acquire simple syllable structures before complex syllable structures (e.g., Fikkert, 1994; Jakobson, 1941; Levelt & Vijver, 1998/2004). Furthermore, phoneme distributions are restricted within syllables. Typologically, consonant clusters in syllable margins are often subject to strong co-occurrence restrictions. In contrast, much weaker restrictions hold between VC sequences, and CV sequences are almost unrestricted. Therefore, phonological theory supplies a hierarchical internal organization of syllables, with nucleus and coda being dominated by a rhyme-

constituent, and onset and rhyme being dominated by the syllable (e.g., Fudge, 1969; Selkirk, 1982) (see Figure 1).

A large body of studies provides evidence that speakers possess hierarchical knowledge of syllable structure. Speech errors, for example, never result in phonotactically illegal structures (e.g., Fromkin, 1971), and when errors occur, syllable constituents remain intact more often than expected if syllables were just strings of *n*-phones (MacKay, 1972). Similarly, in word games, biphones that are part of a constituent are more cohesive than biphones that straddle across constituents (e.g., Treiman, 1983).

In more recent psycholinguistic studies, participants were required to read aloud lists containing repetitions of nonwords in which some phonemes were more frequent in the onset position and other phonemes were more frequent in the coda position. When speech errors occurred, they respected the phonotactics of the native language, and furthermore the errors revealed that the participants had learned the phoneme distribution probabilities from the stimuli (e.g., Dell, Reed, Adams, & Meyer, 2000; Goldrick, 2004).

Importantly, there is evidence for the relevance of hierarchical structural knowledge in STM recall: Treiman and Danis (1988) had participants repeat series of nonwords of CVC, CCV or VCC structures. A count of the erroneous repetitions revealed fewer errors in biphones that are part of a syllable constituent than in biphones that straddle a constituent boundary. Furthermore, words involving more phonemes are recalled less accurately than words involving fewer phonemes. This effect has been ascribed to phonological complexity (Service, 1998).[2] Also, 4-year-old children make fewer errors when repeating nonwords with singleton consonants than with consonant clusters (Gathercole & Baddeley, 1989).

These studies, however, did not control for probabilistic phonotactics, so it might be the case that fewer errors occurred in the biphones of higher frequency—a factor that was not known to influence recall at the time the studies were carried out. Boll-Avetisyan (to appear) demonstrated that both

[2] An alternative account is that this effect is due to stimulus duration (e.g., Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.; and Cowan, N., Wood, N. L., Nugent, L. D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposing effects of duration and complexity. *Psychological Science, 8*, 290–295.) However, later studies have successfully disentangled duration and complexity effects, finding robust effects of complexity (Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51*(A), 283-304.), but only weak effects of duration (Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology, 53*(A), 1-22.) Hence, even if effects from duration and complexity add up, complexity should play the major role.

syllable complexity and probabilistic phonotactics influence STM. Probed STM recognition performance was better for high than for low biphone frequency nonwords, and better for simple than for complex syllables. The facilitatory effect of high biphone frequency even increased with increasing syllable complexity. This suggests that probabilistic phonotactics and abstract hierarchical representations are two separate sub-lexical components that interact in affecting the redintegration process, countering decay.

If it is true that probabilistic phonotactics and hierarchical structural knowledge belong to two separate knowledge components, then it should be possible to acquire them independently. L2 acquisition provides an ideal test case to investigate this possibility. L2 probabilistic phonotactic knowledge can be acquired, improves with L2 proficiency and affects L2 nonword recall (e.g., Lentz & Kager, submitted; Majerus, et al., 2008; Trapman & Kager, 2009). However, L2 learners are well-known for having difficulties with acquiring and processing phonological structures that are "illegal" in the native language. This results in simplifications of complex syllables, for example by creating two simple syllables through vowel epenthesis, both in production (e.g., Broselow, 1984) and perception (see Dupoux, et al., 1999 for a study on non-native listening; Lentz & Kager, submitted for a study on L2 perception). Perceptual epenthesis can be accounted for by phonological models of perception, according to which a perceived utterance goes through a phonological filter that prepares lexical access. This filter checks whether the perceived utterance can be mapped onto a structurally licit form. If the utterance was pronounced in a way that it cannot be mapped onto a structurally licit form, it will be "repaired" (e.g., Boersma, 2000).

Crucially, as yet, studies have only tested L2 sequential knowledge of probabilistic phonotactics for biphones that were structurally licit in the participants' L1. Only a very recent study by Lentz and Kager (submitted) showed that Spanish L2 learners of Dutch were affected by L2 phonotactic probabilities of L1-illegal Dutch /sC/-clusters in an auditory lexical priming task. However, in their study, the probability effect was only present when participants listened to epenthesized variants of the clusters (i.e., /ɛsC/), but not when they heard the clusters as they are actually pronounced in Dutch. The study is the first to suggest that L2 probabilistic phonotactic knowledge can actually be acquired for consonants that are illicit in an L2. Still, as yet, no study has shown that L2 probabilistic phonotactic knowledge of consonant clusters that are illicit in the L1 affects actual speech processing, such as lexical recognition, in an online processing task.

The current study addresses the question of whether L2 learners can, in spite of transfer of L1 structural knowledge, use probabilistic L2 knowledge about biphones which are structurally illicit in their L1 in lexical recognition. It may seem contradictory that L2 n-phone probabilities for L1 illicit sequences should be able to be acquired, if they are filtered by L1 structural knowledge and

"repaired" into L1 licit structures. There are two possibilities that would give reason for such a hypothesis: The first possibility is that the phonological filter might be weakened during L2 acquisition, especially at higher levels of L2 proficiency. Thus, some L1 illicit structures may pass through, and their probability will be acquired. The second possibility is that probabilistic phonotactic knowledge is not grammatical. If it were represented outside the grammar and processed before the filter at a relatively shallow level, then in theory, probabilities could be acquired independent of whether the grammar favors them as a structure or not.

The central hypothesis of this paper is that sub-lexical representations are two-fold with representations of probabilistic phonotactics on the one hand and hierarchical phonological structure on the other hand. The hypothesis of the two different knowledge components leads us to the hypothesis that they can be acquired independently. Regarding L2 sub-lexical knowledge, we state the following hypotheses: L2 learners should have difficulties acquiring L2 structural knowledge due to a transfer of the native phonological grammar. Still, L2 learners should be able to acquire L2 probabilistic phonotactic knowledge even of *n*-phones that are unattested in their L1, as *n*-phones are represented independently of phonological structure.

In line with this, we hypothesize that the two separate sub-lexical knowledge components have independent effects on L2 nonword recognition: First, L2 nonword recognition should be facilitated by L2 probabilistic phonotactic knowledge even of structures that are unattested in the L1. Second, recognition of L2 syllables that are structurally more complex than the L1 allows should still be impeded by transfer of L1 structural knowledge.

The hypotheses were tested in an experiment with adult native speakers of Dutch, and Japanese and Spanish L2 learners of Dutch.

	Japanese	**Spanish**	**Dutch**
CVC	Yes	Yes	Yes
CCVC	No	Yes	Yes
CVCC	No	No	Yes
CCVCC	No	No	Yes

Table 1: Attestation of syllable templates in the languages under test.

These languages were selected because Dutch allows for more complex syllables than either Japanese or Spanish. Dutch allows for complex syllable

onsets and codas (Trommelen, 1983), whereas Japanese only allows for CVC (Kubozono, 1989), and Spanish only allows for complex onsets, but not complex codas (Harris, 1983) (see Table 1).

The stimuli were monosyllabic nonwords manipulated for biphone frequency (high and low) to assess effects of both probabilistic phonotactics and syllable complexity (CVC, CCVC, CVCC & CCVCC) to determine effects of structural knowledge. The items were presented in a probed recognition task, following Sternberg's high-speed, serial, exhaustive scan paradigm (Sternberg, 1966). In the experiment, a series of four stimuli is followed by a probe, and participants are asked to decide whether the probe was in the series or not. Dependent measures are error rates and reaction times (RTs). The task, which has previously been successfully used to reveal effects of probabilistic phonotactics and syllable structure (Boll-Avetisyan, to appear), has two advantages that make it appropriate for the purpose of this study. First, as opposed to the classically used serial nonword repetition task, STM recognition performance is not influenced by language-specific motorical skills, which are non-native-like in L2 learners. The advantage of *probed* over *serial* recognition is that the latter is less susceptible to effects of probabilistic phonotactics and LTM representations in general (Gathercole, et al., 1999), because participants focus more on serial order memory than on the specific properties of the stimuli.

Two predictions can now be stated for this experiment. First, we predict that high biphone frequency will have a facilitatory effect on recognition. We expect this effect to occur in monolingual speakers as well as L2 learners in the recognition of all syllable structure types, even of those that are unattested in the L1 of the learners. Second, we expect an interaction of L1 and syllable structure with worse recognition performance of syllable structures that are unattested in the L1 than of those that are attested. Hence, Japanese should have difficulties with CCVC, CVCC, and CCVCC items and Spanish should have difficulties with CVCC and CCVCC items. Third, we predict an interaction of L1, syllable structure and biphone frequency. This is because previous literature suggests that structural knowledge enhances the effect of probabilistic phonotactic knowledge on recognition, as effects of biphone frequency increase with increasing syllable complexity. We assume that if an L2 learner does not have complex syllables in the L1, then his phonological grammar will not license complex structures. That means that his phonological grammar cannot support that such complex structures will be held in the phonological loop. Hence, regarding the predicted interaction, we predict that the effect of biphone frequency will only be enhanced in complex L2 structures L2 learners know from their L1 (or have acquired in the L2).

The dependent variables were "accuracy rate" and "reaction time" (RT). RTs in general and RTs of erroneous responses in particular are not commonly analyzed and interpreted in studies on STM. Yet several considerations suggest that it might be worthwhile to look at sub-lexical effects on RTs and on RTS in

erroneous responses in particular. First, RTs are known to be a more fine-grained measure of cognitive performance than error rates by themselves. Second, error rates are affected by both RTs in the erroneous and correct responses (see literature on speed-accuracy trade-off and models of drift rate, e.g., Ratcliff, Gomez, & McKoon, 2004). We were interested in the RTs on incorrect responses, as they, by indicating how fast an item has decayed, might shed more light on the factors that lead to failure of holding an item in the phonological loop in addition to the RTs on correct responses, which may reveal factors that facilitate recall. Furthermore, in a word-learning task, Storkel and colleagues (Storkel, Armbrüster, & Hogan, 2006), only found an inhibitory effect of high phonotactic probability among items that were partially remembered, i.e., among errors in which only two out of three phonemes were correctly reproduced. Hence, in this study, RTs were measured not only for the correct, but also for the incorrect responses, in which correct items had erroneously been rejected.

6.2. Experiment

6.2.1. Method

Participants

Participants were 30 native speakers (26 women and 4 men, M_{age} = 23.17, age range: 19-43) and 58 L2 learners of Dutch, of whom 31 were native speakers of Spanish (22 women, 9 men, M_{age} = 36.21, age range: 19-58) and 27 native speakers of Japanese (25 women, 2 men, M_{age} = 41.60, age range: 23-68). We focused on learners of high L2 proficiency, since proficient learners are known to rely more on L2 LTM knowledge than beginning learners, who rely mostly on general STM capacities in recall tasks (Speciale, et al., 2004). We received statements about the length of residence in the Netherlands (LoR) by 21 Japanese (M_{LoR} = 11.9 years; *SD*=9.61) and 28 Spanish participants (M_{LoR} = 10.18 years; *SD* = 8.23). Proficiency level was assessed with a C-test (from Trapman & Kager, 2009).[3] Scores were obtained from 21 Japanese ($M_{answer_correct}$

[3] A C-test is similar to a cloze test with the difference being that a few initial letters are always given, and the participants have to complete the words by filling the gap. Performance in this procedure has been found to correlate with multiple domains of language proficiency Eckes, T. (2010). Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In R. Grotjahn (Ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-test: Contributions from current research* (pp.

= 29.76%, *SD* = 15.11) and 30 Spanish participants ($M_{answer_correct}$ = 34.30%, *SD* = 19.34). All participants reported having normal hearing and were compensated for participation by a fee.

Materials

Materials were the same as in Boll-Avetisyan (to appear). All stimuli were nonwords that are phonotactically legal in Dutch. That is, all phonemes are part of the Dutch phoneme inventory and all biphones are licit sequences in the syllable positions in which they occur. The target stimuli (see Appendix A) were manipulated along two dimensions. First, four different syllable structure types were used: CVC, a legal template in all three native languages, CCVC, legal in Spanish and Dutch but not in Japanese, and CVCC and CCVCC, both legal in Dutch but not in Spanish and Japanese.

Biphone frequency	**High**	**Low**
#CC	5252.72 (4986.55), N=46	580.17 (309.65), N=46
CV	6190.61 (3960.58), N=92	632.47 (444.91), N=92
VC	9132.51 (6510.22), N=92	624.74 (575.24), N=92
CC#	8711.96 (6777.85), N=46	1096.13 (560.07), N=46

Table 2: Averages of token biphone frequencies (#CC= consonant cluster in onset, CC# = consonant cluster in coda position) of the stimuli. Numbers in parentheses represent standard deviations.

Second, targets were composed of Dutch biphones of high versus low frequency. Biphone frequency counts were based on biphone occurrences in a lexicon of 8,305 monomorphemic word types taken from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). All biphones of the low biphone frequency target stimuli were taken from the least frequent 33% of all biphones, and all biphones of the high biphone frequency target stimuli were taken from the most frequent 33% of all biphones. Mean biphone frequencies are given in Table 2. To minimize any effects of unknown factors and to provide consistency between items, the targets were created in such a way that they would only minimally differ from each other, e.g., the low biphone probability nonwords /lum/, /lump/, and /xlump/ or the high biphone probability nonwords

125-192). Frankfurt am Main: Lang.. Here, five short texts with each 20 gaps, i.e., 100 gaps in total, were given.

/vo:k/ and /vo:kt/. Thus, interference of singleton frequency effects, which are known to influence phonotactic well-formedness (e.g., Bailey & Hahn, 2001), was minimized.

Furthermore, the target stimuli were controlled for lexical neighborhood density at all levels of both factors in order to assess pre-lexical probabilistic phonotactic effects on recall while eliminating lexical confounds. Following Luce (1986), a lexical neighbor was defined as any word differing from the nonword by the addition, deletion, or substitution of a single phoneme. Lexical neighborhood density was calculated by summing the logged token frequencies of lexical neighbors for each test item. The summed log token frequencies are given in Table 3.

LND	**High**	**Low**	**Total *M***
CVC	2.13 (0.20)	2.03 (0.20)	2.08 (0.20)
CCVC	2.13 (0.15)	2.07 (0.16)	2.10 (0.16)
CVCC	2.06 (0.17)	2.04 (0.23)	2.05 (0.20)
CCVCC	2.01 (0.20)	2.11 (0.27)	2.06 (0.24)
Total *M*	2.08 (0.18)	2.05 (0.21)	2.05 (0.21)

Table 3: Average Lexical Neighborhood Density (LND) of each syllable type of high and low biphone frequency. Numbers in parentheses represent standard deviations. Number of items is 23 for each cell.

We did not control for Spanish and Japanese probabilistic phonotactics. L2 learners can only transfer L1 probabilistic phonotactic knowledge of biphones that are structurally licit in the L1, while the frequency of structurally illicit biphones (e.g., CC in CCVCC items) is zero. Hence, transfer can only minimally affect the recognition of items that are the main focus of our interest. It might only be a confound for the stimuli that are needed as a baseline for comparison, namely items that are structurally licit in the L1 of the L2 learners (e.g., CVC syllables in Japanese and Spanish and CCVC syllables in Spanish).[4]

[4] It needs to be kept in mind that in Spanish, CC biphones occur in syllable onsets. Hence, if an acquisition of L2 biphone probabilities is independent of structural knowledge, we need to assume that Spanish L2 learners of Dutch might transfer Spanish CC biphone probabilities to the L2. Also, it might be that the acquisition of CC biphones in general will be easier for Spanish than for Japanese L2 learners. The Japanese L2 learners really need to acquire all CC biphones from scratch.

The stimuli were spoken by a Dutch native speaker, who was naive to the purpose of the study. Recordings were made in a sound-proof booth. All nonwords were recorded multiple times, so that we could select two different recordings of the same nonword type as target and probe. As a consequence of this, participants could not rely on cues from complete acoustic overlap of target and probe during recognition.

Lastly, we controlled the items for duration. Speech-rate affects recall latencies (e.g., Baddeley, et al., 1975; Cowan, et al., 1997). High biphone frequency words are generally spoken faster than low biphone frequency words (Kuperman, Ernestus, & Baayen, 2008). We found such a difference in duration in our stimuli. Therefore, we controlled for duration between the high and low frequency items by adjusting the vowel length. Consonants were not changed in duration to maintain the naturalness of the onset and coda constituents, which are the focus of this study. Using the software PRAAT (Boersma & Weenink, 2007), manipulations were carried out on both sides: Stimuli with long durations were shortened, and stimuli with short durations were lengthened, until for each class of syllable structure the average durations of the stimuli did not differ between high and low frequency biphone nonwords. Two native speakers of Dutch confirmed that the resulting stimuli sounded natural. The fillers used in this experiment were randomly selected from a list of Dutch nonwords without further prerequisite.

Procedure

Participants were tested in a probed recognition task, in which a series of four nonwords followed by a probe is presented. The task was to decide whether the probe was in the series or not. Target series contained one target and three fillers, and filler series contained four fillers. The series were designed such that every syllable template occurred once in each trial. Every series thus contained the same number of segments. An example is given in Table 4.

Series	**item 1**	**item 2**	**item 3**	**item 4**	**probe**	**appeared before?**
Target	/do:rk/	**/tɛx/**	/zwa:lm/	/ gɔp/	**/tɛx/**	Yes
Filler	/ brint/	/tɛp/	/ bi:rf /	/ frɑp /	**/lyn/**	No

Table 4: Examples of a target and a filler series.

The experiment consisted of 184 target series. The design had two factors (biphone frequency, syllable structure) with 2x4 levels (high/low; CVC, CVCC, CCVC, CCVCC); accordingly, the total of 184 targets divides into 23 targets of

each type. In addition, 184 filler-series, in which the probe did not match any of the prior four fillers, were included. All series and all items within the series were randomized for every participant in order to avoid item- or series-specific order effects. Serial position was probed at random.

Participants were tested individually. The monolingual Dutch listeners and 13 of the L2 learners were tested in a soundproof booth, and 45 of the L2 learners were tested in a quiet room. The stimuli were played by computer over a set of headphones at a comfortable volume. The participants were facing a computer screen during the experiment. The stimuli were presented with an inter-stimulus-interval of 700 ms. The probe was presented 1400 ms after the last item of the stimulus series. Yes/No-decisions about whether the probe had occurred in the series were made on a button-box. After the end of each series, an icon appeared on the screen indicating the beginning of a new trial. When no decision was made after 3000 ms, the trial was stopped and counted as an error. The experiment took 75 minutes. Three self-timed breaks were included. There were three practice trials before the experiment started. After the experiment, the L2 participants had to fill in the C-test, by which we measured the level of Dutch proficiency, and a questionnaire to collect bibliographical data such as time of residence in the Netherlands and exposure to Dutch.

6.2.2. Results and discussion

All participants were included in the analysis. Fillers were excluded. Reaction times (RTs) were measured from the onset of the target probe.

Accuracy rates

A logistic regression model with the rate of correct responses as the dependent variable and participants and targets as random factors was calculated. The fixed factors were biphone frequency (high/low), syllable structure (CVC, CCVC, CVCC, CCVCC), native language (Dutch, Spanish, Japanese), L2 proficiency and the interactions of these factors. Error data is binomial. Hence, all statistical tests were performed on the logit transformations of the errors. Mean proportions, their logits and the standard errors of the logits are given in Appendix B. Estimates of the proportional means of correct answers by each group for the different stimuli types are given in Table 5.

There was a main effect of biphone frequency, with the proportion of correct answers being lower for low biphone frequency (85) than for high biphone frequency nonwords (90; $\chi^2 = 100.04$, $df = 1$, $p < 0.001$). Biphone frequency interacted with language: biphone frequency had a stronger effect on Japanese than on Spanish and Dutch participants such that Japanese were slightly worse at

recalling low biphone frequency nonwords (χ^2 = 11.65, *df* = 1, $p < 0.001$). The biphone frequency effect, nonetheless, was present in all groups of participants (Japanese: χ^2 = 52.42, *df* = 1, $p < 0.001$; Spanish: χ^2 = 5.82 *df* = 1, $p < 0.05$; Dutch: χ^2 = 33.20, *df* = 1, $p < 0.001$). Furthermore, there was a main effect of L2 proficiency: The higher the c-test scores, the fewer errors participants made ($p > 0.05$). L2 proficiency interacted with high biphone frequency: With increasing L2 proficiency, recognition accuracy increased for high biphone frequency nonwords ($p < 0.05$). The interaction between syllable structure and biphone frequency was nonsignificant.

	Japanese			**Spanish**			**Dutch**		
	Hi	**Lo**	**Δ**	**Hi**	**Lo**	**Δ**	**Hi**	**Lo**	**Δ**
CVC	91	86	05	87	83	04	91	87	04
CCVC	91	85	06	89	86	03	92	87	05
CVCC	94	84	10	87	86	01	92	86	06
CCVCC	93	82	10	89	85	04	93	87	06

Table 5: Proportions of the estimates of the rate of correct answers in the fixed part. Hi = high biphone frequency and Lo = low biphone frequency, Δ = difference between high and low biphone frequency.

Reaction Times

RTs were analyzed using a mixed-effect linear regression model with participants and targets as random factors (Quené & van den Bergh, 2008). The fixed factors were biphone frequency (high/low), syllable structure (CVC, CCVC, CVCC, CCVCC), native language (Dutch, Spanish, Japanese), response type (correct/incorrect) and the interactions of these factors. RTs from items that produced correct responses by Dutch participants on low biphone frequency CVC items were coded as the intercept. The factors LoR and C-test scores were taken out in the subsequent analysis, as they did not reach significance in a prior analysis. The average reaction times are given in Table 6. Recognition latencies of the most complex syllable structure CCVCC was overall significantly slower than for CVC (β = 66.27, (*SE* = 33.14), $p < 0.05$), hence the positive co-efficient β. Furthermore, there was a significant main effect of response type (correct/incorrect). Erroneous responses were 279.74 ms slower than correct responses (β = 279.74, (*SE* = 40.20), $p < 0.001$). Response type interacted with biphone frequency. In the correct responses, estimates of the fixed effects revealed that recognition was facilitated for high biphone frequency items (β = -69.02, (*SE* = 32.85), $p < 0.05$, see Figure 2). In the errors, there was a tendency

to the reverse of this effect (see Table 6 and Figure 3). Moreover, there was an interaction of syllable structure and L1. The Spanish were slower than Japanese and Dutch participants in recognizing CVC syllables (β = -66.25, (SE = 28.71), $p < 0.05$).

		Correct		**Incorrect**	
		Hi	**Lo**	**Hi**	**Lo**
Japanese	**CVC**	1142,36 (47,62)	1193,51 (47,78)	1417,34 (64,75)	1406,40 (59,65)
	CCVC	1139,77 (47,60)	1184,85 (47,82)	1654,21 (65,09)	1517,22 (58,71)
	CVCC	1157,19 (47,53)	1202,80 (47,83)	1493,18 (71,03)	1598,08 (58,99)
	CCVCC	1189,87 (47,58)	1249,85 (47,93)	1434,89 (68,57)	1468,95 (56,84)
Spanish	**CVC**	1162,45 (44,23)	1202,49 (44,38)	1585,13 (56,78)	1497,32 (53,46)
	CCVC	1130,59 (44,24)	1168,81 (44,29)	1471,23 (59,05)	1462,71 (55,45)
	CVCC	1136,77 (44,27)	1194,34 (44,30)	1472,89 (56,48)	1511,96 (55,69)
	CCVCC	1186,33 (44,17)	1202,51 (44,32)	1472,67 (59,51)	1518,64 (54,46)
Dutch	**CVC**	1091,64 (44,73)	1117,34 (44,88)	1424,95 (61,21)	1397,08 (56,64)
	CCVC	1085,99 (44,71)	1118,56 (44,83)	1421,86 (62,79)	1352,71 (57,06)
	CVCC	1096,81 (44,71)	1148,60 (44,88)	1568,39 (63,67)	1417,27 (56,09)
	CCVCC	1114,59 (44,67)	1183,61 (44,84)	1387,70 (65,48)	1527,80 (56,24)

Table 6: Reaction time means in ms and standard errors in brackets measured from the target onset given for Correct and Incorrect answers for each High (Hi) and Low (Lo) biphone frequency in responses by Japanese, Spanish and Dutch participants.

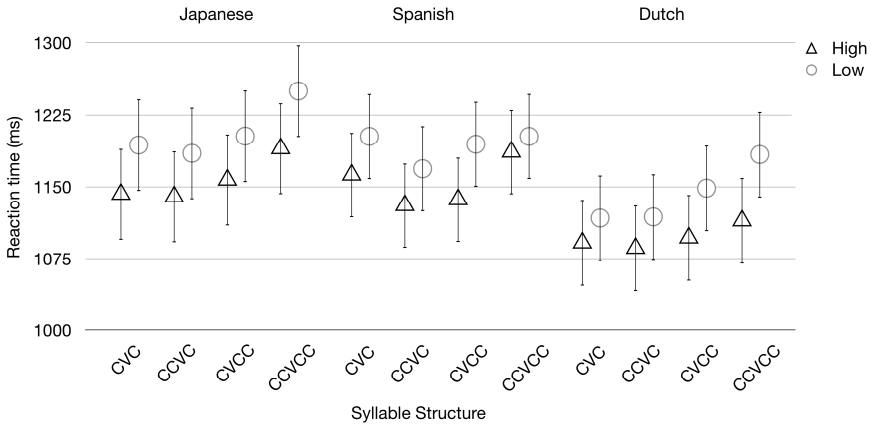

Figure 2: Averages of estimated means and standard errors of RT in ms for each syllable structure (CVC, CVCC, CCVC and CCVCC) for each high (H) and low (L) biphone frequency nonwords within the correct responses.

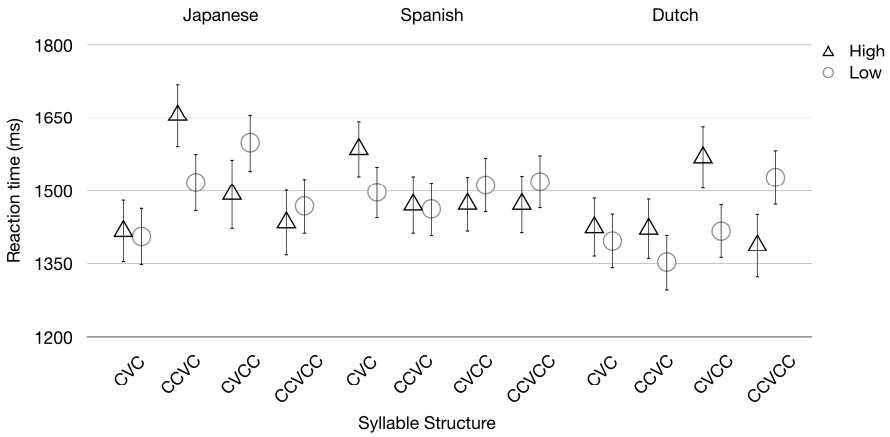

Figure 3: Averages of estimated means and standard errors of RT in ms for each syllable structure (CVC, CVCC, CCVC and CCVCC) for each high and low biphone frequency nonwords within the erroneous responses.

Further estimates of the fixed effects revealed an interaction of syllable structure, L1 and answer type. The interaction occurred in CVCC and CCVC items, which were in their erroneous answers recognized significantly slower by Japanese than Dutch or Spanish participants (CCVC: β = 165.08 (*SE* = 81.21), p = 0.005; CVCC: β = 193.46 (*SE* = 81.00), $p < 0.05$). Moreover, there was an interaction of syllable structure, L1, answer type and biphone frequency. Whereas RTs in the erroneous responses were generally faster in low than in high biphone frequency items, this effect was reversed when Japanese (β = -270.69 (*SE* = 128.95, $p < 0.05$) and Spanish (β = -258,68 (*SE* = 116.67), $p < 0.05$) reacted to CVCC items.

Discussion

The results represent the first evidence that L2 probabilistic phonotactic knowledge is acquired not only of syllable structures that are licit but even of structures that are illicit in the L1, and that this knowledge might facilitate lexical acquisition. As predicted, both groups of L2 learners were more accurate in recognizing high than low biphone frequency nonwords. This replicates findings of previous production-based STM experiments with L2 learners (e.g., Majerus, et al., 2008). The learners did not rely less on biphone frequency than the native speakers of Dutch. However, there was an effect of L2 proficiency: Recognition of high biphone frequency items correlated with proficiency scores, indicating that a reliance on biphone frequency in lexical acquisition improves with proficiency. This supports previous findings of a development of phonotactic knowledge in an L2 (e.g., Trapman & Kager, 2009).

Studies on STM performance predominantly use accuracy as the only dependent measure. For the current study, we measured RTs in addition to accuracy (just as (Sternberg, 1966), since the speed of processing may reveal factors facilitating or inhibiting STM recognition. The RT data in the current study suggests interesting effects that are in line with the predictions. This is not only the case with RTs in the correctly recalled items but also with RTs in the incorrect responses, in which participants had failed at holding a nonword in the phonological loop until they were supposed to recall it. Though RTs in general and RTs of erroneous responses in particular are not commonly analyzed and interpreted in studies on working memory, the data should not be withheld and will be tentatively interpreted below.

Biphone frequency affected recognition latencies in interaction with response type. As for the correct responses, participants were faster at reacting to high biphone frequency nonwords than to low biphone frequency nonwords, indicating that processing is facilitated by high biphone frequency. As for the erroneous responses, this effect was reversed: participants were faster at

erroneously rejecting low than high biphone frequency nonwords. This result can be interpreted as follows. If participants have to remember a nonword composed of low frequency biphones, it decays faster since it receives less support from LTM representations for reconstructing degraded memory traces during rehearsal in STM than nonwords composed of high frequency biphones. Hence, by the time the participant is required to make a decision, the nonword is likely to have already dropped out of memory. This is reflected in a relatively faster error.

In our study, this result is obtained for native speakers as well as L2 learners. As predicted, recognition performance was higher for simple than complex syllables. Recognition of the most complex syllable structure type used in the task, namely CCVCC syllables, was significantly slower than of simpler CVC and CCVC syllables. This supports previous findings of effects of increasing complexity on recall (Service, 1998). We predicted that recognition of syllable structures that are not attested in the L1 would be impeded. There was an interaction of L1 with syllable structure and response type. However, against our predictions, the L2 learners' RTs of the correct responses were native-like and not affected by whether the syllable types were attested in the L1. Still, the Spanish were slower than Dutch and Japanese in recognizing CVC items. This, however, cannot be explained by an influence of L1 structural knowledge. It rather seems to be a lexical neighborhood density effect. Dense lexical neighborhoods have been shown to inhibit nonword repetition in prior studies (e.g., Goh & Pisoni, 2003; Vitevitch & Luce, 1999). Obviously, the CVC items had denser lexical neighborhoods than the complex syllables in Spanish, and probably they had higher densities in Spanish than in Japanese and Dutch.

The predicted interaction of L1 with syllable structure showed up in the RTs in erroneous rejections of syllable types that are illicit in the learners' L1. Japanese participants were slower at erroneously rejecting CVCC and CCVC than CCVCC syllables. This indicates that they receive less support from LTM representations for keeping the most complex CCVCC items in memory, which consequently decay and drop out of the phonological loop faster than syllables of less complexity.

Furthermore, syllable structure interacted with L1 and biphone frequency in the RTs of the erroneous responses: While error latencies for low biphone frequency items are usually faster than those for high biphone frequency items, Spanish and Japanese L2 learners were slower at erroneously rejecting the low biphone frequency CVCC items. That suggests that for this L1 unattested syllable structure type, participants did not receive the additive facilitatory effect from L2 high biphone frequency for keeping information in the phonological loop. In contrast, the Japanese displayed very long RTs when trying to maintain erroneously rejected high biphone frequency CCVC items. This suggests that although the syllable type is unattested in the Japanese learners' L1, biphone frequency facilitated memory of CCVC nonwords.

It is interesting that that L1 structural knowledge does not influence L2 nonword recognition of all L1 unattested structures, but merely in the most complex types CVCC and CCVCC. This might indicate that L2 syllable structure can well be acquired at high L2 proficiency, possibly in a similar order as found in children (e.g., Levelt & Vijver, 1998/2004) from CCVC to CVCC to CCVCC. Note furthermore that error RTs predominantly stem from weaker participants, who make more errors. Hence, although we did not find an interaction of L2 proficiency and syllable structure in the RTs, it might still be that the RTs in the erroneous responses predominantly stem from learners who are less proficient in Dutch.

In contrast to previous L2 studies on STM recognition, the stimuli were controlled for L2 lexical neighborhood density. This indicates that facilitatory effects in this experiment come from a sub-lexical level. Ergo, L2 learners seem to rely on L2 sub-lexical representations they use in a native-like manner in the process of subvocal rehearsal. It needs to be kept in mind that this result might be due to the high proficiency of the L2 learners in our study. This would also explain why we did not replicate a familiarity effect as found in previous studies (e.g., Hulme, et al., 1991).

6.3. General discussion

The central hypothesis of this paper was that sub-lexical representations are two-fold with representations of probabilistic phonotactics on the one hand and hierarchical phonological structure on the other hand. It has been shown in the literature that when holding a nonword, STM draws on both probabilistic phonotactic knowledge (e.g., Thorn & Frankish, 2005) and structural knowledge (Boll-Avetisyan, to appear; Treiman & Danis, 1988). The hypothesis of the two different knowledge components entails that they can be acquired independently. L2 learners are known to have difficulties acquiring phonological structures that are unattested in the L1 (e.g., Broselow, 1984). Therefore, it was hypothesized that L2 learners would be impeded in memorizing syllable types which are unattested in their L1. Still, we hypothesized, they should be able to acquire L2 probabilistic phonotactics represented as *n*-phone patterns that should be unaffected by structural knowledge, if structural knowledge and probabilistic phonotactics are two separate knowledge components. A probed recognition experiment with nonwords manipulated for biphone frequency and syllable complexity was carried out in order to test this hypothesis.

As expected, in a Sternberg task, STM nonword recognition was affected by L2 phonotactic probability in recognizing simple but also complex syllables that are unattested in the learners' L1. This suggests that learners acquire L2

probabilistic phonotactics not only for structures that are attested in the L1 (e.g., Majerus, et al., 2008; Trapman & Kager, 2009), but also for new structures. The fact that the learners were affected by probabilistic phonotactics in syllables that are illicit and therefore have zero frequency in their L1 minimizes the possibility that recognition was aided by L1 transfer. As STM for L2 nonwords and L2 lexical acquisition are related (e.g., Cheung, 1996; Masoura & Gathercole, 1999; Service & Kohonen, 1995; Speciale, et al., 2004), this suggests that L2 probabilistic phonotactic knowledge can facilitate the acquisition of words of all syllable structures, even if they are illicit in the L1. In contrast to previous studies, L2 lexical neighborhood density was controlled for, which is a strong indication that L2 learners rely on L2 sub-lexical representations they use in a native-like manner in the process of subvocal rehearsal. This supports theories of a sub-lexical involvement in lexical acquisition (Gathercole, et al., 1999), even though there might be further measures beyond lexical neighborhood density which could also cause effects from lexical analogy.

Regarding syllable complexity, the L2 learners' error rates and RTs in the correctly recalled items were not affected by L1 structural knowledge. Given the literature about L2 difficulties with acquiring L2 structural knowledge, this seemed surprising. However, when taking into account RTs on items that were erroneously rejected, the picture slightly changes. Here, the learners were not native-like in performance on the most complex syllable types that are not attested in the L1. This result supports the hypothesis of independent sub-lexical representations of structural and probabilistic phonotactic knowledge. Knowledge of the phonological structure of a language has an impeding effect on verbal STM recall. Hence, L1 transfer of structural knowledge can hamper L2 lexical acquisition. However, independently of this, L2 probabilistic phonotactic knowledge including biphones that are structurally illicit in the L1 can still facilitate L2 lexical acquisition.

The fact that L1 structural knowledge only affected the response latencies of the erroneously rejected items but not the RTs of correctly recognized items or error rates is in line with the finding by Storkel and colleagues, who investigated the separate influences of lexical and sub-lexical factors on word learning in adults (Storkel, et al., 2006). They found facilitatory effects of high neighborhood density on word learning, but an inhibitory effect of high phonotactic probability. The effect of phonotactic probability, however, only occurred among items that were partially remembered, i.e., among items in which only two out of three phonemes were correctly reproduced. The authors conclude that novel words with low phonotactic probability might stand out in speech and therefore be relevant in the initial stages of word learning. Later, during the integration of novel forms into the lexicon, lexical representations might be of higher importance.

It is possible that the predicted effects would show up if less proficient learners were tested. The reason for testing highly proficient learners was that they have

been shown to rely more on LTM knowledge during recall than on general STM capacities (Speciale, et al., 2004). However, there is an additional counter-effect that might have played a role in this study: The higher the proficiency in a language, be it the L1 during development as a child (Gathercole, Willis, Emslie, & Baddeley, 1992) or the L2 (Cheung, 1996), the more the strong link between lexicon size and phonological STM diminishes. Ergo, the adult native speakers and advanced L2 learners in the current study might have been too proficient to reveal differences in recognition performance depending on item properties. Hence, it might be worthwhile to run a similar experiment with less advanced learners. Importantly, the highly proficient L2 learners tested in this study turned out to use L2 knowledge of simpler L1 illicit structures for nonword recognition, suggesting that L2 syllable structure can be well-acquired at high L2 proficiency, possibly in a similar order as found in children (e.g., Levelt & Vijver, 1998/2004) from CCVC to CVCC to CCVCC.

It remains to be seen how the two different sub-lexical knowledge components interact in processing during STM rehearsal and lexical acquisition. Possibly, first sub-lexical knowledge of probabilistic phonotactics is accessed. Once a biphone is activated, it makes reference to structural representations. If the activated biphone is structurally licit, it will receive support from the structural knowledge component. If biphones reside in complex syllable constituents, these receive stronger positive feedback from the structural knowledge component, since there are structurally more constrained than if they straddle constituent boundaries. In L2 nonword recall, biphones that are structurally illicit in the L1 will not receive support from structural knowledge LTM. Therefore, structurally illicit syllables decay faster, yielding shorter RTs in the errors. It will be interesting to study this in more detail in the future.

We want to speculate how L2 learners acquire probabilities of biphones that are illicit in their L1. It needs to be kept in mind that L2 learners might apply perceptual repairs to biphones that are illicit in their L1. One account is that L2 learners acquire epenthesized triphones. If that is the case, then their representations of probabilistic phonotactics may be flawed. For example, if they do not represent a biphone /pr/ and its frequency of occurrence, but create a repaired epenthesized entry of a triphone such as /pur/, then the probability of the cluster will be merged with the actual (L1 and) L2 probability of this specific triphone, leading to incorrect entries. A second possibility is that the L1 phonological filter is only radically active at the onset of L2 acquisition, but will be weakened over time with increasing L2 proficiency. A weakened phonological filter may let some L1 illicit structures pass through, and pave the way for an acquisition of their L2 probabilities. A third option is that biphone probabilities represented at a shallow level close to acoustic events and their probabilities are acquired independent of grammatical knowledge. It will be

interesting for future studies to address the question of which of the three possibilities is the most likely.

6.4. Conclusion

To sum up, while the L2 learners' accuracy rates and RTs in the correctly recognized items were not affected by L1 structural knowledge, in RTs of erroneous rejections, learners were not native-like in recognizing syllable types that are not attested in the L1. Hence, L1 transfer of structural knowledge can hamper L2 lexical acquisition. Still, independently of this, L2 probabilistic phonotactic knowledge including biphones that are structurally illicit in the L1 can facilitate L2 lexical acquisition. The current study adds evidence to theories of LTM contributions in working memory, but crucially the results indicate that the LTM knowledge affecting STM is more complex than previously assumed. It suggests that there are two separate sub-lexical knowledge components of probabilistic phonotactics and structural phonological knowledge that may be acquired independently.

6.5. References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philedelphia: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighbourhoods? *Journal of Memory and Language 44*, 568-591.

Boersma, P. (2000). Learning a Grammar in Functional Phonology. In J. Dekkers, F. R. H. Van der Leeuw & J. M. Van der Weijer (Eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford: Oxford University Press.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.09). Retrieved June 24, 2007, from http://www.praat.org/.

Boll-Avetisyan, N. A. T. (to appear). Probabilistic Phonotactics in Lexical Acquisition: The Role of Syllable Complexity. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer & B. Kühnert (Eds.), *Consonants and structural complexity*: Mouton de Gruyter, Interface Explorations Series.

Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics in Language Teaching, 22*, 253-269.

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

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*(89), 183-213.

Cowan, N., Wood, N. L., Nugent, L. D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposing effects of duration and complexity. *Psychological Science, 8*, 290–295.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1355–1367.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77-91.

Eckes, T. (2010). Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In R. Grotjahn (Ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-test: Contributions from current research* (pp. 125-192). Frankfurt am Main: Lang.

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory" capacity just another name for word knowledge? *Journal of Educational Psychology, 82*, 799-804.

Fikkert, P. (1994). *On the acquisition of prosodic structure*. Dordrecht: ICG.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. . *Language Acquisition, 47*, 27-52.

Fudge, E. C. (1969). Syllables. *Journal of Linguistics, 5*, 253-287.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513-543.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84-95.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887-898.

Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly Journal of Experimental Psychology Section A, 56*, 929-954.

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language, 51*, 586-603.

Harris, J. W. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685-701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition., 23*, 1217-1232.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell. Reprint (1969). Frankfurt am Main: Suhrkamp.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech, 32*, 249-278.

Kuperman, V., Ernestus, M., & Baayen, R. H. (2008). Frequency Distributions of Uniphones, Diphones and Triphones in Spontaneous Speech. *Journal of the Acoustical Society of America, 124*(6), 3897-3908.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.

Lentz, T. O., & Kager, R. W. J. (submitted). *Categorical and Gradient Phonotactics Are Represented Differently: Evidence from Word Recognition*.Unpublished manuscript.

Levelt, C., & Vijver, R. v. d. (1998/2004). Syllable types in cross-linguistic and developmental grammars. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Fixing Priorities: Constraints in Phonological Acquisition* (pp. 204-218). Cambridge, England: Cambridge University Press.

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology, 53*(A), 1-22.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Doctoral dissertation, Indian University, Bloomington, IN.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3*, 210-227.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107*, 395-419.

Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language, 51*, 297-306.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*, 383-388.

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effects in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. . *Journal of Experimental Child Psychology, 105*(4), 306-323.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge MA: MIT press.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413-425.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review, 111*(1), 159-182.

Roodenrys, S., & Hinton, M. (2002). Sublexical or Lexical Effects on Serial Recall of Nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 29-33.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall: short-term memory. *Memory and Cognition, 21*, 168-175.

Selkirk, E. O. (1982). The Syllable. In H. v. d. Hulst & N. Smith (Eds.), *The Structure of Phonological Representations, Part II* (pp. 337–383). Dordrecht: Foris.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51*(A), 283-304.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155-172.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics, 25*, 293-321

Sternberg, S. (1966). High speed scanning in human memory. *Science, 5*, 653-654.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321-1337.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotacic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*, 1175-1192.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory and Cognition., 31*(4), 729-735.

Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition, 16*, 178-221.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*, 49-74.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 145-152.

Trommelen, M. (1983). *The syllable in Dutch; with special reference to diminutive formation*. Dordrecht: Foris.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighbourhood Activation in Spoken Word Recognition. *Journal of Memory and Language, 40*, 374-408.

Appendix A

Target Stimuli.

High biphone frequency nonwords.

CVC: beːl, bɛx, deːf, deːk, foːm, foːt, xaːk, hɑs, jaːt, kɑx, laːm, meːf, mɛl, neːk, raːl, raːn, rɪn, rɔf, roːn, sɛŋ, tɛx, tɛŋ, voːk

CCVC: blɑs, brɛx, brɛl, brɪn, broːn, dreːk, froːm, froːn, froːt, xraːk, xroːn, klɑx, klɑr, krɛx, krɔf, praːn, prɪn, slaːm, slaːr, traːl, trɛŋ, trɪn, twɪl

CVCC: beːls, beːrk, bɛxt, deːft, deːks, xaːkt, hɑst, jaːrt, kɑrk, liːnt, liːts, meːft, mɛls, neːks, rɑls, rɛns, rɔxt, sɛŋk, sɛrt, tɛxt, tɛkt, tɔrm, voːkt

CCVCC: blɑst, bleːrk, brɛlt, brɪnt, dreːks, frɑls, frɛns, froːns, xraːkt, klɑrm, kliːnt, krɛxt, pliːts, prɪŋk, slaːrs, stɛŋk, stɛrt, traːls, trɛŋk, trɔxt, trɔrm, twɛkt, twɪlt

Low biphone frequency nonwords.

CVC: bɪr, dyl, hyl, kiːŋ, kux, kym, lux, lum, lʏt, mʏp, myt, ryk, sum, sur, tɪr, vʏŋ, wiːŋ, wur, zaŋ, zus, zʏp, zyx, zyl

CCVC: dwiːŋ, dwuŋ, dwur, dwʏŋ, dwyw, flux, xlun, xlʏt, knux, knʏm, knʏŋ, knʏp, kwiːŋ, smyx, smyt, snum, vlum, vluŋ, wryk, zwʏŋ, zwʏp, zwyx, zwyl

CVCC: bɪrx, dylk, hylm, kiːŋs, kuŋs, kymp, lump, lurx, lʏmp, lʏmt, mʏps, nurn, sums, surp, tɪrf, viːŋt, vʏŋt, vʏrf, wumt, wuŋt, wurn, zaŋt, zylm

CCVCC: dwiːŋt, dwumt, dwuŋt, dwurx, dwʏŋt, dwʏrf, dwywt, xlump, xluŋt, xlʏmp, knʏms, knʏŋs, knʏps, snump, snurn, snurp, vluŋt, vlʏmt, xlʏps, vlyms, vlʏps, dwyŋt, xlumt

Appendix B

Estimates of the rate of correct answers in the fixed part in proportions (Prop) and logits with standard errors in brackets.

		High biphone frequency		Low biphone frequency	
frequency		**Prop.**	**Logit**	**Prop.**	**Logit**
Japanese	**CVC**	0.91	2.30 (0.14)	0.86	1.74 (0.13)
	CCVC	0.91	2.35 (0.15)	0.85	1.70 (0.12)
	CVCC	0.94	2.70 (0.17)	0.84	1.68 (0.12)
	CCVCC	0.93	2.52 (0.16)	0.82	1.50 (0.12)
Spanish	**CVC**	0.87	1.90 (0.12)	0.83	1.60 (0.10)
	CCVC	0.89	2.06 (0.13)	0.86	1.83 (0.12)
	CVCC	0.87	1.86 (0.12)	0.86	1.86 (0.12)
	CCVCC	0.89	2.11 (0.13)	0.85	1.74 (0.11)
Dutch	**CVC**	0.91	2.28 (0.13)	0.87	1.90 (0.12)
	CCVC	0.92	2.39 (0.14)	0.87	1.94 (0.12)
	CVCC	0.92	2.44 (0.14)	0.86	1.85 (0.12)
	CCVCC	0.93	2.55 (0.15)	0.87	1.89 (0.12)

CHAPTER 7

GENERAL CONCLUSION

7.1. Summary

This dissertation addressed phonotactic knowledge as it is put to use in speech segmentation and lexical acquisition. The focus was on phonotactic knowledge that requires representations of a certain degree of complexity and abstraction, while investigating how such knowledge is acquired.

In this section, it will be described how these issues were addressed in the individual chapters, and what the main findings of the studies were. First, §7.1.1. will summarize and interpret the findings regarding the role of segment co-occurrence constraints in speech segmentation, and §7.1.2. will similarly do so for the findings regarding the role of syllable structure and probabilistic phonotactics in lexical acquisition. In the following section (7.2.), it will be discussed how these findings contribute to the fields of phonology and developmental psychology and address some of the open questions that, in my view, should deserve some attention in future studies.

7.1.1. Segment co-occurrence constraints for speech segmentation

Chapters 2, 3, and 4 investigated whether segment co-occurrence constraints can be used as cues for speech segmentation. More precisely, the three chapters dealt with the question of whether phonotactic constraints requiring non-identity between consonants can be used for segmentation. As a method, all three chapters used the artificial language learning paradigm: Infant and adult participants were familiarized with a continuous stream of an artificial language that contained no statistical, phonological, or lexical cues for segmentation other than the constraint under investigation. After familiarization, participants were tested on their segmentation preferences. The test items were trisyllabic sequences that all had occurred in the artificial language. Half of the test items were sequences that the participants should have segmented out of the stream,

whereas the other half should not have been segmented. In all chapters, participants were either native listeners or L2 learners of Dutch.

Chapter 2 focused on infants' knowledge of probabilistic segment co-occurrences as well as the question of whether the source of infants' phonotactic knowledge is continuous speech or the lexicon. It was argued that clues might come from finding out which segmentation strategy infants used: a trough-based (i.e., inserting boundaries at dips in phoneme transitional probability) or chunk-based segmentation strategy (i.e., binding phonemes that occur with a high transitional probability). A trough-based strategy was hypothesized to be preferred by infants based on the observation that it must be difficult for infants to acquire low probability phoneme co-occurrences from their small proto-lexicon.

Two experiments with infants were carried out. Experiment 1 used an artificial language which contained /s/-vowel-/s/ sequences. This sequence has a low probability of occurrence in Dutch infant-directed continuous speech, whereas it is over-represented in the infant-directed lexicon. 15-month-old Dutch infants located boundaries between two occurrences of /s/, demonstrating a preference for a trough-based segmentation strategy. In the second experiment, 9 and 15-month-olds were tested on whether they would use a chunking cue from /p/-vowel-/p/, which is overrepresented in both the infant-directed lexicon and continuous speech. However, the infants did not show any segmentation preference. These results support the hypothesis that infants acquire phonotactic segmentation cues from continuous speech rather than from the lexicon. It suggests that segmentation procedures in infancy predominantly rely on cues from low probability sequences for locating word boundaries.

In **Chapter 2**, it was argued that infants use a cue from an under-representation of two identical non-adjacent consonants (i.e., /s/-vowel-/s/). **Chapter 3 and 4** followed up on this by further investigating whether OCP-PLACE, a typologically recurrent abstract constraint on the co-occurrence of consonants with the shared feature [place], can be used as a cue for speech segmentation. As the Dutch lexicon is restricted by OCP-PLACE, this has the effect that sequences of consonants with shared [place] (e.g., /pVm/) are similarly under-represented in continuous speech, which would potentially make for an effective cue for trough-based segmentation.

Throughout this dissertation, it was hypothesized that phonotactic knowledge is acquired on a language-specific basis from the input, and hence the use of phonotactic knowledge is also language-specific. There are, however, functionalist accounts that assume OCP to be a result of a universal bias against the ability of separately encoding two homorganic consonants (Boersma, 2000; Frisch, 2004).

Chapter 3 showed that effects of OCP-PLACE on segmentation cannot be directly due to universal functional biases by demonstrating OCP's role in speech segmentation in a series of artificial language learning experiments with native speakers of Dutch. The Dutch lexicon is restricted by OCP-LABIAL, but not by OCP-CORONAL. In accordance with distribution, it was found that Dutch listeners use OCP-LABIAL, but not OCP-CORONAL as a cue for speech segmentation. This indicates that OCP is language-specifically correlated with phoneme distributions in the native language. Furthermore, in the artificial languages no real words occurred, while any cue for segmentation from lexical distributions was controlled for. This shows that OCP is useful for finding word boundaries by itself without need for top-down feedback from the lexicon. These findings cannot be accounted for by functionalist theories, which would instead predict that OCP-violating sequences cannot be faithfully perceived (Frisch, 2004) or cannot be perceptually separated as being part of two words (Boersma, 2000), unless top-down feedback from the lexicon (Frisch, 2004) or other phonological cues (Boersma, 2000) indicate a word boundary.

Chapter 4 followed up on this topic by further investigating whether knowledge of OCP-PLACE that informs speech segmentation is due to universal functional biases or, as hypothesized, to language-specific acquisition. First, native speakers of Mandarin Chinese, a language for which the native lexicon is not restricted by OCP-PLACE, were tested on their segmentation preferences. As opposed to the native Dutch listeners tested in Chapter 3, the Mandarin Chinese listeners did not use OCP-PLACE for segmenting an artificial language, as would be predicted if OCP were due to a universal functional bias.

From these results it follows that OCP-PLACE must be acquired. This hypothesis was tested more directly by means of L2 learners of Dutch. With respect to L2 learning, it was hypothesized that the received amount of speech input spoken by native speakers of the target language is crucial for the acquisition of OCP. Experiments were carried out with three groups of L2 learners: advanced Mandarin Chinese L2 learners of Dutch, of whom the first group lived in the Netherlands, and the second group lived in China, and a third group of beginning Dutch learners who lived in the Netherlands. The results confirmed the need for sufficient native spoken input in phonotactic L2 acquisition. The pre-requisite for the acquisition of OCP-PLACE as a speech segmentation cue in an L2 is two-fold: the learner has to be proficient and has to have been exposed to a sufficient amount of input spoken by native speakers.

7.1.2. Syllable structure constraints for lexical acquisition

Previous studies on the role of the sub-lexical level in nonword recognition have solely focused on one sub-lexical component, namely *n*-phone representations of probabilistic phonotactics. The relevance of structural phonotactic knowledge

has been rather neglected. Furthermore, previous studies had mostly used simple CVC stimuli only, falling short of offering a comprehensive view of the role of sub-lexical phonotactic knowledge in lexical acquisition, since in many languages, words of more complex syllable structures have to be learned in addition to CVC words.

Chapters 5 and 6 addressed the question of how knowledge of syllable structure would modulate the facilitatory effect of probabilistic phonotactics in lexical acquisition. As a task, a short-term memory recognition paradigm was used. Performance in nonword recall tasks has often been found to be correlated with lexicon size in adults and children in their native language (L1) (for an overview see Gathercole, 2006) and foreign language (L2) (e.g., Service & Kohonen, 1995). This has been taken as evidence that short-term memory recall performance and lexical acquisition are related. In fact, nonword recall can be seen as the initial step in storing new words in the mental lexicon. The easier it is to hold an item in short-term memory, the easier it is to store it in long-term memory. In both chapters, the same experiment and the same nonword stimuli manipulated for biphone frequency and syllable structure were used.

Chapter 5 investigated the hypothesis that knowledge of structural markedness constraints as part of a phonological grammar modulates the effects from phonotactic probability on lexical acquisition. In line with this hypothesis, I found in two experiments with monolingual Dutch adults that the facilitatory effect of biphone frequency was significantly enhanced when complex syllables were memorized than when nonwords had a simple syllable structure. The results suggest that both knowledge components—structural constraints and probabilistic phonotactics—play a role in lexical acquisition. Importantly, one knowledge component modulates the effect of the other: knowledge of syllable structure can boost the facilitatory effect of high probability phonotactics, and, conversely, put extra weight on the inhibitory effect of low probability phonotactics.

Chapter 6 followed up on this issue by testing the hypothesis that separate storage of the two knowledge components would necessitate that they can be acquired independent of each other. This question was investigated by testing the effects of biphone frequency and syllable structure on L2 lexical acquisition. Participants were L2 learners whose native languages allowed complex maximal syllable templates (Japanese: CVC only; and Spanish: CCVC) than the target language Dutch (CCCVCCC). The L2 learners experienced facilitation of L2 biphone frequency in recognizing nonwords even if they were structurally more complex than their L1 allows. Still, they were faster at making errors in some, i.e., very complex, syllable types that were not attested in their L1. Thus, there was benefit from L2 probabilistic phonotactics in L2 short-term memory recognition despite transferring L1 structures. These results support the

hypothesis that *n*-phone probabilities can be acquired independently of the structure in which they occur.

7.2. Contributions

This dissertation addressed two major issues. First, it was shown that complex phonotactic representations are used for encoding the speech signal (see 7.2.1.). Second, with respect to the acquisition of phonotactics, some findings shed light on the role of learning mechanisms and the source of phonotactic learning (see 7.2.2.). This section will summarize the main conclusions and discuss some issues that remain for future research.

7.2.1. Phonotactic representations and their use

The introduction of this dissertation emphasized that in language acquisition, a successful speech segmentation performance is crucial in order to be able to start building up a lexicon. It was noticed that there is an interesting connection between segmentation and lexical acquisition in that phonotactic knowledge has been found to facilitate both. It was furthermore noticed that previous studies on the use of phonotactics in speech segmentation and lexical acquisition had primarily looked into the role of phonotactic knowledge of *n*-phone probabilities in speech segmentation and lexical acquisition.

This dissertation—besides replicating evidence for a role for phonotactic knowledge of *n*-phone probabilities—showed that phonotactic knowledge of a more abstract and complex nature is also used for speech segmentation and lexical acquisition. This was shown for two types of structural knowledge: constraints on segmental structure requiring non-identity between non-adjacent phonemes (Chapters 2, 3, and 4), and syllable structure (Chapters 5 and 6).

Chapters 2, 3, and 4 showed that infants as well as monolingual and L2 adults possess implicit knowledge of an under-representation of homorganic consonant pairs, and use this knowledge as a cue for the segmentation of an artificial language that is controlled for statistical, phonological, and lexical cues for segmentation. The new finding is that not only constraints favoring featural identity, but also constraints disfavoring featural identity can be used for speech segmentation. This finding would not be predicted by current functionalist theories, which propose that the typological recurrence of OCP is grounded in a universal bias intrinsic to perception, which disfavors the full and separate encoding of two neighboring homorganic consonants (Frisch, 2004; Boersma, 2000). With respect to segmentation, functionalist theories would predict that

listeners can only interpret sequences of homorganic consonants as belonging to two separate elements if additional information—ideally top-down information from the lexicon—is drawn upon. This prediction has been proven incorrect in this dissertation, as listeners use OCP as a cue even when parsing an artificial language from which lexical cues had been eliminated.

The results of Chapter 2 may suggest that OCP-PLACE makes for an ideal cue for segmentation: by banning sequences of consonants with shared [place], OCP-PLACE has the effect that these sequences are *under*-represented in continuous speech, too. We could learn from Chapter 2 that cues from under-representations are especially useful in speech segmentation: infants used knowledge of under-represented but not over-represented phoneme pairs for speech segmentation. These results suggest that infants favor trough-based instead of chunk-based segmentation strategies. This finding has important implications as to how we think of informativeness of phonotactic knowledge. Continuous speech contains fewer troughs than peaks in transitional probability, because speech is composed of words. In continuous speech, more phoneme pairs occur within words than across words, because typically words consist of more than two or three phonemes. Phoneme pairs that occur within words tend to be over-represented. In light of this, it seems logical that it is more useful to attend to the relatively few troughs than to the much larger number of peaks in transitional probability when segmenting speech.

Chapters 5 and 6 particularly assessed what type of phonotactic knowledge may be useful in lexical acquisition. Previous studies that found facilitatory effects of sub-lexical knowledge on lexical acquisition had only looked at the role of probabilistic phonotactics on memorizing novel CVC words. However, in many languages, words of more complex syllable structures have to be learned in addition to CVC words.

It was demonstrated in Chapter 5 and 6 that when memorizing nonwords of a complex syllable structure, structural knowledge comes into play: the facilitatory effect of high probability phonotactics on lexical acquisition increased with increasing syllable complexity, and vice versa. This suggests that two sub-lexical knowledge components representing phonotactics—probabilistic phonotactics on the one hand and structural knowledge of abstract symbolic constraints on the other—play a role in lexical acquisition. More specifically, the phonological grammar representing structural knowledge modulates the effect of probabilistic phonotactics.

In sum, we found converging evidence that the facilitatory use of sub-lexical knowledge in speech segmentation and lexical acquisition is not restricted to *n*-phone probabilities. Abstract structural representations also play a role.

A question regarding probabilistic phonotactic knowledge that remains open at this point is that it might be necessary to distinguish between phonotactic cues

that are usable in speech segmentation and phonotactic knowledge that facilitates processing in other domains such as lexical recognition or lexical acquisition. In this dissertation, we found some evidence that infants use knowledge of an under-represented phoneme pair as a cue for speech segmentation. It was discussed in Chapter 2 that this does not imply that infants do not have knowledge of overrepresented phoneme pairs, too (see, for example, the study by Nazzi, Bertoncini, & Bijeljac-Babic, 2009, who found knowledge of an overrepresented pattern in infants).

I would like to raise attention to the possibility that knowledge of under- and over-represented phonotactic patterns serves different functions in speech processing. It might be that under-representations are more usable in speech segmentation, whereas in lexical acquisition, a clear facilitation comes from knowledge of over-represented patterns. To some extent, the data from the L2 learners in Chapter 6 may indicate the following division of labor holds: memory for nonwords of high biphone probability improved with increasing L2 proficiency of the learners; however, L2 proficiency had no effect on how well nonwords of low biphone probability were remembered. A similar development was reported in a recent study with monolingual and bilingual 4-6 year-old children: over time, children increasingly relied on cues from high phonotactic probability, but not from low phonotactic probability in a nonword recall task (Messer, Leseman, Boom & Majo, submitted). However, to my knowledge, no study as yet has directly compared the functions of knowledge of phonotactic under- and over-representations. It will be interesting for future studies to address this issue.

7.2.2. Phonotactics and its acquisition

Regarding the acquisition of phonotactics, it was looked at whether some abstract phonotactic constraints which are typologically recurrent are directly due to functional biases on perception or whether their use is language-specific, and, hence, acquired from the input—possibly by an innate cognitive biases on acquisition. Here, it was hypothesized that phonotactic knowledge as used in speech segmentation and lexical acquisition is language-specific and hence acquired. Following up on this, a second research question was whether the source of phonotactic learning is the lexicon or continuous speech. Experiments were carried out with language-learning infants and adult L2 learners.

Learning mechanisms

As discussed in the previous section, the results of the experiments reported in Chapters 2, 3 and 4 showed that knowledge of OCP as used in speech segmentation cannot be reduced to universal biases on perception. Instead, we found language-specific effects of OCP on processing: monolingual listeners

whose native language is not restricted by OCP do not use the constraint for processing. Furthermore, we found that the use of OCP-PLACE for speech segmentation can be acquired in an L2. This suggests that the constraint must be acquired from the input.

In light of the complexity of OCP and its typological recurrence, it may be assumed that there is a role for an innate learning mechanism beyond statistical learning that is at play when acquiring OCP-PLACE: a cognitive analytic bias on acquisition. Furthermore, in contrast to previous views (e.g., Clahsen & Muysken, 1989), it needs to be assumed that the influence of such a cognitive bias also facilitates L2 acquisition, since it was found that L2 learners can acquire knowledge of OCP. As discussed in more detail in Chapter 4, this is most likely a cognitive analytic bias on acquisition that draws learners' attention to similarity. It has long been established that humans have a preference for grouping similar elements (Wertheimer, 1923). Remember that it was demonstrated in Chapter 2 that infants already have knowledge of under-represented co-occurrences of identical phonemes. It may be that because of an innate gestalt principle of similarity, sequences of homorganic elements are more salient to the learner. The increased salience should hold for all sequences of homorganic elements, no matter whether they are over- or under-represented. This should make one consider the possibility that a bias for grouping similar elements may help learners extract knowledge of similar elements to be under-represented. It will certainly be interesting to investigate this issue in more detail in future studies in order to obtain a better understanding about the interplay of biases on acquisition and the use of linguistic knowledge.

Continuous speech as the source of phonotactic learning

Another important finding was that at least some phonotactic knowledge must be acquired from continuous speech rather than from the lexicon. In the experiments reported in Chapter 2, infants used an under-represented pattern—which they can only have acquired from continuous speech, as the same pattern was over-represented in the infant-directed lexicon—as a cue for speech segmentation.

It may be speculated that continuous speech is the source of phonotactic knowledge not only in an L1, as shown in Chapter 2, but also in an L2. The results of experiments with L2 learners in this dissertation revealed that L2 learners who had intensive exposure to speech of the target language spoken by native speakers are able to use L2 phonotactic knowledge for speech segmentation and lexical acquisition. The issue of the role of exposure to high quality speech was specifically addressed in Chapter 4. Here, it was shown that advanced L2 learners who have received little or no exposure to speech spoken by native speakers of the target language do not use L2 phonotactics, while

advanced learners that live in a country where the target language is spoken do. Importantly, the two groups of L2 learners only differed by the amount of exposure to the L2 they received. According to the results of a proficiency test, their L2 lexicons were comparable in size. The results suggest that exposure to native continuous speech input may be crucial for acquiring knowledge of a phonotactic constraint on segmental structure in both, L1 and L2 acquisition.

A question that remains open is whether knowledge of over-representations is also acquired from continuous speech. Chapter 2 cannot provide an answer to that question, as it was only found that infants do not use a cue from an over-representation for speech segmentation. This indicates that infants prefer a trough-based over a chunk-based segmentation strategy, but is not yet informative about their knowledge. Relatedly, in Chapter 4 evidence was found for the role of exposure in acquiring an under-represented pattern in an L2. This may suggest that also in an L2, phonotactic constraints are acquired from the continuous speech input. It may be interesting to test this in future studies.

Furthermore, it will be interesting to test whether only under-represented phonotactic patterns, or also over-represented phonotactic patterns are acquired from continuous speech in an L1 and an L2, and to investigate the role of exposure. Chapter 6 may give some indirect clues on this issue: L2 learners were found to have knowledge of L2 phonotactics, which they used for lexical acquisition. This chapter was mostly concerned with the question of whether probabilistic phonotactics can be acquired independently of structural knowledge. It was less concerned with the developmental trajectory in L2 phonotactic acquisition. However, the results revealed the following: the facilitatory effect of high phonotactic probability increased with L2 proficiency, while the effect of low phonotactic probability was stable throughout all learners. The L2 learners in Chapter 6 were all highly proficient and lived in the country where the target language is spoken.

7.3. Conclusion

This dissertation contributes to the debate of how infants and adults use sub-lexical cues for speech segmentation and lexical acquisition. It was shown that not only *n*-phone probabilities, but also abstract representations of segmental (featural) and prosodic structure are put to use. Structural knowledge can enhance the facilitatory effects of probabilistic phonotactics in processing. Phonotactic representations that are used for processing are language-specific. They can be acquired in an L1 and in an L2. From the study of language-learning infants, we found evidence for continuous speech to be the source of phonotactic learning. For a successful acquisition of phonotactic constraints in

an L2, learning conditions with much exposure to native spoken input are essential.

7.4. References

Boersma, P. (2000). The OCP in the perception grammar. *Rutgers Optimality Archive, 435*, 1-52.

Clahsen, H., & Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research, 5*, 1-29.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. M. Kirchner & D. Steriade (Eds.), *Phonetically based Phonology* (pp. 346-371). Cambridge: Cambridge University Press.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513-543.

Messer, M. H., Leseman, P. P. M., Boom, J., & Majo, A. Y. (submitted). Modeling growth in verbal short-term memory and phonotactic knowledge support: A longitudinal study with monolingual and bilingual pre-schoolers. Unpublished manuscript.

Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *Journal of the Acoustical Society of America, 126*, 1440-1144.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155-172.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psycologische Forschung, 4*, 301-350.

SAMENVATTING IN HET NEDERLANDS

Onderzoeksvragen

Talen staan geen willekeurige combinaties van fonemen binnen woorden toe. De manier waarop woorden intern georganiseerd zijn wordt in iedere taal beperkt door klankregels. Eerdere studies hebben aangetoond dat mensen betrouwbare kennis over de klankregels van hun eigen taal hebben. Deze kennis lijkt een belangrijke rol te spelen in spraakverwerking en wordt al vroeg aangeleerd: baby's blijken al kennis van klankregels te hebben voor ze hun eerste woorden hebben verworven. Zelfs tweedetaalleerders (L2-leerders) hebben kennis van de klankregels van de tweede taal, die beter wordt naarmate de vaardigheid in die taal toeneemt.

Dit proefschrift behandelt aspecten van de algemene vraag hoe kennis van klankregels wordt opgedaan, hoe het gerepresenteerd wordt en hoe het in spraakverwerking wordt toegepast, zowel in eerstetaalleerders, volwassen luisteraars, als (volwassen) tweedetaalleerders. De verwerving, representatie en het gebruik van klankregels zijn intrinsiek verbonden factoren. Representaties van klankregels moeten verworven worden om gebruikt te kunnen worden. Bovendien hebben mensen taalkundige representaties nodig om spraak te verwerken. Wat betreft het gebruik van klankregels is het bekend dat het belangrijke aanwijzingen geeft voor spraaksegmentatie en dat het gebruikt wordt om lexicale verwerving te ondersteunen. Zo horen we bijvoorbeeld de zin *Five women left* als een continue stroom: /faɪvwɪmənlɛft/. In deze uitspraak geven klankregels aanwijzingen over woordgrenzen, omdat noch /vw/ in /faɪvwɪmən/, noch /nl/ in /wɪmənlɛft/ mogelijk zijn aan het begin of het eind van een Engels woord. In aanvulling daarop is gebleken dat klankregels de verwerving van woorden door volwassen, baby's en L2-leerders ondersteunen. Het woord *sit* is bijvoorbeeld mogelijk makkelijker te leren dan het woord *these*, ook al zijn ze wat de overige structuur betreft tamelijk gelijk. Dit komt omdat *sit* een samenstelling is van de vaak voorkomende foneemparen /sɪ/ en /ɪt/, terwijl /ðiː/ en /iːz/ in *these* niet erg gewoon zijn. Al voor het lexicale stadium gebruiken baby's klankregels voor spraaksegmentatie en het leren van woorden; dat geldt ook voor volwassen L2-leerders.

Wat betreft het probleem van het leren van een lexicon lijkt er een intrinsieke relatie te bestaan tussen spraaksegmentatie en lexicale verwerving. Een lexicon kan niet verworven worden zonder te weten hoe spraak gesegmenteerd moet worden. Om die reden kijkt dit proefschrift naar het gebruik van klankregels

voor beide taken, spraaksegmentatie (in Hoofdstuk 2, 3 en 4) en lexicale verwerving (in hoofdstuk 5 en 6).

Dit proefschrift behandelt twee vragen over de verwerving van klankregels. De eerste vraag is of baby's en L2-leerders naast kennis van klankdistributies, maar ook structurele kennis kunnen verwerven en gebruiken voor spraaksegmentatie en het leren van woorden. Het is mogelijk dat aangeboren cognitieve voorkeuren de taallerende baby en de L2-leerder helpen om aanwijzingen van klankregels voor spraaksegmentatie in de zogenaamde input (de taal die de leerder waarneemt gedurende de leerperiode) te ontdekken. Het is echter ook mogelijk dat universele eigenschappen van de menselijke waarneming de structurele klankregels (deels) verklaren, wat het onnodig zou maken die klankregels te leren. De tweede vraag over het leren van klankregels betreft de bron van dit leren. De vraag is specifiek of klankregels uit het lexicon worden verworven, of uit continue spraak. Wat betreft de vorm van de representaties van klankregels onderzoekt dit proefschrift of het mogelijk is dat niet alleen de waarschijnlijkheden van klankcombinaties, maar ook abstracte kennis gebaseerd op fonologische structuur of fonologische klankkenmerken spraaksegmentatie en lexicale verwerving vergemakkelijken. Daarbij werd onderzocht of kennis van klankregels voor zowel veelvoorkomende als weinig voorkomende klankcombinaties wordt gebruikt.

Fonologische structuur van klankregels in spraaksegmentatie

De onderzoeksvragen worden in vijf hoofdstukken behandeld. De eerste drie hoofdstukken (2, 3 en 4) behandelen de vraag of klankregels van een bepaalde complexiteit, namelijk over niet-aangrenzende afhankelijkheden, gebruikt kunnen worden als aanwijzing voor spraaksegmentatie.

Alle drie hoofdstukken gebruiken als experimentele methode het kunstmatigetaalleren. Bij die methode luisteren baby's en volwassen proefpersonen eerst enige tijd naar een continue stroom van een kunstmatige taal, die geen statistische, fonologische, of lexicale aanwijzingen voor segmentatie bevatte, op de regel (beperking) die onderzocht werd na. Na die gewenningsfase worden de proefpersonen getest op hun segmentatievoorkeuren. De testitems zijn sequenties van drie lettergrepen die allemaal in de kunstmatige taal voorkwamen. De helft van de testitems zijn sequenties die de proefpersonen uit de stroom hadden kunnen segmenteren, als ze de veronderstelde klankregel zouden gebruiken voor segmentatie. De andere helft zou niet specifiek als eenheid in de stroom moeten worden waargenomen. De proefpersonen waren baby's, volwassenen met Nederlands als eerste taal, of volwassen L2-leerders.

Hoofdstuk 2 richt zich op de kennis van baby's van de waarschijnlijkheid van een overgang tussen fonemen en de vraag of de bron van die kennis het lexicon of continue spraak was. Om het antwoord te vinden, is het nuttig vast te stellen welke segmentatiestrategie door baby's gebruikt werd: een dalgebaseerde (waarbij grenzen worden geplaatst op dieptepunten van de klankovergangswaarschijnlijkheid) of een chunking-gebaseerde strategie (waarbij fonemen die vaak samen voorkomen bij elkaar worden gehouden, oftewel gechunkt). Een dalgebaseerde strategie werd verondersteld de voorkeur te hebben van baby's, omdat het moeilijk is de waarschijnlijkheid van minder voorkomende foneemcombinaties te verwerven uit een klein lexicon, dat nog in opbouw is.

Er zijn twee experimenten uitgevoerd. In experiment 1 is een kunstmatige taal gebruikt die /s/-klinker-/s/-sequenties bevat. Deze sequentie komt relatief weinig voor in continue spraak, maar juist vaak binnen woorden, wanneer gekeken wordt naar het Nederlands zoals dat tot baby's wordt gesproken. Vijftien maanden oude Nederlandse baby's hoorden de grenzen tussen twee voorkomens van /s/, wat een voorkeur voor een dalgebaseerde segmentatiestrategie aantoont. In het tweede experiment werden negen en vijftien maanden oude baby's getest om te kijken of ze de chunking-aanwijzing van /p/-klinker-/p/ gebruiken. Deze sequentie komt relatief vaak voor in zowel de woorden als de continue spraak die tot baby's wordt gericht. Er werd echter geen voorkeur gevonden in de spraaksegmentatie van de baby's om deze combinatie bij elkaar te houden.

De resultaten ondersteunen de hypothese dat baby's segmentatieaanwijzingen gebaseerd op klankregels leren uit continue spraak en niet uit het lexicon en dat de aanwijzingen vooral de vorm hebben van gedetecteerde lage waarschijnlijkheden van foneemcombinaties, die gebruikt worden om woordgrenzen te plaatsen.

Hoofdstuk 3 en 4 borduren voort op het voorgaande door verder te kijken naar het gebruik van OCP-PLACE als aanwijzing voor segmentatie. OCP-PLACE is een abstracte beperking (constraint) die in veel talen voorkomt; de beperking betreft het samen voorkomen van medeklinkers met dezelfde waarde voor het fonologische kenmerk [plaats] (plaats van articulatie). Ook het Nederlandse lexicon wordt beperkt door OCP-PLACE: sequenties van medeklinkers die dezelfde plaats van articulatie hebben, komen minder vaak voor binnen woorden. Dit maakt deze beperking een mogelijke bron van bruikbare aanwijzingen voor segmentatie.

In dit proefschrift wordt de hypothese opgeworpen dat kennis van klankregels wordt verworven op taalspecifieke basis, uit de input. Het gebruik van kennis van klankregels is dan ook taalspecifiek. Er zijn echter wel functionalistische benaderingen waarbij wordt aangenomen dat OCP het gevolg is van een universele moeilijkheid in de waarneming om twee medeklinkers die met hetzelfde spraakorgaan worden gemaakt als verschillende medeklinkers te coderen.

Hoofdstuk 3 laat zien dat de effecten van OCP-PLACE op segmentatie niet het directe gevolg kunnen zijn van universele functionele voor- of afkeuren, door de rol van OCP in spraaksegmentatie aan te tonen in een reeks experimenten met kunstmatigetaalleren, met moedertaalsprekers van het Nederlands. Het Nederlandse lexicon wordt namelijk beperkt door OCP-LABIAL, maar niet door OCP-CORONAL. In overeenstemming met de verdeling bleken Nederlandse sprekers OCP-LABIAL wel en OCP-CORONAL niet te gebruiken als aanwijzing voor spraaksegmentatie. Dit laat zien dat OCP verband houdt met taalspecifieke foneemverdelingen. Omdat de kunstmatige talen bovendien geen bestaande woorden en ook geen statistische aanwijzingen uit het lexicon bevatten, tonen de resultaten aan dat OCP bruikbaar is om woordgrenzen te vinden zonder invloed van het hogere lexicale niveau op lagere niveaus van waarneming (top-downinvloed).

De resultaten kunnen niet door functionalistische theorieën verklaard worden, omdat die zouden voorspellen dat sequenties die OCP overtreden niet correct waargenomen kunnen worden, of niet kunnen worden gesplitst (als onderdeel van twee woorden) in de waarneming, tenzij top-downinvloed of andere fonologische aanwijzingen een woordgrens aangeven.

Hoofdstuk 4 gaat verder in op dit onderwerp door nader te kijken of kennis van OCP-PLACE die informatie geeft voor spraaksegmentatie veroorzaakt wordt door functionele voorkeuren, of door taalspecifieke verwerving. Eerst werden moedertaalsprekers van het Mandarijn getest op hun segmentatievoorkeuren. Het Mandarijn is een taal waarvan het lexicon niet onderhevig is aan beperkingen door OCP-PLACE. In tegenstelling tot de Nederlandse moedertaalsprekers in Hoofdstuk 3 gebruikten de Mandarijn-sprekende luisteraars geen OCP-PLACE om de kunstmatige taal te segmenteren, wat wel voorspeld zou worden als OCP het gevolg is van een universele functionele voorkeur.

Uit deze resultaten blijkt dat kennis van OCP-PLACE moet worden verworven, oftewel aangeleerd. Deze theorie werd op meer directe wijze getest met L2-leerders van het Nederlands. De hypothese die getest werd is dat de ontvangen hoeveelheid input van spraak (gesproken taal van moedertaalsprekers) cruciaal is voor het verwerven van OCP. Er werden experimenten uitgevoerd met drie groepen L2-leerders van het Nederlands. Er waren twee groepen vergevorderde L2-leerders met Mandarijn als moedertaal, waarvan één groep in Nederland woonde en de tweede groep in China. De derde groep bestond uit beginnende leerders van het Nederlands die in Nederland woonden. De resultaten bevestigden het belang van input van spraak geproduceerd door moedertaalsprekers voor het leren van de klankregels van een tweede taal. De voorwaarde voor het verwerven OCP-PLACE als een spraaksegmentatieaanwijzing voor de tweede taal is tweeledig: de leerder moet

voldoende taalvaardig zijn en daarnaast blootgesteld zijn aan voldoende input van moedertaalsprekers.

Beperkingen van lettergreepstructuur bij het leren van woorden

Hoofdstukken 5 en 6 behandelen de invloed van kennis van lettergreepstructuur en het ondersteunende effect van de waarschijnlijkheid van klankcombinaties op het leren van woorden. De nadruk ligt op de vraag of de waarschijnlijkheid van klankcombinaties en lettergreepstructuur deel zijn van twee aparte kenniscomponenten en of gescheiden opslag leidt tot gescheiden effecten van de twee componenten. De hypothese is dat ondersteunende effecten van klankcombinatiewaarschijnlijkheden versterkt kunnen worden door structurele kennis.

De gebruikte methode is kortetermijnherkenning. In taken waarbij non-woorden herinnerd moeten worden zijn de prestaties vaak gecorreleerd met de grootte van het lexicon, bij volwassenen en baby's en in de eerste en tweede taal. Dit wordt gezien als bewijs dat de kortetermijnherinnering in verband staat met het leren van woorden. In feite kan het herinneren van woorden op de korte termijn gezien worden als de eerste stap naar het opslaan van nieuwe woorden in het mentale lexicon. Hoe makkelijker het is om woorden in het kortetermijngeheugen op te slaan, des te makkelijker het is om het op te slaan in het langetermijngeheugen. In beide hoofdstukken worden hetzelfde experiment en dezelfde non-woordstimuli gebruikt, waarbij de stimuli verschillen in de lettergreepstructuur en in de frequentie van de bifonen (combinaties van twee fonen, spraakgeluiden).

Hoofdstuk 5 beschrijft twee experimenten met ééntalige Nederlandse volwassenen die lieten zien dat het ondersteunende effect van bifoonfrequentie significant wordt versterkt als complexe lettergrepen worden gememoriseerd, vergeleken met simpele lettergrepen. De resultaten ondersteunen de hypothese dat er een rol is bij het leren van woorden voor twee kenniscomponenten die klankregels representeren. Vooral belangrijk is dat één kenniscomponent het effect van de andere moduleert: kennis van de lettergreepstructuur kan het ondersteunende effect van de waarschijnlijkheid van klankcombinaties versterken.

Hoofdstuk 6 verdiept dit punt door de hypothese te testen dat de gescheiden opslag van de twee componenten ook betekent dat ze onafhankelijk van elkaar verworven kunnen worden. Dit is onderzocht door de effecten van bifoonfrequentie en lettergreepstructuur te testen in tweedetaalverwerving. Deelnemers waren L2-leerders met een moedertaal die minder complexe lettergrepen toestaat dan het Nederlands (CCCVCCC, waarbij C voor consonant oftewel medeklinker staat en V voor vocaal oftewel klinker). De moedertalen waren Japans (met als maximaal complexe lettergreepstructuur CVC) en Spaans (CCVC). De L2-leerders werden ondersteund door hoge frequentie van de

bifonen in de tweede taal, zelfs als de non-woorden structureel complexer waren dan de eerste taal toeliet. Ze maakten echter ook sneller fouten in sommige complexe lettergreeptypes die in hun eerste taal niet voorkwamen. Ze hadden dus een voordeel van de waarschijnlijkheden van klankcombinaties in de tweede taal ondanks het gebruik van structuren uit de eerste taal. Deze resultaten ondersteunen de hypothese dat kennis van klankcombinatiewaarschijnlijkheden apart verworven wordt van kennis van de structuur waarin ze optreden.

Conclusie

Wat betreft de representatie van de kennis van klankregels laat dit proefschrift zien dat niet alleen de waarschijnlijkheid van klankcombinaties, maar ook klankregels van een meer abstracte of complexe natuur wordt gebruikt voor het segmenteren van spraak en het leren van woorden. Dit is aangetoond voor twee soorten structurele kennis: de beperkingen op de foneemstructuur betreffende de overeenkomsten tussen niet-aangrenzende medeklinkers (hoofdstukken 2, 3 en 4) en beperkingen op lettergreepstructuur (hoofdstukken 5 en 6).

Wat betreft het gebruik van kennis van klankregels kan uit hoofdstuk 2 afgeleid worden dat aanwijzingen van relatieve lage waarschijnlijkheden bij uitstek nuttig zijn voor spraaksegmentatie: Baby's gebruikten hun kennis van relatief zeldzame foneemparen, maar niet van relatief frequente paren. In continue spraak zijn de meeste foneemparen onderdeel van hetzelfde woord, omdat woorden over het algemeen uit meer dan twee of drie fonemen bestaan. Foneemparen die binnen woorden kunnen voorkomen zijn meestal relatief waarschijnlijk, waardoor er minder dalen dan pieken in overgangswaarschijnlijkheid optreden in continue spraak. Om die reden is het nuttiger op de dalen te letten.

Aan de andere kant kan bij het leren van woorden juist een hoge waarschijnlijkheid nuttig zijn. Dit wordt ook gesuggereerd door de resultaten van hoofdstuk 6. L2-leerders werden bij het leren van woorden beter in het gebruik van aanwijzingen van hoge, niet lage, waarschijnlijkheid van de klankcombinatie.

Bij de verwerving van kennis van klankregels werd gekeken naar de mogelijk functionele oorsprong van abstracte beperkingen op klankcombinaties die in meerdere talen terugkomen. Er kunnen functioneel gemotiveerde eigenschappen van de waarneming aan die beperkingen ten grondslag liggen, maar de beperkingen kunnen ook taalafhankelijk zijn. In het laatste geval zijn ze verworven uit de input, mogelijk met gebruik van aangeboren eigenschappen van het leerproces.

De resultaten van de experimenten uit de hoofdstukken 2, 3 en 4 laten zien dat kennis van OCP niet teruggebracht kan worden tot algemene eigenschappen van het waarnemingsproces. Huidige functionalistische theorieën zouden

voorspellen dat luisteraars sequenties van medeklinkers met dezelfde plaats van articulatie alleen kunnen interpreteren als twee elementen als er extra informatie is om die interpretatie te ondersteunen, idealiter top-downinformatie uit het lexicon. Die voorspelling blijkt nu echter incorrect, aangezien luisteraars OCP ook gebruiken bij het segmenteren van een kunstmatige taal waar alle mogelijke aanwijzingen van lexicale kennis uit verwijderd zijn en zulke top-downinformatie dus niet kan optreden. Bovendien is aangetoond dat de effecten van OCP op taalverwerking taalspecifiek zijn: luisteraars met een moedertaal die niet beperkt is door OCP gebruiken de beperking ook niet in verwerking van spraak. Bovendien hebben we vastgesteld dat het gebruik van OCP-PLACE voor spraaksegmentatie verworven kan worden in een tweede taal. Dit suggereert dat de beperking uit de input wordt geleerd.

Een andere belangrijke vaststelling over de verwerving van klankregels is dat kennis van die regels tenminste deels uit continue spraak moet komen, in plaats van uit het lexicon. In de experimenten in hoofdstuk 2 gebruiken baby's bij het segmenteren van spraak kennis van het relatief weinig voorkomen van een patroon. Deze kennis kunnen ze alleen verworven hebben uit continue spraak, aangezien hetzelfde patroon juist relatief vaak voorkomt in de woorden die tot de baby gericht worden.