

# The Induction of Phonotactics for Speech Segmentation

Converging evidence from computational and  
human learners

Published by  
LOT  
Trans 10  
3512 JK Utrecht  
The Netherlands

phone: +31 30 253 6006  
fax: +31 30 253 6406  
e-mail: [lot@uu.nl](mailto:lot@uu.nl)  
<http://www.lotschool.nl>

Cover illustration: Elbertus Majoor, *Landschap-fantasie II* (detail), 1958, gouache.

ISBN: 978-94-6093-049-2  
NUR 616

Copyright © 2011: Frans Adriaans. All rights reserved.

# The Induction of Phonotactics for Speech Segmentation

Converging evidence from computational and  
human learners

## Het Induceren van Fonotactiek voor Spraksegmentatie

Convergerende evidentie uit computationele en menselijke  
leerders

*(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. J.C. Stoof, ingevolge het besluit  
van het college voor promoties in het openbaar te verdedigen  
op vrijdag 25 februari 2011 des middags te 2.30 uur

door

Frans Willem Adriaans

geboren op 14 maart 1981  
te Ooststellingwerf

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

*To my parents, Ank & Piet*



# CONTENTS

---

ACKNOWLEDGEMENTS	xi
1 INTRODUCTION	1
1.1 The speech segmentation problem . . . . .	1
1.2 Segmentation cues and their acquisition . . . . .	3
1.2.1 The role of segmentation cues in spoken word recognition	4
1.2.2 The acquisition of segmentation cues by infants . . . . .	7
1.3 The induction of phonotactics for speech segmentation . . . . .	11
1.3.1 Bottom-up versus top-down . . . . .	11
1.3.2 Computational modeling of phonotactic learning . . . . .	12
1.3.3 Two hypotheses regarding the acquisition of phonotac-	
tics by infants . . . . .	16
1.4 Learning mechanisms in early language acquisition . . . . .	17
1.4.1 Statistical learning . . . . .	18
1.4.2 Generalization . . . . .	19
1.4.3 Towards a unified account of infant learning mechanisms	22
1.5 Research questions and scientific contribution . . . . .	23
1.6 Overview of the dissertation . . . . .	26
2 A COMPUTATIONAL MODEL OF PHONOTACTIC LEARNING AND SEG-	
MENTATION	31
2.1 Introduction . . . . .	31
2.1.1 Models of speech segmentation using phonotactics . . . . .	32
2.1.2 Models of constraint induction . . . . .	33
2.2 The OT segmentation model . . . . .	35
2.3 The learning model: StaGe . . . . .	38
2.3.1 Statistical learning . . . . .	39
2.3.2 Frequency-Driven Constraint Induction . . . . .	40
2.3.3 Single-Feature Abstraction . . . . .	42
2.4 An example: The segmentation of plosive-liquid sequences . . . . .	45
2.5 General discussion . . . . .	49
3 SIMULATIONS OF SEGMENTATION USING PHONOTACTICS	53
3.1 Introduction . . . . .	53
3.2 Experiment 1: Biphones in isolation . . . . .	54
3.2.1 Method . . . . .	55
3.2.2 Results and discussion . . . . .	58
3.3 Experiment 2: Thresholds . . . . .	61
3.3.1 Method . . . . .	61
3.3.2 Results and discussion . . . . .	62
3.4 Experiment 3: Biphones in context . . . . .	65

## CONTENTS

3.4.1	Method . . . . .	65
3.4.2	Results and discussion . . . . .	66
3.5	Experiment 4: Input quantity . . . . .	69
3.5.1	Modeling the development of phonotactic learning . . . . .	69
3.5.2	Method . . . . .	70
3.5.3	Results and discussion . . . . .	70
3.6	General discussion . . . . .	72
4	MODELING OCP-PLACE AND ITS EFFECT ON SEGMENTATION . . . . .	75
4.1	Introduction . . . . .	76
4.1.1	OCP effects in speech segmentation . . . . .	79
4.1.2	Modeling OCP-PLACE with STAGE . . . . .	81
4.2	Methodology . . . . .	86
4.2.1	Materials . . . . .	86
4.2.2	Procedure . . . . .	87
4.2.3	Evaluation . . . . .	88
4.3	The induction of OCP-PLACE . . . . .	90
4.3.1	Predictions of the constraint set with respect to word boundaries . . . . .	94
4.4	Experiment 1: Model comparisons . . . . .	99
4.4.1	Segmentation models . . . . .	99
4.4.2	Linear regression analyses . . . . .	100
4.4.3	Discussion . . . . .	102
4.5	Experiment 2: Input comparisons . . . . .	103
4.5.1	Segmentation models . . . . .	103
4.5.2	Linear regression analyses . . . . .	105
4.5.3	Discussion . . . . .	112
4.6	General discussion . . . . .	114
5	THE INDUCTION OF NOVEL PHONOTACTICS BY HUMAN LEARNERS . . . . .	121
5.1	Introduction . . . . .	121
5.2	Experiment 1 . . . . .	128
5.2.1	Method . . . . .	130
5.2.2	Results and discussion . . . . .	133
5.3	Experiment 2 . . . . .	137
5.3.1	Method . . . . .	138
5.3.2	Results and discussion . . . . .	141
5.4	Experiment 3 . . . . .	144
5.4.1	Method . . . . .	144
5.4.2	Results and discussion . . . . .	146
5.5	General discussion . . . . .	148



## CONTENTS

6	SUMMARY, DISCUSSION, AND CONCLUSIONS	153
6.1	Summary of the dissertation . . . . .	154
6.1.1	A combined perspective . . . . .	154
6.1.2	The learning path . . . . .	156
6.1.3	The proposal and its evaluation . . . . .	157
6.2	Discussion and future work . . . . .	159
6.2.1	Comparison to other constraint induction models . . . . .	159
6.2.2	What is the input for phonotactic learning? . . . . .	162
6.2.3	Which mechanisms are involved in phonotactic learning and segmentation? . . . . .	164
6.2.4	The formation of a proto-lexicon . . . . .	168
6.2.5	Language-specific phonotactic learning . . . . .	170
6.2.6	A methodological note on the computational modeling of language acquisition . . . . .	171
6.3	Conclusions . . . . .	172
A	THE STAGE SOFTWARE PACKAGE	173
B	FEATURE SPECIFICATIONS	181
C	EXAMPLES OF SEGMENTATION OUTPUT	183
D	EXPERIMENTAL ITEMS	187
	References	191
	SAMENVATTING IN HET NEDERLANDS	205
	CURRICULUM VITAE	209



## ACKNOWLEDGEMENTS

---

And so this journey ends... Many people have supported me in doing my research over the past few years, and in writing this book. Some gave critical feedback, or helped solving various problems. Others were simply a pleasant source of distraction. All of these people have provided essential contributions to this dissertation, and ensured that I had a very enjoyable time while doing my PhD research in Utrecht.

This book could not have materialized without the inspiration and feedback from my supervisor, René Kager. When I started my project (which was the ‘computational’ sub-project within René’s NWO Vici project), I did not know that much about linguistics, let alone psycholinguistics. René provided me with an environment in which I could freely explore new ideas (whether they were computational, linguistic, psychological, or any mixture of the three), while providing me with the essential guidance for things to make sense in the end. I want to thank René for his enthusiasm, and for always challenging me with the right questions. In addition to many hours of scientific discussion, we also got along on a more personal level. Not many PhD students have the opportunity to hang out with their supervisors in record stores in Toulouse, while discussing jazz legends such as Mingus and Dolphy. I consider myself lucky.

I was also fortunate to be part of a research group of very smart people, who became close friends. I am especially grateful to my fellow ‘groupies’ Natalie Boll-Avetisyan and Tom Lentz. The research presented in this book has clearly benefited from the many discussions I had with them throughout the years. These discussions sometimes took place in exotic locations: Natalie and I discussed some of the basic ideas of our projects by the seaside in Croatia; Tom and I came up with new research plans at the San Francisco Brewing Company. Aside from obvious fun, these trips and discussions helped me to form my ideas, and to look at them in a critical way.

I would like to thank Diana Apoussidou, who joined the group later, for helpful advice on various issues, and for discussions about computational modeling. Thanks also to the other members of the group meetings (and eatings): Chen Ao, Frits van Brenk, Marieke Kolkman, Liquan Liu, Bart Penning de Vries, Johannes Schliesser, Tim Schoof, Keren Shatzman, Marko Simonović, Jorinde Timmer. I also want to thank two students that I have supervised, Joost Bastings and Andréa Davis, for taking an interest in my work, and for providing me with a fresh perspective on my research.

For this book to take on its current shape, I am very much indebted to Tikitu de Jager for proofreading, and for helping me with various  $\LaTeX$  layout issues. Also thanks to Gianluca Giorgolo for help with technical issues, but mainly for the excellent suggestion of watching Star Trek while writing a

#### ACKNOWLEDGEMENTS

dissertation. ('These people get things *done!*') Thanks to Donald Roos for editing the cover illustration. I also want to thank Ingrid, Angelo, and Arifin at Brandmeester's for a continuous supply of delicious caffeine throughout the summer.

Several people at the Utrecht Institute of Linguistics OTS have helped me with various aspects of my research. Elise de Bree and Annemarie Kerkhoff gave me very useful advice on designing experiments. Jorinde Timmer assisted me in running the experiments. I want to thank Jacqueline van Kampen for being my tutor. I am also grateful to Maaïke Schoorlemmer, Martin Everaert and the UiL OTS secretaries for providing me with a pleasant work environment.

Fortunately, there was more to life in Utrecht than just work. I would like to thank Hannah De Mulder, Loes Koring (a.k.a. 'Lois'), Arjen Zondervan, Sander van der Harst, Arnout Koornneef, Daria Bahtina, Marko Simonović, and, again, Tom, Natalie, and Diana, for many VrijMiBo's, lunches, dinners, concerts, etc. Thanks also to Gianluca Giorgolo and Gaetano Fiorin for a rather unique jam session at dB's. I want to thank the other PhD students and postdocs at UiL OTS for occasional meetings, drinks, futsal games, and many enjoyable conversations at Janskerkhof and Achter de Dom: Ana Aguilar, Ivana Brasileiro, Desiree Capel, Anna Chernilovskaya, Alexis Dimitriadis, Xiaoli Dong, Jakub Dotlacil, Lizet van Ewijk, Nicole Grégoire, Nadya Goldberg, Bettina Gruber, Kate Huddlestone, Naomi Kamoen, Cem Keskin, Anna Kijak, Huib Kranendonk, Kiki Kushartanti, Bert LeBruyn, Emilienne Ngangoum, Rick Nouwen, Andreas Pankau, Liv Persson, Min Que, Dagmar Schadler, Rianne Schippers, Marieke Schouwstra, Giorgos Spathas, Roberta Tedeschi, Christina Unger, Sharon Unsworth, Rosie van Veen, Anna Volkova, Eline Westerhout, Marie-Elise van der Ziel, and Rob Zwitterlood. Thanks to Luca Ducceschi for providing me with bass talk, Italian coffee and Tony Allen during the last months at Achter de Dom.

Then there is a group of people that deserve mentioning because they have supported me throughout the years, and continue to make my life enjoyable in many different ways. I would like to thank Eeke Jagers for being a great 'gozor', and for making sure that I got a regular dose of excellent food, wine, and The Wire; Marijn Koolen for almost a decade of inspiring discussions that somehow always switch unnoticeably between scientific topics and P-Funk; Berend Jan Ike for many soothing drinks, conversations and jam sessions; Niels Schoenmaker and Freek Stijntjes for keeping me on the beat (both musically and intellectually); Julien Oomen and Sanne van Santvoort for many years of enduring friendship; Nanda Jansen for support, especially when I most needed it; Tom, Belén, Marijn, Anna, Tikitu, and Olga for a seemingly never-ending series of Dialogue Dinners; Snoritas (Chuntug, Daan, Walter,

#### ACKNOWLEDGEMENTS

Maarten, and others) for fun times in Amsterdam; Cylia Engels for a room with a stunning view and many cups of tea in Amsterdam-Noord. A special thanks to Ilke for adding an unexpected sense of holiday to my summer in Utrecht, and for being incredibly supportive in these stressful times.

Last but not least, I would like to thank my family. I want to thank my brothers Nanne and Rik, who always keep me sharp, and who never cease to make me laugh. Both of them are just incredibly smart and incredibly funny. Thanks to Talita for being part of the family, and to little Naäma for providing me with some interesting data on infant speech (*'Nee-nee-nee-nee!'*). In the end, I owe everything to my parents, who have always inspired me and encouraged me to follow my path in life. None of this could have been accomplished without your support. I love you deeply, and I dedicate this book to you.



# 1

## INTRODUCTION

---

### 1.1 THE SPEECH SEGMENTATION PROBLEM

Listening to speech is automatic and seemingly effortless. It is an essential part of our daily communication. Whenever a speaker produces a sentence, an acoustic signal is created which contains a sequence of words. As listeners, we generally have no problem hearing and understanding these spoken words. That is, of course, provided we are familiar with the language of the speaker, and provided that there are no exceptionally noisy conditions (such as during a conversation at a loud rock concert). The ease with which we listen to speech is quite remarkable when one considers the processes that are involved in decoding the speech stream. The acoustic signal that is produced by a speaker does not consist of words which are separated by silences, but rather of a continuous concatenation of words. As a consequence, word boundaries are usually not clearly marked in the speech signal (Cole & Jakimik, 1980; Klatt, 1979). Spoken language is thus quite different from written language. While the written words that you are reading in this dissertation are visually separated by blank spaces, words in spoken language typically have no clearly delineated beginnings or endings. *One could thus think of the speech signal as a text without spaces.*

In order to understand speech, the listener has to break down the continuous speech signal into a discrete sequence of words. As adult listeners, we are greatly helped by our knowledge of the native language vocabulary. We have the ability to retrieve word forms that we have stored in our mental lexicon, allowing us to recognize such word forms in the speech signal. For example, when hearing the continuous sentence listed above, we recognize English words such as *could* and *signal*. At the same time, there are chunks of speech in the signal which do not correspond to words. For example, *nalas* is not an English word, nor is *speechsi*. Indeed, models of spoken word recognition have traditionally relied on finding matches between stretches of sound in the speech signal and word forms in the lexicon (e.g., Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). In these models, parts of the speech signal are matched with words that we have stored in our mental lexicon, until every part of the speech signal corresponds to a meaningful word.

## INTRODUCTION

The problem of speech segmentation is particularly challenging for language learners. From the moment they are born, infants are confronted with speech input that is continuous. Even in speech specifically directed toward the infant, most words are embedded in a continuous sentence, rather than being presented in isolation (van de Weijer, 1998; Woodward & Aslin, 1990). Infants develop the ability to segment words from continuous speech already in the first year of life (e.g., Jusczyk & Aslin, 1995). A fundamental difference, however, in speech segmentation by adults and infants is that infants do not yet possess the complete vocabulary of the native language (e.g., Fenson et al., 1994). In fact, one of the major tasks that infants face is to *learn* the words of their native language. The development of the mental lexicon crucially relies on infants' ability to break up the speech stream into word-like units.

While infants do not yet have access to a large lexicon of word forms, they are helped by various *sublexical* cues for segmentation. Sublexical segmentation cues convey information about the structure of the speech stream, but are not linked to specific words in the lexicon. Such cues may provide infants with an initial means to break up the speech stream, and could allow them to bootstrap into lexical acquisition (e.g., Mehler, Dupoux, & Segui, 1990; Christophe, Dupoux, Bertoncini, & Mehler, 1994). The sublexical segmentation cue that will be the focus of this dissertation is *phonotactics*. Phonotactic constraints state which sound sequences are allowed to occur within the words of a language. This information is useful for segmentation, since it highlights possible locations of word boundaries in the speech stream. In general, if a sequence occurs in the speech stream that is not allowed to occur within the words of a language, then such a sequence is likely to contain a word boundary. For example, the sequence /pf/ does not occur in Dutch words. Therefore, when /pf/ occurs in the speech stream, listeners know that /p/ and /f/ belong to separate words. To illustrate this, consider the following continuous Dutch utterance:

(1.1) /dɔləmpfil/      (*de lamp viel*, 'the lamp fell')

The phonotactic knowledge that /pf/ cannot occur within words can be used by listeners to segment this utterance. A word boundary (denoted by '.') can be hypothesized within the illegal sequence:

(1.2) /dɔləmp.fil/

With this simple phonotactic constraint the listener has separated /dɔləmp/ from /fil/. Note that knowledge of phonotactics is language-specific. A German listener not familiar with the Dutch vocabulary may be inclined to think that /pfil/ is a word, since the phonotactics of German allows words to start with /pf/ (e.g., *pfeffer*, 'pepper'). Languages thus differ in the sound



## 1.2 SEGMENTATION CUES AND THEIR ACQUISITION

sequences that are allowed within words. The consequence is that phonotactic knowledge has to be acquired from experience with a specific language.

Phonotactics is used in the segmentation of continuous speech by both adults and infants (McQueen, 1998; Mattys & Jusczyk, 2001b). For the case of speech perception by adults, knowledge of phonotactics supports the retrieval of words in the mental lexicon (Norris, McQueen, Cutler, & Butterfield, 1997). Spoken word recognition thus involves a combination of sublexical segmentation cues and lexical lookup (Cutler, 1996; Mattys, White, & Melhorn, 2005). For the case of segmentation by infant language learners, phonotactics may provide a basis for the development of the mental lexicon (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). In the current example, the infant could add /dɒləmp/ and /fɪl/ as initial entries in a 'proto-lexicon'. During later stages of lexical acquisition, the infant further decomposes such lexical entries, and attributes meaning to the words.

The relevance of phonotactics for language development raises the question of how phonotactics is acquired. The central question in this dissertation is the following: How do learners acquire the knowledge of phonotactics that helps them to break up the speech stream? The induction of phonotactic constraints for speech segmentation will be addressed through a combination of computational modeling of infant learning mechanisms, computer simulations of speech segmentation, and language learning experiments with human participants. The goal is to provide insight into some essential aspects of early language acquisition. The proposed model is explicit, since it is implemented as a computer program, and is supported by psychological evidence of human learning capacities.

In what follows, the role of segmentation cues in spoken word recognition, and the acquisition of these cues by infants will be discussed (Section 1.2). Two hypotheses are then outlined regarding the induction of phonotactics for speech segmentation (Section 1.3), followed by a discussion of mechanisms that are available to infants for phonotactic learning (Section 1.4). Finally, the research questions and scientific contribution are stated (Section 1.5), and an overview of the different chapters in the dissertation is given (Section 1.6).

## 1.2 SEGMENTATION CUES AND THEIR ACQUISITION

Spoken word recognition is aided by several types of linguistic cues. Each of these cues helps the listener to break down the continuous speech stream into discrete, word-sized units, thereby facilitating the activation of words in the mental lexicon. Cues that have been demonstrated to affect speech segmentation broadly fall into three categories: metrical cues, fine-grained acoustic cues, and phonotactics.

1.2.1 *The role of segmentation cues in spoken word recognition*

A large body of research has focused on the role of *metrical cues* in segmentation. Metrical cues concern the ordering of strong and weak syllables in a language. For example, most English words have a trochaic stress pattern, where word-initial strong syllables (which contain full vowels) are optionally followed by one or more weak syllables (which contain reduced vowels). The Metrical Segmentation Strategy (MSS, Cutler, 1990) postulates that strong syllables occurring in the speech stream are taken as word onsets by the listener, thereby providing a starting point for lexical access. The MSS is supported by corpus analyses showing that the majority of English lexical words (approximately 85%) starts with a strong syllable (Cutler & Carter, 1987). If English listeners insert word boundaries before strong syllables, they would thus make considerable progress in segmenting the speech stream (see e.g., Harrington, Watson, & Cooper, 1989).

Evidence for the MSS comes from perception studies. Cutler and Norris (1988) show that listeners are faster to spot English words in speech when they are embedded in a sequence of a strong and a weak syllable (e.g., *mint* in *mintesh*) than when embedded in two strong syllables (e.g., *mintayve*). They argue that this is due to the insertion of a word boundary before strong syllables. That is, the occurrence of a strong syllable initiates a new attempt at lexical access, whereas the occurrence of a weak syllable does not. Listeners are slower at recognizing *mint* in *mintayve*, since recognition in this case requires the assembly of a word form across a hypothesized word boundary (*min.tayve*). Cutler and Norris suggest that the slowing-down effect may be the result of competing lexical hypotheses (competing for the shared /t/). Indeed, recognition only slowed down when the target word crossed a syllable boundary. That is, no difference in response was found for words that did not cross the hypothesized word boundary (e.g., *thin* in *thintayve* vs. *thintef*).

In a follow-up study, Cutler and Butterfield (1992) found that English listeners insert boundaries before strong syllables, and, in addition, delete boundaries before weak syllables. Analyses were conducted on data sets of erroneous segmentations ('slips of the ear') made by English listeners. The analyses showed that many misperceptions could be explained by the MSS. That is, incorrect boundaries were generally placed before strong syllables, and many missed boundaries occurred before weak syllables. In addition, boundaries inserted before strong syllables generally occurred at the beginning of lexical (content) words, whereas boundaries inserted before weak syllables occurred at the beginning of grammatical (function) words. These perception studies support the view that stress patterns affect the segmentation of continuous speech, as predicted by the metrical segmentation hypothesis.

A different type of segmentation cue is *fine-grained acoustic information* in the speech signal. For example, word onsets in English are marked by aspiration for voiceless stops (Christie, 1974; Lehiste, 1960), and by glottal stops and laryngealization for stressed vowels (Nakatani & Dukes, 1977). Another cue for word onsets is duration. The duration of a segment is generally longer for word-initial segments than for segments in word-medial or word-final position (Oller, 1973; Umeda, 1977). Hearing such acoustic events in the speech stream may lead listeners to hypothesize word beginnings, thereby facilitating lexical access (e.g., Gow & Gordon, 1995).

Segment duration indeed has an effect on the segmentation of continuous speech (e.g., Quené, 1992; Shatzman & McQueen, 2006). In two eye-tracking experiments, Shatzman and McQueen (2006) found that a shorter duration of /s/ facilitates the detection of a following stop-initial target (e.g., *pot* in *eenspot*). The short duration indicates that /s/ is likely to occur in a word-final position. Conversely, a longer duration indicates a word-initial position, which leads listeners to hypothesize a word starting with an '/s/ + consonant' cluster (e.g., *spot*). The consequence is that the detection of the target word *pot* slows down. Listeners thus rely on segment duration to resolve ambiguities that arise in the segmentation of continuous speech.

Finally, there are *phonotactic cues* that affect spoken word recognition. Studies on the role of phonotactics in speech segmentation have typically assumed a view based on sequential constraints. Such constraints encode an absolute or gradient avoidance of certain sequences within the words of a language. For example, the sequence /mr/ is not allowed in languages like English and Dutch. In a word spotting study, McQueen (1998) shows that Dutch listeners are sensitive to phonotactics and use phonotactic information to segment the speech stream. Words embedded in nonsense speech were either aligned or misaligned with a phonotactic boundary. Listeners were faster to spot words when a consonant cluster formed an illegal onset cluster, indicating a phonotactic boundary corresponding to the word onset (e.g., *rok* 'skirt' in *fiemrok*), than when the cluster formed a legal onset cluster (e.g., *fiedrok*). In the latter case the phonotactic boundary (*fie.drok*) did not match with the onset of the target word, resulting in a slower response in detecting the word. In a follow-up study, Lentz and Kager (in preparation) show that these effects are due to sublexical phonotactics providing an initial chunking of the speech stream before lexical look-up.

Several studies have shown that the phonotactic knowledge that is used for segmentation is more fine-grained than a classification into legal and illegal sequences. Listeners also use phonotactic *probabilities* for segmentation. That is, legal sequences can be scaled according to how likely they are to occur in a language. During segmentation, high-probability sequences provide a

## INTRODUCTION

stronger cue that a sequence is word-internal, while low-probability sequences are less likely to be word-internal. A word spotting study by Van der Lugt (2001) provides evidence that the likelihood of the occurrence of CV onset sequences affects segmentation by Dutch listeners. Words were easier to spot when they were followed by a high-probability CV onset (e.g., *boom* 'tree' in *boomdif*) than when followed by a low-probability onset (e.g., *boomdouf*). Importantly, phonotactic probabilities have been shown to affect the processing of nonwords independently of lexical effects such as lexical neighborhood density (Vitevitch & Luce, 1998, 1999, 2005). This supports a sublexical view of phonotactics in which probabilistic phonotactic knowledge is represented independently of the mental lexicon.

While these studies show that constraints on specific sequences of phonetic segments affect speech segmentation, there is evidence that segmentation is also affected by more complex phonological processes. For example, Finnish listeners use vowel harmony as a cue for segmentation (Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & Gelder, 1998). Furthermore, segmentation in Korean is affected by an interaction of multiple phonological processes underlying the surface structures (Warner, Kim, Davis, & Cutler, 2005). These studies show that phonotactic constraints do not merely operate on the sequences that surface in the speech stream, but also target the phonological structure of segments, taking into account the features that define them (see e.g., Chomsky & Halle, 1968).

As a consequence of the language-specific nature of phonotactics, native (L1) language phonotactic constraints potentially interfere with spoken word recognition in a second (L2) language (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Weber & Cutler, 2006). Weber and Cutler (2006) show that the detection of English words by German listeners who were advanced learners of English was influenced by German phonotactics. Thus, even advanced learners of a second language are affected by their native language phonotactics in L2 speech segmentation. The German learners of English, however, were also able to exploit English phonotactic cues for boundaries. These findings suggest that, while L1 phonotactics may interfere with L2 listening, advanced learners of a second language can acquire the phonotactic structure of the target language (Weber & Cutler, 2006; Trapman & Kager, 2009).

A more general phonotactic constraint, which seems to hold in most languages, is the Possible-Word Constraint (Norris et al., 1997; McQueen, Otake, & Cutler, 2001). This constraint states that word recognition should only produce chunks of speech which conform to a minimal word structure. Specifically, words minimally contain a vowel. A single consonant can therefore not by itself be a word. The segmentation mechanism should thus avoid

producing occurrences of isolated consonants. Norris et al. (1997) found that listeners were faster at detecting *apple* in *vuffapple* than in *vapple*. They argue that this is due to the fact that the residue of *vuffapple* is a possible word (*vuff*), whereas the residue of *vapple* is a single consonant (*v*), and therefore violates the PWC. While the PWC was originally proposed as a universal constraint on spoken word recognition, the constraint can be overruled in languages that allow single-consonant words, such as Slovak (Hanulíková, McQueen, & Mitterer, 2010).

Taken together, these studies show that sublexical cues facilitate the recognition of words in continuous speech. In order to obtain a more complete picture of spoken word recognition in adult listeners, it is interesting to look at some of the relative contributions of the different types of cues to word recognition. For example, Vitevitch and Luce (1998, 1999) show that effects of probabilistic phonotactics are mainly found in the processing of nonwords, while lexical effects, such as neighborhood density (i.e., the number of words in the lexicon that differ from the target word by a single phoneme) mainly arise in the processing of actual lexical items. For the case of speech segmentation, Mattys et al. (2005) found evidence for a hierarchy in which lexical cues (the actual words) dominate segmental cues (e.g., phonotactics, allophony). In turn, segmental cues are more important than prosodic cues (word stress). Whenever one type of cue fails (for instance, due to noise in the speech stream, or due to lack of lexical information), participants would rely on the next cue in the hierarchy.

The studies discussed so far concern the role of segmentation cues in spoken word recognition by adult listeners. However, the relative importance of lexical and sublexical cues for segmentation may be quite different for infant language learners. During early language acquisition, infants are starting to build up a vocabulary of words. It thus seems that infants, at least initially, do not rely on lexical strategies to discover words in continuous speech. Rather, infants will mainly rely on sublexical cues for speech segmentation (e.g., Cutler, 1996). Below evidence is discussed showing that infants in their first year of life acquire segmentation cues. Such cues help them in constructing a lexicon from continuous speech input.

### 1.2.2 *The acquisition of segmentation cues by infants*

Infants acquire knowledge about the sublexical properties of their native language during the first year of life. The primary purpose of the sublexical knowledge that infants possess (metrical patterns, fine-grained phonetic cues, and phonotactics) is to segment and recognize words in fluent speech (e.g.,

Jusczyk, 1997). Segmentation cues thus guide infants' search for words in continuous speech, and facilitate the development of the mental lexicon.

With respect to metrical cues, infants have a sensitivity to the basic rhythmic properties of their native language from birth (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998). More fine-grained knowledge of native language patterns of stressed (strong) and unstressed (weak) syllables develops between the age of 6 and 9 months (Jusczyk, Cutler, & Redanz, 1993; Morgan & Saffran, 1995). Jusczyk, Cutler, and Redanz (1993) show that 9-month-old infants listen longer to words that follow the predominant stress pattern of English (strong-weak) than to words that follow the opposite pattern (weak-strong). Younger infants (6-month-olds) did not show any significant listening preferences, suggesting that infants learn about language-specific stress patterns as a result of exposure to the native language during the first year of life. The ability to use metrical patterns as a cue for segmenting word-like units from speech is present at the age of 7.5 months (Jusczyk, Houston, & Newsome, 1999). In accordance with the Metrical Segmentation Strategy, infants segmented bisyllabic words conforming to the English strong-weak pattern from continuous speech. In addition to behavioral studies, there is electrophysiological (ERP) evidence that Dutch 10-month-old infants rely on stress patterns for segmentation (Kooijman, Hagoort, & Cutler, 2009).

Fine-grained acoustic cues that play a role in infant language development include co-articulation, allophony, and duration. Specifically, infants have been shown to be sensitive to co-articulation between segments at the age of 5 months (Fowler, Best, & McRoberts, 1990), and to context-sensitive allophones at the age of 10.5 months (Jusczyk, Hohne, & Bauman, 1999). These phonetic factors also have an impact on the segmentation of continuous speech by infants (Johnson & Jusczyk, 2001; Jusczyk, Hohne, & Bauman, 1999; Mattys & Jusczyk, 2001a). Johnson and Jusczyk (2001) show that 8-month-old infants use co-articulation as a cue in extracting trisyllabic artificial words from continuous speech. In addition, Jusczyk, Hohne, and Bauman (1999) found evidence suggesting that, by the age of 10.5 months, infants use information about allophonic variation to extract words from speech. Such knowledge allows infants to distinguish between different context-sensitive realizations of a phoneme, and thus allows them to segment phrases that are phonemically equivalent, but made up of different allophones (e.g., different realizations of *t* and *r* in *night rates* versus *nitrates*). Finally, durational cues corresponding to phonological phrase boundaries are used for segmentation in infants of 10 and 13 months of age (Gout, Christophe, & Morgan, 2004).

In addition to metrical and acoustic cues, infants learn about the phonotactic patterns of their native language (Friederici & Wessels, 1993; Jusczyk, Friederici, et al., 1993; Jusczyk, Luce, & Charles-Luce, 1994), and use phono-

tactic cues for segmentation (Johnson, Jusczyk, Cutler, & Norris, 2003; Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001b; Myers et al., 1996). Jusczyk, Friederici, et al. (1993) show that 9-month-old infants listen longer to words from their native language than to nonnative words (which violate the phonotactic constraints of the language). Similar results were obtained in studies using nonwords: Infants listen longer to nonwords that respect the native language phonotactics than to nonwords that violate the phonotactic constraints of the native language (Friederici & Wessels, 1993). These studies show that infants prefer to listen to words that conform to the phonotactic patterns of the native language.

Infants' sensitivity to phonotactics is more fine-grained than the ability to distinguish between legal and illegal sound sequences. Infants have been shown to be sensitive to phonotactic probabilities. That is, infants are able to distinguish between legal sequences that differ in how likely they are to occur in the native language. Jusczyk et al. (1994) found that infants have a preference for nonwords that were made up of high-probability sound sequences (e.g., *riss*) over nonwords that consisted of low-probability sequences (e.g., *yowdge*). While infants at 9 months of age listened longer to lists of high-probability nonwords than to lists of low-probability nonwords, 6-month-old infants did not show this sensitivity. This finding indicates that, in the period between the first 6 and 9 months of life, infants start to learn about the phonotactic patterns of their native language from experience with speech input.

Knowledge of phonotactics is useful for infants, since it indicates which sequences in the continuous speech stream are likely to contain boundaries, and which sequences are not. Sequences that violate the native language phonotactics are plausible locations for word boundaries in the speech stream. From a probabilistic perspective, low-probability sequences are more likely to contain word boundaries than high-probability sequences. Mattys and Jusczyk (2001b) show that 9-month-old infants use probabilistic phonotactic cues to segment words from continuous speech. Items were embedded in either between-word consonant clusters (i.e., consonant clusters which occur more frequently between words than within words), or embedded in within-word consonant clusters (i.e., consonant clusters which occur more frequently within words than between words). For example, during familiarization the item *gaffe* would occur in the context of a phrase containing *bean gaffe hold* (where /ng/ and /fh/ are between-word clusters, and thus provide a phonotactic cue for the segmentation of *gaffe*), or in the context of *fang gaffe tine* (where /ŋg/ and /ft/ are within-word clusters, and thus provide no indication of word boundaries at the edges of *gaffe*). During the test phase infants listened longer to words which had been embedded in between-word consonant clusters

## INTRODUCTION

than to words which had been embedded in within-word consonant clusters. Infants thus appear to have knowledge of the likelihood of boundaries in segment sequences in continuous speech.

In addition to language-specific phonotactic cues, there is evidence that infants use the more general Possible-Word Constraint in segmentation. Johnson et al. (2003) found that 12-month-old infants listened longer to target words embedded in 'possible' contexts (leaving a syllable as a residue) than to words embedded in 'impossible' contexts (leaving a single consonant as a residue). The PWC, militating against single-consonant residues in segmentation, thus appears to play a role in segmentation by both adults and infants.

All of these studies provide evidence that infants use a variety of cues for the segmentation of continuous speech. There is also evidence that infants integrate these cues, and that infants attribute more importance to some cues than to other cues, when these are brought into conflict. For example, Johnson and Jusczyk (2001) show that both metrical cues and coarticulation cues are relied upon more strongly by 8-month-old infants than a segmentation strategy based on syllable probabilities. However, a study by Thiessen and Saffran (2003) shows that younger infants (7-month-olds) rely more strongly on distributional cues than on stress, suggesting that infants might initially rely on statistical cues in order to learn stress patterns. In addition, when cues are put in conflict, stress patterns have been found to override phonotactic patterns (Mattys et al., 1999). At the same time, phonotactics is required to refine metrical segmentation by determining the exact locations of word boundaries (e.g., Myers et al., 1996). It thus appears that segmentation cues join forces to optimize the detection of word boundaries in the continuous speech stream.

In sum, infants rely on various sublexical cues to extract wordlike units from continuous speech. Most of these cues are specific to the language to which the infant has been exposed during the first months of life. This raises an important issue. The task of segmenting a continuous speech stream into words is one that all infants face, regardless of their native language background. Nevertheless, most segmentation cues are language-specific. Infants thus face the challenge of acquiring segmentation cues from experience with the target language. Little is known about how infants tackle this problem. In order to obtain a thorough understanding of infants' early language development (specifically, the contribution of sublexical cues to learning a lexicon), it is essential to account for the acquisition of segmentation cues. Crucially, there has to be a *learning mechanism* that allows infants to acquire cues for segmentation. This mechanism needs to be able to learn segmentation cues *before the lexicon is in place*. The work described in this dissertation addresses this problem for the case of phonotactics.



### 1.3 THE INDUCTION OF PHONOTACTICS FOR SPEECH SEGMENTATION

#### 1.3 THE INDUCTION OF PHONOTACTICS FOR SPEECH SEGMENTATION

The main interest of this dissertation is in the learning mechanisms that are used by infants to acquire phonotactics. In addition, while a complete model of infant speech segmentation would require accounting for the integration of multiple segmentation cues (i.e., metrical cues, acoustic cues, phonotactic cues), this dissertation focuses on the contribution of phonotactics to solving the speech segmentation problem. By studying phonotactic cues in isolation, this dissertation assesses the contributions of different mechanisms to learning phonotactics, and to solving the speech segmentation problem. As will become clear later on, the resulting model of phonotactic learning is fundamentally different from earlier accounts of phonotactic learning, since those models assume a lexicon of word forms as the basis for phonotactic learning (e.g., Hayes & Wilson, 2008; Pierrehumbert, 2003).

##### 1.3.1 *Bottom-up versus top-down*

Although the segmentation studies clearly indicate a role for phonotactics in the segmentation of continuous speech by both adults and infants, they leave open the question of how the phonotactic patterns of the native language are acquired. Analyses of parents' estimation of their infant's receptive vocabulary indicate that infants have a vocabulary smaller than 30 words by the age of 8 months, which increases to about 75 words by the end of the first year (Fenson et al., 1994). Since infants by the age of 9 months already possess fine-grained knowledge of the phonotactic probabilities of their native language (Jusczyk et al., 1994), it seems unlikely that they derive this knowledge from their miniature lexicons. Rather, it seems that infants induce phonotactic patterns by monitoring sequences that they hear in their continuous speech input.

The bottom-up learning of phonotactics from continuous speech is in line with the view that infants use phonotactics to bootstrap into word learning. The general idea is that infants rely on phonotactics in order to learn words, and therefore do not rely on words to learn phonotactics. In order to avoid this potential chicken-and-egg problem, several computational studies have adopted a bottom-up view in which segmentation cues are derived from unsegmented input (e.g., Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Perruchet & Vinter, 1998; Swingley, 2005). This view is supported by evidence from infant studies suggesting that sublexical knowledge of phonotactics emanates from patterns in fluent speech, rather than from isolated lexical items (e.g., Mattys et al., 1999).

A bottom-up approach to phonotactic learning may seem counterintuitive at first. Phonological constraints (phonotactics, stress) typically refer to words

and word edges: ‘Words should start with a strong syllable’, ‘words should not start with /pf/’, etc. It would thus seem logical to assume that such knowledge is derived in a top-down fashion from stored word forms in the lexicon. Indeed, several models of phonotactic learning have been based on a view in which phonotactic constraints are learned from the lexicon (e.g., Hayes & Wilson, 2008; Pierrehumbert, 2001, 2003; Tjong Kim Sang, 1998). The initial lexicons that infants have, however, provide them with a rather limited source of data for phonotactic learning. While infants may know some words (e.g., *mommy*, *daddy*, and their own names) which may serve as top-down anchors in segmentation, the limited size of the infant vocabulary makes it plausible that infants initially rely on bottom-up information in speech segmentation (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Fenson et al., 1994). This makes sense when one considers the amount of input data that can be derived from an infant’s lexicon versus continuous speech input. In contrast to their small lexicons, infants are exposed to a vast amount of continuous speech input (see e.g., van de Weijer, 1998). It will be argued that this information is rich enough for the induction of phonotactics in a bottom-up fashion using mechanisms that have been shown to be available to infant language learners. Moreover, the induction of phonotactics from continuous speech allows the infant to make considerable progress in cracking the speech code.

A bottom-up learning procedure thus has two desirable properties. First, it is compatible with the demonstrated sensitivity and segmentation skills of infant language learners, who have not yet mastered the native language vocabulary. Second, a bottom-up approach allows for a beneficial role of phonotactics in segmentation and lexical acquisition. That is, the bottom-up learning of segmentation cues from continuous speech helps the learner to acquire words from speech. It has been argued that during language acquisition learners maximally exploit the phonology of their native language in order to optimize their segmentation skills (Cutler, Mehler, Norris, & Segui, 1986). In line with this view, the bottom-up approach assumes that infants maximally exploit the phonotactic structures that they encounter in continuous speech, since these phonotactic constraints allow them to segment the speech stream more efficiently.

### 1.3.2 *Computational modeling of phonotactic learning*

The main challenge for a bottom-up approach to phonotactic learning is the following: How can infants learn about the phonotactic structure of words without knowing the words themselves? This issue will be addressed by means of computational modeling. The algorithmic nature of computational models enable the researcher to formulate explicit theories of language acquisition.

Moreover, the predictions that follow from computational models allow for straightforward empirical testing of the theoretical proposal. Computational models of language acquisition that are based on psycholinguistic findings, and are tested empirically, therefore provide a valuable tool in the study of human learning mechanisms.

Importantly, computational models can be evaluated on natural language data, such as speech corpora. In order to allow full control of factors which might affect participants' performance, psycholinguistic experiments typically involve simplified and/or unnatural data (such as artificial sequences of CV syllables). When using accurately transcribed speech corpora, theoretical proposals can be tested on how they perform on more realistic data, including settings that involve variations that occur in the pronunciation of natural speech (such as acoustic reductions and assimilations). Computational models may therefore provide valuable insights which complement psycholinguistic findings that are obtained in laboratory settings.

Earlier work on computational modeling of phonotactic cues for segmentation (see Chapter 2) has revealed two different bottom-up approaches. Phonotactic segmentation models are either based on employing utterance boundaries (Brent & Cartwright, 1996; Daland, 2009), or on exploiting sequential probabilities in continuous speech (e.g., Cairns et al., 1997; Swingley, 2005). Utterance boundary models are based on the observation that utterance-initial segments are by definition also possible word-initial segments, simply because utterance beginnings are also word beginnings. The same line of reasoning applies to utterance-final segments. If the probability of a segment at an utterance boundary is interpreted as the probability at a word boundary, an estimation can be made with respect to the likelihood of word boundaries in continuous speech (Daland, 2009). That is,  $Prob(x.y) = Prob(x) \cdot Prob(.y)$ , where  $Prob(x.)$  and  $Prob(.y)$  are based on utterance boundaries.

Other models have taken a different approach by exploiting the occurrence of sequences *within* continuous speech utterances (rather than restricting the model to sequences that occur at utterance edges). That is, the probability of a word boundary can be estimated using the co-occurrence probability of two segments (i.e., 'biphone' probability) in continuous speech. A low biphone probability indicates a likely word boundary between the two segments in the biphone (e.g., Cairns et al., 1997).

The difference between the two approaches can be illustrated by the following example. Consider again the Dutch utterance /dələmpfil/. A model based on utterance boundaries would extract only the initial and final segments of the utterance, along with their alignment (/d/ and /l./). In contrast, a sequential model would extract all segment pairs within the utterance (/də/, /əl/, /lə/, /əm/, /mp/, /pf/, /fi/, /il/). While both types of information may

be useful for segmentation, the example shows that sequential phonotactics provides the learner with more data points. In this short utterance there are 8 biphone tokens. In contrast, the utterance has (by definition) only 2 edges. Utterance-internal sequences thus potentially constitute a rich source of information for phonotactic learning. At the very least, it seems unlikely that the learner would ignore all the information that is contained within the utterance itself. In fact, there is abundant evidence that infants indeed have the ability to extract purely sequential information from continuous speech. This ability has been demonstrated by a large body of psycholinguistic research on statistical learning from continuous speech by infants (discussed in Section 1.4).

There is a substantial gap, however, between phonotactic models that have been proposed for the purpose of speech segmentation, and phonotactic models that have been proposed in the field of linguistics. Phonologists have traditionally defined phonotactics in terms of positional constraints which operate on a hierarchical structure (e.g., Clements & Keyser, 1983; Selkirk, 1982). Specifically, the hierarchical structure states that words consist of syllables, and that syllables have constituents such as onsets, nuclei, and codas. For example, syllables in English can have the phoneme /ŋ/ as a coda, but not as an onset. Phonotactic constraints at the level of such positional constituents have been used in earlier studies, for example, as an account of wellformedness judgments of nonwords by adult participants (e.g., onset and rhyme probabilities, Coleman & Pierrehumbert, 1997). In addition, phonological theories assume that constraints are represented in terms of phonological features (e.g., Chomsky & Halle, 1968; Prince & Smolensky, 1993). Phonological theory thus assumes a much richer, abstract representation of phonotactics than the simple segment-based probabilities that are used in models of segmentation.

Recent developments in phonological research have led to models which account for the induction of abstract feature-based phonotactic constraints from input data (e.g., Albright, 2009; Hayes & Wilson, 2008, see Chapter 2). For example, Hayes and Wilson (2008) propose a model in which phonotactic constraints are selected from a space of possible constraints. Constraints are selected and incorporated into the grammar according to their accuracy (with respect to the lexicon of the language) and their generality. This dissertation aims to connect constraint induction models with phonotactic models of segmentation. Specifically, the sequential approach that has been used in earlier segmentation models will be adopted (e.g., Cairns et al., 1997; Swingley, 2005), and will be combined with feature-based generalization (e.g., Albright, 2009; Albright & Hayes, 2003). The approach results in abstract sequential constraints which indicate word boundaries in the speech stream.

An attractive property of sequential constraints is that they are learnable using mechanisms that have received much attention in infant studies (e.g.,

statistical learning, Saffran, Aslin, & Newport, 1996). Building on these findings, knowledge of phonotactics will take the form of constraints stating which sequences should be avoided within words, and which sequences should be preserved (see Chapter 2 for a more elaborate description of the structure of constraints). The relevance of such phonotactic constraints for segmentation was first noted by Trubetzkoy, who referred to these constraints as *positive phonematische Gruppensignale* and *negative phonematische Gruppensignale*, respectively (Trubetzkoy, 1936).

In line with earlier models of constraint induction, an encoding of phonotactic constraints in terms of phonological features will be assumed (e.g., Albright, 2009; Hayes & Wilson, 2008). Albright (2009) shows that a model with feature-based generalizations is able to account for human judgments of nonword wellformedness. In his model, the learner abstracts over sequences of segments and extracts the shared phonological feature values to construct feature-based generalizations. Importantly, while segment-based models do not generalize to unattested clusters, the feature-based model has proven to be a good predictor for both attested and unattested clusters. As will be discussed in more detail in Section 1.4, several infant studies provide evidence that is consistent with a feature-based view of phonotactic constraints (Cristià & Seidl, 2008; Maye, Weiss, & Aslin, 2008; Saffran & Thiessen, 2003). The approach taken in this dissertation is to maximally exploit the benefits of sequential constraints and feature-based generalization for the induction of phonotactic constraints, and for the detection of word boundaries in continuous speech.<sup>1</sup>

In sum, this dissertation will investigate whether the statistical learning of biphone constraints from continuous speech can be combined with the construction of more abstract feature-based phonotactic constraints. It will be shown that generalizations over sequential constraints allow the learner to make considerable progress with respect to phonotactic learning and the formation of a mental lexicon. As will become clear in the next section, the approach is supported by findings on infant learning mechanisms. In order to investigate the linguistic relevance of the model, Chapter 4 addresses the issue of whether the model can learn a sequential phonotactic constraint (originally proposed in research on theoretical phonology) restricting the co-occurrence of segments with the same place of articulation (OCP-PLACE, Frisch, Pierrehumbert, & Broe, 2004; McCarthy, 1988).

---

<sup>1</sup> An interesting characteristic of the model is that the sequential constraints that are induced by the model can give rise to positional (alignment) effects without explicitly encoding word edges in the structure of constraints. See Chapter 4.

1.3.3 *Two hypotheses regarding the acquisition of phonotactics by infants*

The bottom-up view sketched above can be summarized as follows: Infants induce phonotactic constraints from continuous speech, and use their acquired knowledge of phonotactics for the detection of word boundaries in the speech stream. In this view, emerging phonotactic knowledge informs the learner's search for words in speech. The view implies that at least some phonotactic learning precedes segmentation and word learning. The result is a proto-lexicon based on bottom-up segmentation cues. This view will be referred to as the Speech-Based Learning (SBL) hypothesis:

(1.3) *Speech-Based Learning hypothesis:*

*continuous speech* → *phonotactic learning* → *segmentation* → *proto-lexicon*

An alternative hypothesis can be formulated in which the dependency between phonotactic learning and word learning is reversed. That is, one could argue that the initial contents of the infant's lexicon are used to bootstrap into phonotactic learning (rather than vice versa). This top-down view resembles the approach taken in earlier models of phonotactic learning. Models of phonotactic learning have typically assumed that phonotactic constraints are the result of abstractions over statistical patterns in the lexicon (Frisch et al., 2004; Hayes & Wilson, 2008). While these studies have not been concerned with the topic of early language development, the developmental view that would follow from such models is that phonotactics is acquired from a proto-lexicon, rather than from the continuous speech stream. In this case the proto-lexicon could, for example, have been formed using alternative segmentation mechanisms which do not involve phonotactics. This alternative hypothesis will be called the Lexicon-Based Learning (LBL) hypothesis:

(1.4) *Lexicon-Based Learning hypothesis:*

*continuous speech* → *segmentation* → *proto-lexicon* → *phonotactic learning*

This dissertation aims to promote the Speech-Based Learning hypothesis. In order to provide evidence for the SBL hypothesis, the dissertation proposes a learning model which induces phonotactics from continuous speech (Chapter 2). The model is implemented as a computer program, and is based on learning mechanisms that have been shown to be available to infant language learners. The model produces a set of phonotactic constraints which can be used to discover word boundaries in continuous speech. While this model shows the potential of learning phonotactics from continuous speech, it leaves open the question of how plausible this approach is as an account of infant phonotactic learning. The plausibility will be addressed via various empirical

#### 1.4 LEARNING MECHANISMS IN EARLY LANGUAGE ACQUISITION

studies which aim to provide support for the model. These studies test the viability of the SBL hypothesis by measuring the extent to which it helps the learner in discovering word boundaries (Chapter 3). In other words, does continuous speech contain enough information to learn phonotactics which may serve to segment speech? Computer simulations are conducted which specifically aim to demonstrate the added value of phonological generalizations in segmentation. In addition, a series of simulations examines the extent to which the bottom-up account can learn phonotactic constraints which have been studied in the traditional phonological literature (Chapter 4). Finally, the plausibility of the approach is tested against human data. The bottom-up (speech-based) versus top-down (lexicon-based) learning approaches are compared with respect to their ability to account for human segmentation data (Chapter 4). In addition, a series of experiments with adult participants provides a final test for the induction of novel phonotactics from continuous speech by human learners (Chapter 5). These experiments have been set up in such a way as to maximally reduce possible top-down influences, and thus provide strong support for the claim that human learners can induce phonotactics in a bottom-up fashion.

In order to make claims about infant language learning by means of computational modeling, it is essential to look at learning mechanisms which have been shown to be available to infants. Ignoring such evidence could easily lead to a very efficient learning model, but would not necessarily tell us anything about how infants acquire phonotactic constraints for speech segmentation. Specifically, the implementation of a bottom-up approach to phonotactic learning requires a careful consideration of mechanisms which can operate on continuous speech input. The dissertation thus takes a computational angle by investigating how learning mechanisms which have been shown to be available to infants might interact in the learning of phonotactic constraints from continuous speech. A large amount of psycholinguistic research points towards important roles for two learning mechanisms: statistical learning and generalization. Below the relevance of these mechanisms for infant language acquisition will be discussed. The mechanisms will form the basis of the computational model that will be presented in Chapter 2.

#### 1.4 LEARNING MECHANISMS IN EARLY LANGUAGE ACQUISITION

Given the input that is presented to the infant, how is the input processed, and under which conditions does the infant derive phonotactic knowledge from the input? From a computational point of view, what is the algorithm underlying infants' learning behavior? Understanding the mechanisms involved in acquiring linguistic knowledge from input data is perhaps the most

challenging part in the study of infant language acquisition. While two main mechanisms have been identified, many questions remain with respect to how these mechanisms operate. Bringing together these mechanisms in a computer model brings us one step closer to understanding infant language acquisition.

#### 1.4.1 *Statistical learning*

A well-studied learning mechanism is statistical learning. Infants have the ability to compute the probability with which certain linguistic units co-occur. For example, infants compute transitional probabilities of adjacent syllables when listening to speech from an artificial language (Aslin, Saffran, & Newport, 1998; Goodsitt, Morgan, & Kuhl, 1993; Saffran, Aslin, & Newport, 1996). Such probabilistic information may assist the segmentation of continuous speech. Transitional probabilities are generally higher for sequences within words than for sequences between words. For example, in the phrase *pretty baby*, the transitional probability from *pre* to *ty* is higher than the probability from *ty* to *ba*. Low probabilities are thus indicative of word boundaries. Saffran, Aslin, and Newport (1996) exposed 8-month-old infants to a 2-minute nonsense stream of continuous speech (e.g., *bidakupadotigolabubidaku...*) in which the transitional probabilities between syllable pairs had been manipulated. Infants were able to distinguish ‘words’ (containing high probability sequences) from ‘part-words’ (containing a low probability sequence, and thus straddling a statistical boundary), indicating that statistical learning is used to decompose the speech stream into word-like units.

Many studies subsequently have shown that statistical learning is a domain-general learning mechanism, allowing for the learning of dependencies of various representational units. These units can be linguistic in nature (e.g., consonants and vowels, Newport & Aslin, 2004), or can be non-linguistic units, such as musical tones (Saffran, Johnson, Aslin, & Newport, 1999) and visual stimuli (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002). Recent evidence shows that infants can learn the statistical dependency in two directions. Although transitional probabilities are typically calculated as forward conditional probabilities (i.e., the probability of *y* given *x*, e.g., Pelucchi, Hay, & Saffran, 2009b), there is evidence that both adult and infant learners can learn backward probabilities (i.e., the probability of *x* given *y*, Pelucchi, Hay, & Saffran, 2009a; Perruchet & Desaulty, 2008; Saffran, 2002). These studies indicate that statistical learning is a versatile and powerful mechanism which may play an important role in language acquisition.

Several sources of evidence indicate that statistical learning could also be involved in infants’ learning of phonotactics. One source of evidence concerns a prerequisite for phonotactic learning, namely that infants should be able



to *perceive* segments. A study by Maye, Werker, and Gerken (2002) shows that 6- and 8-month-olds are able to discriminate between phonetic categories after exposure to bimodal distributions of phonetic variation. While infants may not yet have acquired the full segment inventory, the ability to perceive at least some phonetic categories at a young age allows for the possibility that infants learn the co-occurrence probabilities of such categories. More direct evidence for the role of statistical learning in acquiring phonotactics comes from studies showing that 9-month-old infants are sensitive to native language probabilistic phonotactics (Jusczyk et al., 1994; Mattys & Jusczyk, 2001b). These studies suggest that infants are capable of learning segment co-occurrence probabilities. The ability to learn statistical dependencies between segments has also been demonstrated by White, Peperkamp, Kirk, and Morgan (2008), who found that 8.5-month-old infants' responses in learning phonological alternations were driven by segment transitional probabilities.

It should be mentioned that statistical learning by itself is not a full theory of language acquisition (e.g., Soderstrom, Conwell, Feldman, & Morgan, 2009; Johnson & Tyler, 2010). While the calculation of co-occurrence probabilities can provide a useful means for detecting linguistic structures in the input, it does not lead to the more abstract linguistic representations that are commonly assumed in linguistic theories of language acquisition (e.g., Chomsky, 1981; Prince & Tesar, 2004; Tesar & Smolensky, 2000; see also, Jusczyk, Smolensky, & Allocco, 2002). Such linguistic frameworks assume that there is abstract linguistic knowledge which traces back to the learner's innate endowment (Universal Grammar). Studies of infant learning mechanisms, however, indicate that there may be a second learning mechanism which may provide a means to learn abstract knowledge without referring to the innateness hypothesis. It will be argued that, at least for the case of phonotactic learning, a generalization mechanism can join forces with the statistical learning mechanism to allow for the induction of more abstract constraints. Such a mechanism thus may provide a crucial addition to statistical approaches to language acquisition, allowing for an emergentist view that maintains the assumption of linguistic abstractness (e.g., Albright & Hayes, 2003; Hayes & Wilson, 2008).

#### 1.4.2 *Generalization*

In addition to statistical learning, there appears to be a role for a different form of computation in phonotactic acquisition. This learning mechanism allows the infant to generalize beyond observed (co-)occurrence probabilities

## INTRODUCTION

to new, unobserved instances.<sup>2</sup> Generalizations learned by infants can take on the form of phonetic categories (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Maye et al., 2002, 2008; Werker & Tees, 1984), phonotactic patterns (e.g., Chambers, Onishi, & Fisher, 2003; Saffran & Thiessen, 2003), lexical categories (e.g., Gómez, 2002; Gómez & Lakusta, 2004; Gómez & Maye, 2005), and artificial grammars (e.g., Gómez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999).

Gómez and Gerken (1999) show that infants are able to generalize over observed training utterances to new, unobserved word strings. In their study, 12-month-old infants were exposed to word strings generated by a finite-state grammar. After a familiarization of only 2 minutes, infants were able to distinguish between grammatical and ungrammatical items. Importantly, one of their experiments used test items that had not occurred during familiarization. As a consequence, transitional probabilities between word pairs were zero for both grammatical and ungrammatical sequences. Infants generalized from the word strings heard during familiarization (e.g., *fim sog fim fim tup*) to word strings from a novel vocabulary (e.g., *pel tam pel pel jic*), indicating that they had learned not just the specific word pairs, but also more abstract structures (e.g., Berent, Marcus, Shimron, & Gafos, 2002; Marcus et al., 1999; cf. Altmann, 2002). These results suggest that generalization may play a major role in infant language acquisition, in addition to statistical learning.

Saffran and Thiessen (2003) showed that 9-month-old infants can induce phonotactic patterns that are more general than the occurrence patterns of the specific phonological segments to which they were exposed. During a pattern induction phase, the infant was familiarized with the phonotactic regularity. Familiarization was followed by a segmentation phase, in which infants could segment novel words from a continuous speech stream by employing the phonotactic pattern to which they had been familiarized. Finally, the test phase served to determine whether the infant indeed was able to distinguish novel test items which conformed to the phonotactic pattern from test items which did not conform to the pattern. Infants acquired a phonotactic generalization about the positional restrictions on voiced and voiceless stops after a brief training period. In contrast, they could not learn patterns of segments which were not phonetically similar. These results indicate that there is more to phonotactic acquisition than the learning of constraints on the co-occurrences of specific segments. It seems that infants are able to abstract over the similarity between segments to construct phonotactic generalizations.

---

<sup>2</sup> The term 'generalization' will be used throughout the dissertation as a broad term to indicate the learner's ability to process items based on similarity to familiar items. The question of whether infants store generalizations as abstract representations, or, for example, recompute generalizations repeatedly from stored exemplars during online speech processing is considered an open issue.

Evidence for infants' generalization on the basis of phonetic similarity (voicing, manner of articulation, etc.) has been reported in various studies (e.g., Jusczyk, Goodman, & Baumann, 1999; White et al., 2008; Maye et al., 2008). Little is known, however, about how phonotactic generalizations are represented by infants. Traditionally, generative phonologists have assumed that generalizations can be stated in terms of abstract features (Chomsky & Halle, 1968). Several recent studies have indeed argued that infants form generalizations that are abstract at the level of the feature. For example, Maye et al. (2008) show that exposing 8-month-old infants to a bimodal distribution of Voice Onset Time (VOT) for one place of articulation (e.g., *da/ta* contrast for dentals) facilitates the discrimination of a voicing contrast for a different place of articulation (e.g., *ga/ka* contrast for velars). This finding is consistent with the view that infants construct a generalization at the level of an abstract feature *voice* as a result of exposure to a specific contrast along the VOT dimension.

In a study on phonotactic learning, Cristià and Seidl (2008) exposed 7-month-old infants to CVC words from an artificial language that either had onsets with segments that formed a natural class (plosives + nasals, which share the specification '–continuant') or onsets with segments that did not form a natural class (fricatives + nasals, which do not form a coherent class). During the test phase, infants listened to words that contained novel plosives and fricatives, not heard during training. When trained on the natural class, infants distinguished between words that did or did not conform to the phonotactic structure of the training language. In contrast, infants trained on the incoherent class were not able to distinguish legal from illegal test items. In a control experiment, they showed that the difference in learning difficulty was not due to an inherent difference between plosives and fricatives. These results provide evidence that infants learn phonotactic constraints which are more general than the specific segments that were used during training, and support the view that infants represent phonotactic generalizations at the level of phonological features.

These studies raise the question of how such features could become available to infants. Specifically, features could either be innate, or learned from exposure to the native language. It should be noted that features are not necessarily innate (e.g., Mielke, 2008). Abstract phonological features have acoustic and perceptual correlates in the speech signal (e.g., Stevens, 2002), which is an important prerequisite if one wants to argue that features can be learned from input data. In addition, while the traditional function of features in phonological theory has been to distinguish the meanings of words in the lexicon (e.g., the words *pet* and *bet* are distinguished by the feature *voice*), attempts have been made to induce abstract features from raw acoustic input

data (Lin & Mielke, 2008). This indicates the possibility of acquiring features before the lexicon is in place.

#### 1.4.3 *Towards a unified account of infant learning mechanisms*

Statistical learning allows the learner to accumulate frequency data (conditional probabilities) over observed input. A second mechanism, generalization, allows the learner to abstract away from the observed input, leading to the formation of categories, patterns, and grammars. While the importance of both learning mechanisms for language acquisition has been widely acknowledged (Endress & Mehler, 2009; Gómez & Gerken, 1999; Marcus et al., 1999; Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; Toro, Nespor, Mehler, & Bonatti, 2008; White et al., 2008), surprisingly little is known about how these two mechanisms, statistical learning and generalization, interact. For example, how can statistical learning provide a basis for the construction of generalizations? How do generalizations affect the probabilistic knowledge of the learner? Do the generalizations in any way reflect abstract constraints that have been studied in the field of linguistics? Explicit descriptions and analyses of interactions between statistical learning and generalization would greatly enhance our understanding of language acquisition.

In addition, while infants' capacity to use probabilistic phonotactics in speech segmentation has been demonstrated (Mattys & Jusczyk, 2001b), it is not clear whether infants also use phonotactic generalizations to discover words in continuous speech. Although the study by Saffran and Thiessen explores this possibility by presenting infants with a segmentation task after familiarization, the authors themselves mention that the infants may have applied the patterns that were induced during familiarization to the test items directly, i.e. regardless of the segmentation task (Saffran & Thiessen, 2003, p.487). Infants' use of phonotactic generalizations in speech segmentation thus remains to be demonstrated. This dissertation addresses this issue indirectly by determining the potential use of such generalizations. That is, the benefits of both segment-specific and more abstract, feature-based constraints in speech segmentation will be assessed. The dissertation thus aims at providing an account of phonotactic learning which includes a role for both phonotactic probabilities and phonological similarity. The research presented in this dissertation thereby offers insight into the learning mechanisms and segmentation strategies that play a role in language acquisition.

A final note concerns the input on which the two mechanisms operate. It should be stressed that, while the learning of co-occurrence probabilities from continuous speech has been widely acknowledged, the learning of linguistic generalizations from continuous speech has not earlier been proposed. In

## 1.5 RESEARCH QUESTIONS AND SCIENTIFIC CONTRIBUTION

fact, there have been studies arguing that generalizations cannot be learned from continuous speech, and that word boundaries are a prerequisite for the learning of generalizations (Peña, Bonatti, Nespor, & Mehler, 2002; Toro et al., 2008). One of the goals of the dissertation is to explore whether phonotactic generalizations can be based on sublexical regularities in continuous speech input, rather than on word forms in the lexicon.

### 1.5 RESEARCH QUESTIONS AND SCIENTIFIC CONTRIBUTION

The main issue in the dissertation is:

- *How do language learners induce phonotactic constraints for speech segmentation?*

This issue will be addressed by proposing a computational model (based on the Speech-Based Learning hypothesis), and subsequently providing evidence for the model using simulations involving computational learners, and psycholinguistic experiments involving human learners. The research questions that form the core of the dissertation are the following:

#### 1. Computational learners

- Can phonotactic constraints be induced from continuous speech using the mechanisms of statistical learning and generalization?
  - *What is the role of statistical learning in the induction of phonotactic constraints?*
  - *What is the role of generalization in the induction of phonotactic constraints?*
  - *To what extent do induced constraints resemble the sequential constraints that have appeared in the traditional phonological literature?*
- How do the different learning mechanisms affect the segmentation of continuous speech?
  - *Do feature-based generalizations improve the learner's ability to detect word boundaries in continuous speech?*

#### 2. Human learners

- Do human learners induce phonotactics from continuous speech?
  - *Do adult participants learn specific (segment-based) constraints from continuous speech?*

## INTRODUCTION

- *Do adult participants learn abstract (feature-based) constraints from continuous speech?*
- What kind of phonotactic knowledge is used by human learners for the segmentation of continuous speech?
  - *Do adult learners use specific constraints, abstract constraints, or both?*
  - *Do adult learners use constraints induced from the lexicon, or constraints induced from continuous speech?*

### *Scientific contribution*

The dissertation contributes to the general field of cognitive science by taking an approach to language acquisition which incorporates ideas from both computational linguistics, theoretical linguistics, and psycholinguistics. That is, a computational model of phonotactic learning is presented which is supported by psycholinguistic findings, and which produces phonotactic constraints which have a linguistic interpretation. The dissertation thus combines and compares the outcomes of computer simulations, linguistic analyses, and psycholinguistic experiments. The contribution of the dissertation to each of these different perspectives on language acquisition is briefly specified below.

#### *(i) Contribution to computational linguistics*

A computational model of speech segmentation is presented, which can be applied to transcribed speech corpora. Several segmentation models have been proposed which make use of probabilistic phonotactics (e.g., Cairns et al., 1997; Brent, 1999a). The current study complements such studies by adding a generalization component to the statistical learning of phonotactic constraints (Chapter 2), and by quantifying the benefits of abstract phonotactic constraints for speech segmentation (Chapter 3). In addition, the simulations reported in this dissertation are valuable since they involve fairly accurate representations of spoken language. While earlier studies typically used orthographic transcriptions that were transformed into canonical transcriptions using a phonemic dictionary, the simulations reported in this dissertation involve speech which has been transcribed to a level which includes variations that are typically found in the pronunciation of natural speech, such as acoustic reductions and assimilations.

In order to allow for replication of the findings, and to encourage further research on phonotactic learning and speech segmentation, a software package has been created, which has been made available online:

<http://www.hum.uu.nl/medewerkers/f.w.adriaans/resources/>

The software can be used to run phonotactic learning and speech segmentation simulations, as described throughout this dissertation. In addition, it allows the user to train models on new data sets, with the possibility to use other statistical measures, thresholds, and a different inventory of phonological segments and features. The package implements several learning models (STAGE, transitional probability, observed/expected ratio) and segmentation models (OT segmentation, threshold-based segmentation, trough-based segmentation). The details of these models are explained in Chapters 2 and 3. The software package is accompanied by a manual which explains the user-defined parameters, and explains how the model can be applied to new data sets (see Appendix A).

*(ii) Contribution to theoretical linguistics*

The dissertation provides a learnability account of phonological constraints, rather than assuming innate constraints. While most previous accounts of phonotactic learning defined the learning problem as finding the appropriate ranking for a universal set of constraints (e.g., Tesar & Smolensky, 2000; Prince & Tesar, 2004), the learning model presented here learns the constraints themselves, as well as their rankings. The research thereby adds to a growing body of research which aims at minimizing the role of Universal Grammar in phonological acquisition, while still acknowledging the existence of abstract representations and constraints ('constraint induction'; e.g., Hayes, 1999; Albright & Hayes, 2003; Hayes & Wilson, 2008). The induction approach is demonstrated in a case study on the induction of OCP-PLACE, a phonological constraint restricting the co-occurrence of consonants sharing place of articulation across intervening vowels (Chapter 4).

An important difference with earlier models of constraint induction is that the proposed model is unsupervised. Previous models have been based on inducing the phonotactic structure of words from actual forms (lemmas, onsets) in the lexicon (e.g., Albright, 2009; Hayes & Wilson, 2008), which is a case of supervised learning. Since it is assumed that the learner has not yet acquired a lexicon (or that the learner's lexicon is too small to support the learning of phonotactic constraints), the learner has no way of determining how good, or useful, the resulting generalizations will be. In contrast, supervised models induce phonotactic constraints through evaluation of the accuracy of generalizations with respect to the lexicon. The unsupervised approach results in a constraint induction model for the initial stages of language acquisition.

*(iii) Contribution to psycholinguistics*

Focusing on human speech-based learning capacities, the dissertation investigates whether human learners can induce novel phonotactic constraints from continuous speech (Chapter 5). This issue is addressed in a series of artificial language learning (ALL) experiments with adult participants. The learning conditions of these experiments present a greater challenge to participants than in earlier studies. The ALL experiments contribute to earlier phonotactic learning experiments (e.g., Onishi, Chambers, & Fisher, 2002) by presenting participants with a continuous speech stream, rather than with isolated words. The experiments contribute to earlier speech-based learning experiments (e.g., Bonatti, Peña, Nespor, & Mehler, 2005; Newport & Aslin, 2004) by presenting participants with artificial languages that display relatively small differences in ‘within-word’ and ‘between-word’ consonant transitional probabilities. In addition, the languages contain randomly inserted vowels, and thus do not contain reoccurring word forms. Learners’ ability to learn the phonotactic structure of these languages is evaluated by testing for generalization to novel words (which had not occurred in the familiarization stream). A side issue of interest is that the experiments add to a growing body of experimental research demonstrating a large influence of the native language phonotactics on the segmentation of a novel speech stream (Boll-Avetisyan & Kager, 2008; Finn & Hudson Kam, 2008; Onnis, Monaghan, Richmond, & Chater, 2005).

## 1.6 OVERVIEW OF THE DISSERTATION

*Chapter 2: A computational model of phonotactic learning and segmentation*

This chapter introduces the OT segmentation model, which is a modified version of Optimality Theory, and which is used for regulating interactions between phonotactic constraints in speech segmentation. The constraint induction model, STAGE, provides an account of how language learners could induce phonotactic constraints from continuous speech. The model assumes that the learner induces biphone constraints through the statistical analysis of segment co-occurrences in continuous speech (Frequency-Driven Constraint Induction, FDCI), and generalizes over phonologically similar biphone constraints to create more general, natural class-based constraints (Single-Feature Abstraction, SFA).

FDCI employs two statistical thresholds (on the observed/expected ratio) to distinguish between high and low probability phonotactics. High probability biphones trigger the induction of a contiguity constraint (CONTIG-IO( $xy$ )) which states that the biphone should be kept intact during speech processing. Conversely, low probability biphones result in the induction of a markedness



constraint ( $*xy$ ), which states that the biphone should be broken up through the insertion of a word boundary. No constraints are induced for biphones that have neutral probability (more formally, observed  $\approx$  expected).

SFA inspects the similarity between biphone constraints, which is quantified as the number of shared phonological feature values. In case of a single feature difference between constraints, the learner constructs a more general constraint which disregards this feature. The result is that the generalizations cover sequences of natural classes rather than sequences of specific segments. The main benefit for the learner is that the abstract constraints make a segmentation statement about biphones of neutral probability. Phonological similarity thus complements statistical learning in speech segmentation. As a result of (over-)generalization, conflicts arise between markedness and contiguity constraints. Such conflicting segmentation predictions are resolved by numerically ranking the constraints and evaluating constraint violations according to the principle of strict domination.

A worked out example illustrates how STAGE and the OT segmentation model together can provide a mapping from continuous speech to segmented speech through the induction of phonotactic constraints. The chapter shows that phonotactic constraints can be learned from continuous speech in a psychologically motivated and formally explicit way. The examples show that the model has a straightforward linguistic interpretation.

### *Chapter 3: Simulations of segmentation using phonotactics*

This chapter addresses the ability of STAGE to detect word boundaries in continuous speech. Specifically, the question whether feature-based generalizations improve the segmentation performance of the learner is addressed in a series of computer simulations. The simulations address the induction of native language (henceforth, L1) Dutch constraints on biphone occurrences. Different learning models are evaluated on transcriptions of Dutch continuous speech. The crucial comparison is between STAGE, which acknowledges a role for both statistical learning and generalization, and segmentation models which rely solely on statistical learning. Experiment 1 compares the segmentation performance of STAGE to statistical learning models that use threshold-based segmentation. Experiment 2 tests a wide range of statistical thresholds for both STAGE and the statistical learning models. Experiment 3 compares the segmentation performance of STAGE to statistical learning models that use trough-based segmentation. All three experiments zoom in on the complementary roles of statistical learning and generalization, addressing both the need for statistical thresholds and the need for generalization. Finally,

## INTRODUCTION

Experiment 4 shows how the model's segmentation performance develops as a function of input quantity.

The main finding of Experiments 1-3 is that STAGE outperforms purely statistical models, indicating a potential role for phonotactic generalizations in speech segmentation. The experiments also show that the induction thresholds are necessary to create a good basis for the construction of phonotactic generalizations. That is, generalization only improves segmentation performance when a reliable distinction between high and low probability phonotactics is made. Experiment 4 shows that the learning of contiguity constraints precedes the learning of markedness constraints. That is, the model learns where not to put boundaries first, and later learns where to put boundaries. This finding is in line with developmental studies showing that infants under-segment: They learn larger (proto-)words first, which are later broken down into smaller word chunks.

Taken together, the experiments provide support for the Speech-Based Learning hypothesis, in which phonotactic learning facilitates the development of the mental lexicon, rather than vice versa. The simulations show that segmentation benefits from both abstract and specific constraints.

### *Chapter 4: Modeling OCP-PLACE and its effect on segmentation*

This chapter addresses two issues: (i) To what extent can the model induce constraints which have been proposed in theoretical phonology? (ii) To what extent can the model account for human segmentation behavior? The chapter provides an important contribution to the dissertation by looking at the linguistic relevance of the constraints that are learned through generalization, and by looking at the psychological plausibility of those constraints. Specifically, it is investigated whether STAGE can provide a learnability account of an abstract phonotactic constraint which has been shown to affect speech segmentation: OCP-PLACE. OCP-PLACE is a phonological constraint which states that sequences of consonants sharing place of articulation should be avoided. The chapter connects to the work of Boll-Avetisyan and Kager (2008) who show that OCP-PLACE affects the segmentation of continuous artificial languages by Dutch listeners.

Due to the underrepresentation of several specific labial-labial pairs (across vowels), the model induces feature-based generalizations which cause a general avoidance of labial-labial pairs. However, the constraint set does not exactly mimic the predictions of OCP-PLACE. Two studies address how well the learned constraint set reflects Dutch listeners' phonotactic knowledge of non-adjacent consonant dependencies. It was found that generalization improves the fit to the human segmentation data of Boll-Avetisyan and Kager

(2008), as compared to models that rely exclusively on consonant probabilities (Experiment 1). Since the approach also outperforms a single, pre-defined OCP-PLACE, the results suggest that the success of the approach can be attributed to the mix of specific and abstract constraints. In addition, it was found that a continuous-speech-based learner has a comparable fit to the human data to a learner based on word types (Experiment 2). Again, the success of both models can be attributed to the mixture of specific and general constraints. Interestingly, a token-based learner fails, regardless of the specific threshold configuration that is used.

The chapter provides the second piece of evidence (in addition to the simulations in Chapter 3) that feature-based generalization plays a role in segmentation. More specifically, the simulations of human segmentation data provide additional evidence that segmentation involves both abstract and specific constraints. It is again demonstrated that continuous utterances could be used by learners as a valuable source of input for phonotactic learning. The SBL hypothesis is thus supported by an account of human segmentation behavior.

#### *Chapter 5: The induction of novel phonotactics by human learners*

Building on the assumptions and findings from Chapter 4, the SBL hypothesis is tested with human participants. Specifically, the chapter investigates whether human learners can learn novel phonotactic structures from a continuous speech stream. Using the artificial language learning (ALL) approach, the chapter provides a more direct test for the psychological plausibility of the phonotactic learning approach. In addition, the chapter provides a test for the assumption that learners can induce phonotactic constraints on consonants *across intervening vowels* from continuous speech.

Continuous artificial languages are constructed in such a way that they exhibit both statistical and phonological structure. Experiment 1 is an attempt to extend earlier studies to a case involving more word frames, more words, and smaller differences in TP, but with greater phonological similarity. The test phase focuses on the statistical learning component. Experiment 2 drops the notion of ‘word’ altogether (by using randomly selected vowels), thereby focusing on purely phonotactic learning, rather than word learning. The experiment tests for generalization to novel ‘words’ with the same phonotactic (consonant) structure. Finally, Experiment 3 tests whether human participants perform feature-based generalization on top of the statistical learning of consonant dependencies. The experiment tests for generalization to novel segments from the same natural class.

## INTRODUCTION

Experiment 1 shows a large influence of L1 phonotactics on artificial language segmentation. While this finding is in line with the findings from Chapter 4, it hinders the induction of novel phonotactic constraints. That is, no learning effect is found. Experiment 2 shows that, when carefully controlling for L1 factors, human learners can indeed learn novel phonotactics from continuous speech. The significant learning effect gives direct support for the SBL learning hypothesis: It shows phonotactic learning from continuous speech (i.e., without referring to a lexicon). In addition, it provides additional evidence for the learning of phonotactic constraints across intervening vowels. Furthermore, the experiment demonstrates learners' capacity to generalize phonotactics, learned from continuous speech, to novel words. This indicates that the acquired phonotactic knowledge is represented outside of the mental lexicon, and is therefore not merely a by-product of word forms in the lexicon. Whereas the results of Chapter 4 could also be explained by a type-based learner, the results in Experiment 2 cannot be attributed to lexicon-based learning. Experiment 3, however, failed to provide evidence for feature-based generalization to novel segments. Possible explanations for this null result are discussed.

### *Chapter 6: Summary, discussion, and conclusions*

The final chapter summarizes the main points of the dissertation, critically assesses the findings, and provides suggestions for future work in this area. The main accomplishment of the dissertation is the demonstration that phonotactics can be learned from continuous speech by combining mechanisms that have been shown to be available to infant and adult language learners. Together, the mechanisms of statistical learning and generalization allow for the induction of a set of phonotactic constraints with varying levels of abstraction, which can subsequently be used to successfully predict the locations of word boundaries in the speech stream. In doing so, the model provides a better account of speech segmentation than models that rely solely on statistical learning. This result was found consistently throughout the dissertation, both in computer simulations and in simulations of human segmentation data. With respect to human learning capacities, the dissertation shows that participants can learn novel segment-based phonotactic constraints from a continuous speech stream from an artificial language. By combining computational modeling with psycholinguistic experiments, the dissertation contributes to our understanding of the mechanisms involved in language acquisition.

## A COMPUTATIONAL MODEL OF PHONOTACTIC LEARNING AND SEGMENTATION<sup>1</sup>

---

In this chapter a computational model is proposed for the induction of phonotactics, and for the application of phonotactic constraints to the segmentation of continuous speech. The model uses a combination of segment-based statistical learning and feature-based generalization in order to learn phonotactic constraints from transcribed utterances of continuous speech. The constraints are interpreted in a segmentation model based on the linguistic framework of Optimality Theory (OT). In Section 2.1, an overview is given of previous work on computational modeling of speech segmentation and phonotactic learning. Section 2.2 presents the OT segmentation model, which formalizes how phonotactic constraints can be used to detect word boundaries in continuous speech. The learning model, STAGE, is presented in Section 2.3. A worked out example in Section 2.4 illustrates the types of constraints that are learned by the model. Finally, some implications of the model are discussed in Section 2.5.<sup>2</sup>

### 2.1 INTRODUCTION

Computational models of speech segmentation are typically trained and tested on transcribed utterances of continuous speech. The task of the model is either to learn a lexicon directly, through the extraction of word-like units, or to learn to predict when a word boundary should be inserted in the speech stream. Here the latter task will be discussed, focusing on models that learn phonotactics in order to detect word boundaries in continuous speech. (For more general overviews of computational models of segmentation, see Batchelder, 2002; Brent, 1999b.) Various segmentation models have been proposed that make use of either phonotactics based on utterance boundaries (Brent & Cartwright, 1996), or phonotactics based on sequence probabilities (e.g., Cairns et al., 1997).

---

<sup>1</sup> Sections of this chapter were published in: 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.

<sup>2</sup> The model can be downloaded from:  
<http://www.hum.uu.nl/medewerkers/f.w.adriaans/resources/>

2.1.1 *Models of speech segmentation using phonotactics*

Brent and Cartwright (1996) propose that phonotactic constraints can be learned through inspection of consonant clusters that appear at utterance boundaries. The model is based on the observation that clusters at the edges of utterances are necessarily also allowed at the edges of words. Their segmentation model evaluates candidate utterance parsings using a function based on Minimum Representation Length (MRL). The phonotactic constraints act as a filter to eliminate utterance parsings which would produce phonotactically ill-formed words. A disadvantage of this approach is that it is based on categorical phonotactics. That is, a cluster is either allowed or not, based on whether it occurs at least once at an utterance boundary. As a consequence, the approach is rather vulnerable to occurrences of illegal clusters at utterance edges (which may occur as a result of acoustic reductions). In general, categorical phonotactics fails to make a prediction in cases of ambiguous clusters in the speech stream, since such clusters have multiple phonotactically legal interpretations. Moreover, infants have been demonstrated to use sequential phonotactic probabilities to segment speech (Mattys et al., 1999; Mattys & Jusczyk, 2001b). An approach based on categorical phonotactics derived from utterance boundaries therefore at best only provides a partial explanation of infants' knowledge of phonotactics. (See Daland, 2009, for a more promising, probabilistic approach to using utterance boundaries for segmentation.)

Models that rely on sequential phonotactic cues either explicitly implement segment co-occurrence probabilities (e.g., Brent, 1999a; Cairns et al., 1997), or use neural networks to learn statistical dependencies (Cairns et al., 1997; Christiansen, Allen, & Seidenberg, 1998; Elman, 1990). Two different interpretations exist with respect to how co-occurrence probabilities affect speech segmentation (Rytting, 2004). Saffran, Newport, and Aslin (1996) suggest that word boundaries are hypothesized at troughs in transitional probability. That is, a word boundary is inserted when the probability of a bigram is lower than those of its neighboring bigrams. This *trough-based* segmentation strategy thus interprets bigram probabilities using the context in which the bigram occurs. Computational models have shown this interpretation to be effective in segmentation (e.g., Brent, 1999a). A trough-based approach, however, is not capable of extracting unigram words, since such words would require two adjacent local minima (Rytting, 2004; Yang, 2004). The implication is that the learner is unable to discover monosyllabic words (Yang, 2004, for syllable-based statistical learning) or single-phoneme words (Rytting, 2004, for segment-based statistical learning).

Studies addressing the role of probabilistic phonotactics in infant speech segmentation (e.g., Mattys & Jusczyk, 2001b) indicate that the probability

of a bigram can also affect speech segmentation directly, i.e. regardless of neighboring bigrams. The interpretation of probabilities in isolation has been modeled either by inserting boundaries at points of low probability (Cairns et al., 1997; Rytting, 2004), or by clustering at points of high probability (Swingley, 2005). In both cases the learner relies on a threshold on the probabilities in order to determine when a boundary should be inserted, or when a cluster should be formed. This *threshold-based* segmentation strategy gives rise to a new problem for the learner: How can the learner determine what this threshold should be? Although the exact value of such a statistical threshold remains an open issue, it will be argued in Section 2.3 that a classification of bigrams into functionally distinct categories can be derived from the statistical distribution.

While the relevance of segment co-occurrence ('biphone') probabilities for segmentation is well-established, the potential relevance of more abstract, feature-based constraints in segmentation has not been explored in previous segmentation models.<sup>3</sup> The model presented in this chapter complements previous modeling efforts by adding a generalization component to the statistical learning of phonotactic constraints. The existence of such general ('abstract') constraints is widely accepted in the field of phonology. However, the link to psycholinguistically motivated learning mechanisms, such as statistical learning and generalization, has not been made, and the learning of phonotactic constraints from continuous speech has not been explored. In fact, constraints in linguistic frameworks such as Optimality Theory (Prince & Smolensky, 1993) are typically assumed to be innate (as a part of Universal Grammar). An increasing number of phonological studies, however, have assumed that constraints are not innate, but rather that constraints have to be learned (Albright, 2009; Boersma, 1998; Hayes, 1999; Hayes & Wilson, 2008). In this view, learning constraints from experience is an essential part of language acquisition.

### 2.1.2 *Models of constraint induction*

Recent work in phonology has focused on the induction of phonotactic constraints, either from articulatory experience (Hayes, 1999), or from statistical regularities in the lexicon (e.g., Hayes & Wilson, 2008). These models reduce Universal Grammar to a given set of phonological features, and a fixed format of phonotactic constraints. The constraints themselves are induced by training the model on input data. Hayes (1999) proposes that learners can construct

<sup>3</sup> A study by Cairns et al. (1997) used a Simple Recurrent Network with feature-based input representations. However, their feature-based network did not perform better than a model based on simple biphone probabilities. The study thus provides no evidence that feature-based generalizations improve segmentation performance, as compared to segment-based probabilities.

phonological constraints on the basis of experienced phonetic difficulty in perceiving and producing sounds. That is, the child gets feedback from her own production/perception apparatus. The experience is reflected in a map of phonetic difficulty which is used by the learner to assess possible phonotactic constraints. That is, the learner constructs a space of logically possible constraints with features as primitive elements. These candidate constraints are then assessed for their degree of phonetic grounding. The result is a set of phonetically grounded markedness constraints which can be incorporated into the formal constraint interaction mechanism of Optimality Theory (Prince & Smolensky, 1993).

Other models have been based on the idea that phonotactic patterns are abstractions over statistical patterns in the lexicon (Albright, 2009; Frisch et al., 2004; Hayes & Wilson, 2008; Pierrehumbert, 2003). Hayes and Wilson (2008) propose a model in which phonotactic constraints are selected from a constraint space (provided by Universal Grammar), and are assigned weights according to the principle of maximum entropy. Specifically, constraints are selected according to the accuracy of the constraint with respect to the lexicon of the native language, and according to the generality of the constraint. It should be noted that the generalizations in the models by Hayes (1999) and Hayes and Wilson (2008) are constructed before the learner is presented with any input. That is, all logically possible generalizations are *a priori* represented as candidate constraints.

A different approach to the construction of generalizations is taken in models by Albright and Hayes (2002, 2003) and Albright (2009). These models gradually build up more abstract constraints through Minimal Generalization (MG) over processed input data, rather than basing generalizations on a pre-given set of constraints. The learner abstracts over sequences of segments and extracts the shared phonological features to construct a new generalization. The procedure is minimal in the sense that only *shared* feature values are used for the construction of generalizations, thus avoiding overgeneralizations which are not supported by the data.

From a cognitive point of view, a learning mechanism that constructs generalizations as a response to input data (such as in Albright & Hayes, 2002, 2003) seems more plausible as a model of infant learning than a mechanism that operates on a space of logically possible constraints (such as in Hayes & Wilson, 2008). The model that is proposed in Section 2.3 therefore takes an approach to the construction of phonotactic generalizations which is similar to MG. That is, generalizations are constructed on the basis of similarities in the input. The model needs no *a priori* abstractions, since it generalizes over statistically learned biphone constraints. The model is also similar to MG in the sense that similarity is quantified in terms of shared phonological features. As



a result, generalizations affect natural classes, rather than individual segments. In line with previous induction models, the model presented in this chapter works with a set of phonological features that is given to the learner prior to phonotactic learning. It is currently an open issue whether these features are innate or learned from the acoustic speech stream (see e.g., Lin & Mielke, 2008). Regardless of the origin of phonological features, it will be assumed here that features are available for the construction of phonotactic generalizations.

An important difference with earlier models of constraint induction is that the model presented in this chapter is unsupervised. While earlier induction models learn by evaluating the accuracy of generalizations with respect to the lexicon (which is a case of supervised learning), it is assumed here that the learner has not yet acquired a lexicon, or that the learner's lexicon is too small to support the learning of phonotactic constraints. In fact, the proposal is that the learner uses phonotactic generalizations, learned from continuous speech, as a source of knowledge to extract words from the speech stream. Therefore, the learner has no way of determining how good, or useful, the resulting generalizations will be. The unsupervised approach to learning generalizations is another step in the direction of a cognitively plausible model of phonotactic learning, especially for the case of phonotactic learning by infants.

Modeling phonotactic constraints for speech segmentation requires the formalization of both a learning model, accounting for the constraints, and a segmentation model, explaining how the constraints are used to predict the locations of word boundaries in the speech stream. Before describing the learning model, *STAGE* (STATistical learning and GENeralization), a formal model will be presented for the interpretation of constraints in speech segmentation. Note that no earlier formal models of speech segmentation have been proposed which use abstract constraints. The segmentation model thus opens up the possibility of modeling abstract phonotactic constraints for speech segmentation.

## 2.2 THE OT SEGMENTATION MODEL

The proposal is to use a modified version of Optimality Theory (OT, Prince & Smolensky, 1993) for regulating interactions between phonotactic constraints in speech segmentation. Optimality Theory is based on the idea that linguistic well-formedness is a relative notion, as no form can possibly meet all demands made by conflicting constraints. The optimal form is one which best satisfies a constraint set, taking into account the relative strengths of the constraints, which is defined by a strict ranking. In order to select the optimal form, a set of candidate forms is first generated. This candidate set contains all

logically possible outputs for a given input. All candidates are evaluated by the highest-ranked constraint. Candidates that violate the constraint are eliminated; remaining candidates are passed on to the next highest-ranked constraint. This assessment process goes on until only one candidate remains. This is the optimal form. The optimal form thus incurs minimal violations of the highest-ranked constraints, while taking any number of violations of lower-ranked constraints for granted. This principle of constraint interaction is known as *strict domination*.

The version of OT that is adopted here retains the assumptions of constraint violability and strict domination, but is otherwise quite different. Whereas OT learners have the task of learning the appropriate ranking for an *a priori* given set of constraints, the task of our learner is (i) to learn the constraints themselves, as well as (ii) to rank these constraints. Crucially, the OT segmentation model does not employ a universal constraint set ('CON') that is given to the learner. Rather, constraints are induced by employing the mechanisms of statistical learning and generalization (cf., Hayes, 1999; Hayes & Wilson, 2008).

The constraint ranking mechanism in our model is also fundamentally different from mechanisms employed in earlier OT learners (in particular, the Constraint Demotion Algorithm, Tesar & Smolensky, 2000; and the Gradual Learning Algorithm, Boersma, 1998; Boersma & Hayes, 2001). Rather than providing the learner with feedback about optimal forms, i.e. segmented utterances, our model assumes unsupervised constraint ranking, since the input to the learner consists exclusively of unsegmented utterances. Each constraint is accompanied by a numerical ranking value, which is inferred by the learner from the statistical distribution, and which expresses the strength of the constraint (see also, Boersma & Hayes, 2001; Boersma, Escudero, & Hayes, 2003).

Finally, note that, although the OT segmentation model is based on strict constraint domination, one could conceive of other constraint interaction mechanisms (such as constraint weighting in Harmonic Grammar, Legendre, Miyata, & Smolensky, 1990) to segment the speech stream. That is, the numerical values of the induced constraints are not committed to the specific interpretation of strict domination. The issue of how the choice of constraint interaction mechanism would affect performance of the model remains open. For the current study, strict domination at least offers a useful mechanism to regulate the interaction between constraints in speech segmentation.

The OT segmentation model is illustrated in Figure 2.1. The learner processes utterances of continuous speech through a biphone window. That is, for each biphone in the utterance, the learner needs to decide whether a boundary should be inserted or not. Input to the OT segmentation model

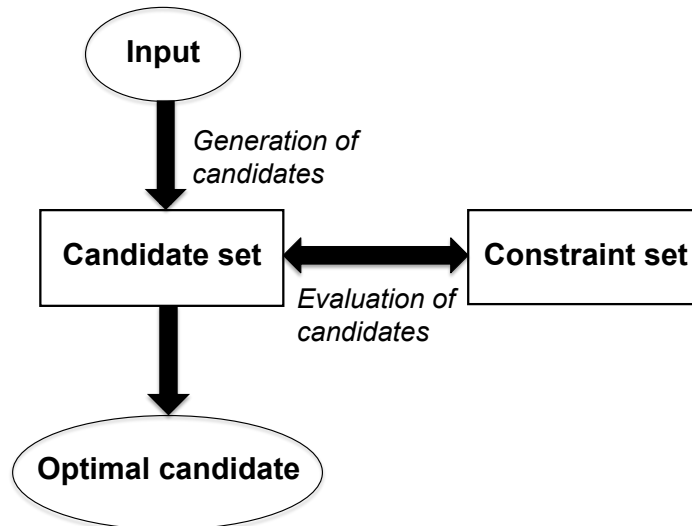


Figure 2.1: The OT segmentation model. The model is presented with biphones in isolation ( $xy$ -sequences), or embedded in context ( $wxyz$ -sequences). Segmentation candidates indicate possible boundary locations. The candidate set is  $\{xy, x.y\}$  for biphones in isolation, and  $\{wxyz, w.xyz, wx.yz, wxy.z\}$  for biphones in context. Candidates are evaluated using the phonotactic constraint set. A boundary is inserted into the speech stream if the constraint set favors segmentation of a biphone (i.e., if  $x.y$  or  $wx.yz$  is the optimal candidate).

consists of biphones, either presented to the model in isolation ( $xy$ -sequences; see Chapter 3 – Experiments 1, 2, and 4), or embedded in context ( $wxyz$ -sequences; see Chapter 3 – Experiment 3). In the latter case, the learner inspects not just the current biphone under consideration ( $xy$ ), but also its immediate neighbors ( $wx$  and  $yz$ ) when making a segmentation decision for the current biphone.

Segmentation candidates are generated which are possible interpretations of the input sequence. This is implemented using a very simple version of OT’s candidate generator (GEN): Each candidate contains either a word boundary at one of the possible boundary locations, or no boundary at all. For biphones in isolation, two candidates are generated:  $xy$  and  $x.y$ . For biphones

in context, the candidates are:  $wxyz$ ,  $w.xyz$ ,  $wx.yz$ , and  $wxy.z$ . Candidates are evaluated using the constraint set, which contains induced, numerically ranked constraints. A boundary is inserted into the speech stream whenever the constraint set favors segmentation of the current biphone under inspection. For the case of biphones in isolation, this means that a boundary is inserted whenever  $x.y$  is the optimal candidate (i.e., it is preferred over  $xy$ ). For the case of biphones embedded in context, a boundary is inserted whenever  $wx.yz$  is the optimal candidate (i.e., it is preferred over  $wxyz$ ,  $w.xyz$ , and  $wxy.z$ ). If multiple candidates remain active after inspection of the constraint set, a winner is chosen at random.

### 2.3 THE LEARNING MODEL: STAGE

STAGE (STATistical learning and GENERALization) learns specific and abstract phonotactic constraints, as well as ranking values for those constraints, from continuous speech. The model keeps track of biphone probabilities in the input (statistical learning). Biphone probabilities trigger the induction of segment-specific constraints whenever the probabilities reach a specified threshold. These thresholds capture the distinction between high- and low-probability phonotactics. The learner interprets low-probability biphones as likely positions for word boundaries. Conversely, the learner interprets high-probability biphones as unlikely positions for word boundaries. The learner thus infers the likelihood of word boundaries from segment co-occurrence probabilities in continuous speech; a process which will be referred to as Frequency-Driven Constraint Induction.

The learner constructs generalizations whenever phonologically similar biphone constraints (of the same phonotactic category, i.e., ‘high’ or ‘low’ probability) appear in the constraint set. Similarities are quantified as the number of shared values for phonological features. In case of a single-feature difference between constraints, the learner abstracts over this feature, and adds the generalization to the constraint set; this will be referred to as Single-Feature Abstraction. The abstract constraints affect sequences of natural classes, rather than sequences of specific segments. In addition to learning constraints, the learner infers ranking values from the statistical distribution. These ranking values determine the strength of the constraint with respect to other constraints in the constraint set.

The general architecture of the model is illustrated in Figure 2.2. Below, the main components of STAGE (i.e., statistical learning, Frequency-Driven Constraint Induction, and Single-Feature Abstraction) are discussed in more detail.

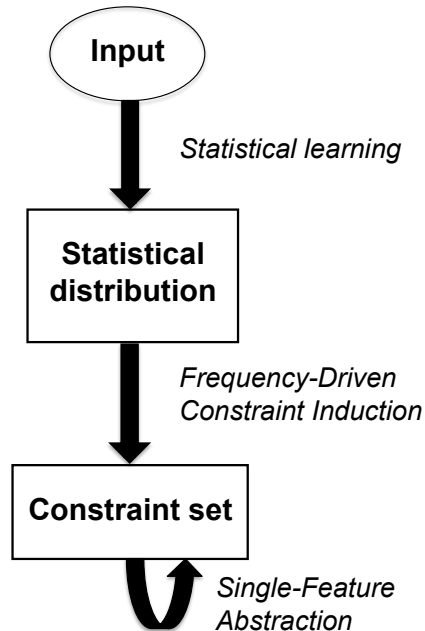


Figure 2.2: The architecture of STAGE. The learner builds up a statistical distribution (Statistical learning) from which biphone constraints are induced which either favor or restrict the occurrence of a biphone (Frequency-Driven Constraint Induction). Generalizations are constructed whenever phonologically similar constraints appear in the constraint set (Single-Feature Abstraction).

### 2.3.1 Statistical learning

Following many psycholinguistic studies STAGE implements a statistical learning mechanism. In this case statistical learning expresses how likely it is that two segments co-occur in continuous speech. The most well-known formula implementing such statistical dependencies is transitional probability (e.g., Saffran, Newport, & Aslin, 1996; Newport & Aslin, 2004):

$$TP(xy) = \frac{Prob(xy)}{\Sigma Prob(xY)} \quad (2.1)$$

where  $Y$  can be any segment following  $x$ . However, as several authors have noted (e.g., Aslin et al., 1998; Perruchet & Peereman, 2004), there exists a variety of formulas that can be used to model statistical learning. STAGE implements a slightly different measure of co-occurrence probability, the observed/expected (O/E) ratio (Pierrehumbert, 1993):

$$\frac{O(xy)}{E(xy)} = \frac{Prob(xy)}{\Sigma Prob(xY) \cdot \Sigma Prob(Xy)} \quad (2.2)$$

where  $Y$  can be any segment following  $x$ , and  $X$  can be any segment preceding  $y$ . A third, closely related measure is mutual information (MI), which corresponds to the  $\log_2$  value of the observed/expected ratio, and which has been used in several computational segmentation studies (Brent, 1999a; Rytting, 2004; Swingley, 2005). The difference between transitional probability and observed/expected ratio (or MI) is the direction of the statistical dependency (Brent, 1999a). Transitional probability expresses a distribution over all elements ( $y \in Y$ ) that *follow* a certain element  $x$ . In contrast, observed/expected ratio is a bidirectional dependency, expressing the likeliness of two elements *co-occurring*. Recent studies have shown that infants and adults can also compute backward transitional probabilities (Pelucchi et al., 2009a; Perruchet & Desaulty, 2008). Learners thus possibly compute both forward and backward probabilities, or perhaps compute a single bidirectional measure. In earlier computational studies, the choice of formula had little impact on the outcome (Brent, 1999a; Swingley, 1999), although Brent (1999a) found somewhat better performance for mutual information than for transitional probability. STAGE uses the O/E ratio, because this measure has been used in earlier studies on phonotactics (Pierrehumbert, 1993; Frisch et al., 2004), and it has a straightforward interpretation in terms of phonotactic constraints (see below).

### 2.3.2 Frequency-Driven Constraint Induction

STAGE classifies probabilities obtained through statistical learning into three distinct categories: ‘low probability’, ‘high probability’, and ‘neutral probability’. Such a categorization follows from the statistical distribution. The O/E ratio, for example, expresses whether a biphone is underrepresented or overrepresented in continuous speech. That is, biphones occur either more or less often than would be expected on the basis of the occurrence frequencies of the individual segments.

Biphones that occur more often than expected are considered ‘high probability’ biphones (overrepresentations) by the learner. In speech segmentation, the learner tries to keep high-probability biphones intact. This is done through the induction of a phonotactic constraint, which states that no boundary

should be inserted (a ‘contiguity’ constraint, see McCarthy & Prince, 1995; Trubetzkoy, 1936):

(2.3) CONTIG-IO( $xy$ ) ‘Sequence  $xy$  should be preserved.’

In contrast, ‘low probability’ biphones (underrepresentations) are likely to contain word boundaries. For this phonotactic category, the learner induces a constraint that favors the insertion of a word boundary (a ‘markedness’ constraint, stating which sequences are not allowed in the language):

(2.4)  $*xy$  ‘Sequence  $xy$  should not occur.’

Markedness and contiguity constraints thus represent two opposing forces in speech segmentation, both of which are derived directly from the statistical distribution. Contiguity constraints exert pressure towards preservation of high-probability biphones, whereas markedness constraints exert pressure towards the segmentation of low-probability biphones. When interpreted within the framework of OT, a contiguity (or markedness) constraint says that the insertion of a word boundary should be avoided (or enforced), *whenever possible* (that is, unless some other, higher-ranked constraint would require otherwise).

The strength of a constraint is expressed by its expected frequency ( $E(xy)$ ). Since the expected frequency of a biphone is defined as the product of individual segment probabilities, the strength of a constraint is in fact determined by the frequencies of its segment constituents. If two phonemes, in spite of their high frequencies in isolation, seldom occur in conjunction, then there is a strong constraint restricting their co-occurrence (that is, a highly ranked markedness constraint). Similarly, if two frequent segments occur much more often than expected, then there is a strong constraint favoring the co-occurrence of these phonemes (that is, a highly ranked contiguity constraint). The ranking value ( $r$ ) of a markedness or contiguity constraint is thus based on the same statistical measure:

$$r = E(xy) = \Sigma Prob(xY) \cdot \Sigma Prob(Xy) \quad (2.5)$$

The third category concerns biphones of ‘neutral’ probability, whose observed frequency is equal to their expected frequency. Such biphones provide the learner with no phonotactic information (which is reflected in the corresponding mutual information value, which is zero in these cases). Therefore, on the basis of the statistical distribution the learner has no reason to induce either type of constraint for such biphones. In a real-life setting, however, the observed frequency will never *exactly* match the expected frequency. The classification of biphones on the basis of their statistical values is illustrated in Table 2.1. STAGE induces contiguity constraints for biphones whose observed

Table 2.1: Classification of biphones according to their statistical values.

Category	O/E ratio	MI	Interpretation	Constraint
low	$O(xy) \ll E(xy)$	$MI(xy) \ll 0$	Pressure towards segmentation	* $xy$
high	$O(xy) \gg E(xy)$	$MI(xy) \gg 0$	Pressure towards contiguity	CONTIG-IO( $xy$ )
neutral	$O(xy) \approx E(xy)$	$MI(xy) \approx 0$	No pressure	–

frequency is *substantially* higher than their expected frequency. In contrast, markedness constraints are induced for biphones with a much lower observed frequency than expected. Finally, no constraints are induced for biphones which carry little probabilistic information.

The decision of when exactly to induce a constraint can be modeled by setting thresholds on the probabilities. Two parameters are introduced in the model: a threshold for the induction of markedness constraints ( $t_M$ ), and a threshold for the induction of contiguity constraints ( $t_C$ ). Introducing these parameters raises an important issue: How do these thresholds become available to the learner? Are they fixed values, possibly due to biological factors? Or can they be induced from the statistical distribution? Since there is currently no way of resolving this issue, the thresholds are set manually. As a first estimation, the notion ‘substantially’ is implemented as a factor-two deviation from the expected value. That is, a markedness constraint is induced whenever the observed frequency is less than half of the expected frequency ( $t_M = 0.5$ ). A contiguity constraint is induced whenever observed is more than twice the expected frequency ( $t_C = 2.0$ ). A wide range of possible threshold values is tested in Chapter 3 (Experiment 2).

### 2.3.3 Single-Feature Abstraction

The categorization of biphones into markedness and contiguity constraints provides the learner with a basis for the construction of phonotactic generalizations. Such generalizations state which classes of sound sequences should in general be segmented, or be kept intact, respectively. The learner constructs a generalization whenever phonologically similar constraints of the same category arise in the constraint set. Generalizations are added to the constraint set, while keeping existing constraints intact.



In modeling the unsupervised learning of phonotactic generalizations, a very basic measure of similarity is implemented, adopting the notion of ‘constraint neighbors’ (Hayes, 1999). Two constraints are said to be neighbors when they have different values for one single feature only. The following set of features is adopted: *syllabic, consonantal, approximant, sonorant, continuant, nasal, voice, place, anterior, lateral* for consonants, and *high, low, back, round, long, tense, nasalized* for vowels (see Appendix B). For example, the constraints CONTIG-IO(pl) and CONTIG-IO(bl) are neighbors, since they have a single-feature difference (*voice* in the first segment). Generalization consists of abstracting over these differences: A new constraint is created where this feature has been neutralized. That is, only shared feature values remain. The resulting constraint (e.g., CONTIG-IO( $x \in \{p,b\}; y \in \{l\}$ )) affects a sequence of natural classes, rather than a sequence of individual segments. This more general constraint is added to the constraint set. The algorithm is recursive: The existence of another phonologically similar biphone constraint that is a neighbor of this abstract constraint would trigger a new generalization. In this case, the feature difference between the abstract constraint (CONTIG-IO( $x \in \{p,b\}; y \in \{l\}$ )) and the biphone constraint (e.g., CONTIG-IO(br)) is assessed as the total number of different feature values for features that have not been neutralized in the generalization (e.g., *voice* has been neutralized in position  $x$ , therefore only the single-feature difference between /l/ and /r/ is taken into account in computing similarity between the two constraints). Abstraction over the single-feature difference creates an even more general constraint (CONTIG-IO( $x \in \{p,b\}; y \in \{l,r\}$ )), which is again added to the constraint set. This constraint states that any sequence of /p/ or /b/, followed by /l/ or /r/ (i.e., /pl/, /pr/, /bl/, /br/) should not be broken up by a word boundary. The model thus creates constraints that have a wider scope than the specific constraints that cause the construction of the generalization. In the current example, /pr/ is included in the generalization, while there is no specific constraint affecting this biphone. No more new generalizations are created if there are no more biphone constraints from the same constraint class (markedness or contiguity) within a single-feature difference from this constraint.

Generalizations are ranked according to the expected frequencies (as defined in formula 2.5) of the biphone constraints that support the generalization, averaged over the total number of biphones that are affected by the generalization. For example, the contiguity constraints CONTIG-IO(pl), CONTIG-IO(bl), and CONTIG-IO(br) support the generalization CONTIG-IO( $x \in \{p,b\}; y \in \{l,r\}$ ) (in addition to the less general CONTIG-IO( $x \in \{p,b\}; y \in \{l\}$ ) and CONTIG-IO( $x \in \{b\}; y \in \{l,r\}$ )). While this abstract constraint is based on three statistically induced constraints, it affects a total of four biphones: /pl/, /bl/,

/br/, /pr/. In this hypothetical example, the fourth biphone, /pr/, does not support the generalization. This is because the learner did not assign this biphone to the contiguity category. More formally, it did not pass the statistical threshold for contiguity constraints ( $t_C$ ), meaning that it was either assigned to the low-probability category (i.e., \*pr), or to the neutral probability category (in which case no specific constraint was induced for /pr/). Therefore, the ranking value of the generalization  $\text{CONTIG-IO}(x \in \{p,b\}; y \in \{l,r\})$  is the summed ranking values (i.e., expected frequencies) of  $\text{CONTIG-IO}(pl)$ ,  $\text{CONTIG-IO}(bl)$ , and  $\text{CONTIG-IO}(br)$ , divided by 4. Note that the generalization would have been given a higher ranking value by the learner if there would have been a contiguity constraint  $\text{CONTIG-IO}(pr)$ . In that case, the ranking value would be calculated as the summed values of  $\text{CONTIG-IO}(pl)$ ,  $\text{CONTIG-IO}(bl)$ ,  $\text{CONTIG-IO}(br)$ , and  $\text{CONTIG-IO}(pr)$ , divided by 4. Generalizations with stronger statistical support in the constraint category are thus ranked higher than generalizations with weaker support.

The ranking of constraints based on statistical support within constraint categories is crucial, since it ensures that statistically induced constraints will generally be ranked higher than abstracted constraints. This has two important consequences. The first concerns resolving conflicts between constraints. A conflict arises whenever a biphone is affected by both a markedness constraint and a contiguity constraint. In such a case, the markedness constraint favors segmentation of the biphone, whereas the contiguity constraint favors keeping the biphone intact. If two constraints are in conflict, the highest ranked constraint determines the outcome (assuming OT's strict domination). STAGE's constraint ranking allows the learner to represent exceptions to phonotactic regularities, since such exceptions (i.e., specific constraints) are likely to be ranked higher than the regularity (i.e., the abstract constraint).

The second, related consequence concerns constraining the generalization mechanism. STAGE's single-feature abstraction is unsupervised. Since all phonological similarities (with recursive single-feature differences) within a constraint category result in new constraints (which are simply added to the constraint set without any further consideration), the model is likely to overgeneralize. However, overly general constraints will have little statistical support in the data (i.e., relatively few biphone-specific constraints will support them). Since the numerical ranking of the constraints is based on exactly this statistical support, overly general constraints will end up at the bottom of the constraint hierarchy. Thus, general constraints like \*CC,  $\text{CONTIG-IO}(CC)$ , \*CV, etc., are likely to be added to the constraint set, but their impact on segmentation will be minimal due to their low ranking values. Note that specific constraints are not *by definition* ranked higher than more general ones. Specifically, biphone constraints that are made up of low-frequency segments

#### 2.4 AN EXAMPLE: THE SEGMENTATION OF PLOSIVE-LIQUID SEQUENCES

are likely to be outranked by a generalization, since such biphones have low expected frequencies. The numerical ranking values, which are inferred by the learner in an unsupervised fashion, thus resolve conflicts between markedness and contiguity constraints, while constraining generalization at the same time.

As a consequence of generalization, the ‘neutral’ probability biphones are now likely to be pulled into the class of either markedness or contiguity constraints. In fact, this may be the main advantage of generalization for the learner: The middle part of the statistical distribution, where the observed frequency approximates the expected frequency, consists of values that are neither high nor low (the ‘neutral’ category in Table 2.1). Biphones in such a ‘grey’ area carry little probabilistic information. Hence, on the basis of statistical learning alone, the learner would have to make a guess whether or not to segment such a biphone, or would have to inspect the values of neighboring biphones in order to estimate the likelihood of a word boundary. Alternatively, when our model encounters a statistically neutral biphone during segmentation, it can use a more general constraint to determine whether the biphone should be segmented or not. The advantage for the learner is thus that biphones for which no reliable statistical information is available can still be reliably segmented (or be kept intact) due to similarity to other biphones.

#### 2.4 AN EXAMPLE: THE SEGMENTATION OF PLOSIVE-LIQUID SEQUENCES

In this section, an illustration is given of how *STAGE* builds up a constraint set, and how this constraint set is used in speech segmentation using an example of plosive-liquid sequences in Dutch. The example is based on a simulation in which *STAGE* is applied to consonant clusters (CC biphones) in transcribed utterances of unsegmented speech in the Spoken Dutch Corpus (Goddijn & Binnenpoorte, 2003). As a test case, the model is presented with the problem of predicting word boundaries in the following hypothetical Dutch utterance:

(2.6) dat lɛx ik zo tryx brurtjə (‘I’ll put that right back, little brother’)

To the learning infant this sounds like:

(2.7) datlɛxɪkzotryxbrurtjə

The model processes utterances through a biphone window, and is faced with the task of deciding whether or not a boundary should be inserted for each biphone in the utterance. At the onset of learning, the learner knows nothing, and he/she would have to insert boundaries at random. The focus here is on



<i>Input:</i> t1, tr, br	CONTIG-IO(br) ( $r = 344.50$ )
? t1 ? t.l	
? tr ? t.r	
 br b.r	*

Figure 2.3: An OT tableau showing segmentation using a single, specific constraint. The upper left cell contains the input. Segmentation candidates for the input are listed in the first column. The upper row shows the induced constraint set, with corresponding ranking values ( $r$ ) in parentheses. Ranking is irrelevant at this point, since there is only one constraint. The star ‘\*’ indicates a violation of a constraint by a segmentation candidate. The index finger ‘’ indicates the optimal candidate.

the plosive-liquid sequences in the utterance, representing undecided segmentations as ‘?’:

$$(2.8) \text{ d} \boxed{t?l} \text{ ɛxɪkzo} \boxed{t?r} \text{ yx} \boxed{b?r} \text{ ʉrtjə}$$

Through statistical learning the learner builds up a distribution from which constraints are derived. The learner induces a phonotactic constraint whenever a biphone passes the threshold for markedness constraints ( $t_M = 0.5$ ), or the threshold for contiguity constraints ( $t_C = 2.0$ ). For example, the learner discovers that /br/ is overrepresented in the input (i.e.,  $\frac{O(br)}{E(br)} > 2.0$ ), and induces CONTIG-IO(br) with ranking value  $r = E(br) = 344.50$ . (For simplicity, static ranking values are assumed here. However, since ranking values are derived from expected frequencies, ranking values may change due to changes in the statistical distribution.) Figure 2.3 shows how this specific constraint affects segmentation of the plosive-liquid sequences in the utterance, using the OT segmentation model. The learner decides that no boundary should be inserted into /br/. With respect to the other sequences, the learner remains ignorant:

$$(2.9) \text{ d} \boxed{t?l} \text{ ɛxɪkzo} \boxed{t?r} \text{ yx} \boxed{br} \text{ ʉrtjə}$$

As a result of multiple specific plosive-liquid overrepresentations in the statistical distribution, the learner induces multiple similar contiguity constraints:

2.4 AN EXAMPLE: THE SEGMENTATION OF PLOSIVE-LIQUID SEQUENCES

<i>Input:</i> tl, tr, br	CONTIG-IO( $x \in \{p,b,t,d\}; y \in \{l,r\}$ ) ( $r = 360.11$ )	CONTIG-IO(br) ( $r = 344.50$ )
$\text{☞}$ tl t.l	*	
$\text{☞}$ tr t.r	*	
$\text{☞}$ br b.r	*	*

Figure 2.4: An OT tableau showing segmentation using an abstract constraint. The upper row shows the induced constraint set, with corresponding ranking values ( $r$ ) in parentheses. Constraints are ranked in a strict domination, shown from left to right. Violation of a higher-ranked constraint eliminates a candidate. Ranking is irrelevant here, since the constraints are not in conflict. The index finger ‘☞’ indicates the optimal candidate.

CONTIG-IO(pl), CONTIG-IO(pr), CONTIG-IO(bl), CONTIG-IO(dr). Through Single Feature Abstraction the learner infers a general constraint, CONTIG-IO( $x \in \{p,b,t,d\}; y \in \{l,r\}$ ) (in addition to less generalized versions of the constraint, which are not shown in this example). On the basis of statistical support (the specific contiguity constraints for /br/, /pl/, /pr/, /bl/, and /dr/), the learner calculates a ranking value for this abstract constraint. Since the constraint affects a total number of 8 biphones, the ranking value is equal to the summed ranking values of the 5 contiguity constraints, divided by 8. In this case, the strength of the constraint is  $r = 360.11$ . The effect of the generalization is illustrated in Figure 2.4. The generalization has substantial strength: It is ranked slightly higher than the specific constraint for /br/. However, since the constraints do not make conflicting predictions, their respective ranking has no effect on segmentation.

Generalization is both helpful and potentially harmful in this case. As a consequence of generalization, the learner is able to make a decision for all plosive-liquid sequences in the utterance. That is, no more undecided segmentations (‘?’) remain:

$$(2.10) \text{d}\boxed{\text{t}}\text{l}\text{e}\text{x}\text{i}\text{k}\text{z}\text{o}\text{t}\boxed{\text{r}}\text{y}\text{x}\text{b}\boxed{\text{r}}\text{u}\text{r}\text{t}\text{j}\text{o}$$

The generalization helps the learner, since it correctly predicts that /tr/, which is statistically neutral in continuous speech and has no specific constraint affecting it, should be kept intact. However, the constraint CONTIG-IO( $x \in \{p,b,$

<i>Input:</i> tl, tr, br	*tl ( <i>r</i> = 2690.98)	CONTIG-IO( $x \in \{p,b,t,d\}; y \in \{l,r\}$ ) ( <i>r</i> = 360.11)	CONTIG-IO(br) ( <i>r</i> = 344.50)
tl	*		
<sup>ES</sup> t.l		*	
<sup>ES</sup> tr		*	
<sup>ES</sup> br			*
b.r		*	

Figure 2.5: An OT tableau showing interaction between specific and abstract constraints. The upper row shows the induced constraint set, with corresponding ranking values (*r*) in parentheses. Constraints are ranked in a strict domination, shown from left to right. Violation of a higher-ranked constraint eliminates a candidate. Ranking is relevant here, since the constraints are in conflict with respect to /tl/. The index finger ‘<sup>ES</sup>’ indicates the optimal candidate.

t,d};  $y \in \{l,r\}$ ) at the same time overgeneralizes with respect to /tl/ and /dl/. While plosive-liquid sequences are in general well-formed, these specific sequences typically do not occur within Dutch words. (Rare exceptions include *atlas* and *atleet*.)

Since the learner is keeping track of both high and low probabilities in the statistical distribution, the learner also induces markedness constraints, stating which sequences should not occur in the language. For example, the learner induces the constraint \*tl, since /tl/ passes the threshold for markedness constraints ( $\frac{O(tl)}{E(tl)} < 0.5$ ). The ranking value of \*tl is high, due to its high expected frequency (i.e., /t/ and /l/ are frequent phonemes). Therefore, \*tl ends up at the top of the constraint hierarchy (see Figure 2.5). Note that the learner has learned an exception to a generalization: The learner will not insert boundaries into plosive-liquid sequences, *unless* it concerns the specific sequence /tl/. In sum, the model has correctly inferred that /br/ should not contain a boundary (due to the induction of a specific constraint, and confirmed by a generalization). In addition, the model has learned that /tr/ should not contain a boundary (due to the abstract constraint). And finally, the model has correctly inferred that /tl/ is an exception to the generalization, and that /tl/ should therefore be broken up by a word boundary. The learner predicts the correct segmentation for all the plosive-sequences it was presented with:

$$(2.11) \text{ da } \boxed{\text{t.l}} \text{ exikzo } \boxed{\text{tr}} \text{ yx } \boxed{\text{br}} \text{ urtjə}$$

## 2.5 GENERAL DISCUSSION

In this chapter, a computational model has been proposed for the induction of phonotactic knowledge from continuous speech. The model, STAGE, implements two learning mechanisms that have been shown to be accessible to infant language learners. The first mechanism, statistical learning, allows the learner to accumulate data and draw inferences about the probability of occurrence of such data. The second mechanism, generalization, allows the learner to abstract away from the observed input and construct knowledge that generalizes to unobserved data, and to probabilistically neutral data. We integrate these mechanisms into a single computational model, thereby providing an explicit, and testable, proposal of how these two mechanisms might interact in infants' learning of phonotactics.

Several properties of STAGE's frequency-driven constraint induction are worth noticing. First, it should be stressed that the learner does not process or count any word boundaries for the induction of phonotactic constraints. It has been argued that phoneme pairs typically occur either only within words or only across word boundaries, and that keeping track of such statistics provides the learner with a highly accurate segmentation cue (Christiansen, Onnis, & Hockema, 2009; Hockema, 2006). However, such a counting strategy requires that the learner is equipped with the ability to detect word boundaries *a priori*. This is a form of supervised learning that is not representative of the learning problem of the infant, who is confronted with unsegmented speech input. In contrast, STAGE draws inferences about the likelihood of boundaries based on probabilistic information about the occurrences of segment pairs in unsegmented speech. The model categorizes biphones without explicitly taking any boundary information into account, and thus represents a case of unsupervised learning.

Second, while earlier segmentation models employed either boundary detection strategies (Cairns et al., 1997; Rytting, 2004) or clustering strategies (Swingley, 2005), STAGE integrates these approaches by acknowledging a role for both edges of the statistical distribution. A low probability biphone is likely to contain a word boundary, which is reflected in the induction of a markedness constraint. Conversely, a high probability biphone is interpreted as a contiguous cluster, reflecting the low probability of a boundary breaking up such a biphone.

Finally, the assumption of functionally distinct phonotactic categories is also what sets the approach apart from earlier constraint induction models (Hayes, 1999; Hayes & Wilson, 2008). Whereas these models induce only markedness constraints, penalizing sequences which are ill-formed in the language, STAGE uses both ends of a statistical distribution, acknowledging

that other sequences are *well*-formed. While ill-formedness exerts pressure towards segmentation, well-formedness provides counter pressure *against* segmentation. This distinction provides a basis for the construction of phonotactic generalizations, while at the same time providing counter pressure against overgeneralization.

By adding generalization to the statistical learning of phonotactic constraints, the model achieves two things. First, through generalization over observed data, the learner has acquired abstract knowledge. While most theories of abstract linguistic knowledge assume abstract representations to be innate, the model shows that it is possible to derive abstract phonotactic constraints from observed data. This finding is in line with recent studies (e.g., Hayes & Wilson, 2008) that aim at minimizing the role of Universal Grammar in language acquisition, while still acknowledging the existence of abstract representations and constraints. Second, the interaction of markedness and contiguity constraints, with varying degrees of generality, and the use of strict constraint domination in speech segmentation, allows the learner to capture generalizations, as well as exceptions to these generalizations. Similar to earlier work by Albright and Hayes (2003), the model thus makes no principled distinction between ‘exceptions’ and ‘regularities’ (cf. Pinker & Prince, 1988). That is, regularities and exceptions are modeled in a single formal framework. The linguistic relevance of the constraints induced by STAGE will be addressed in more detail in Chapter 4.

An important property of STAGE is that the model is unsupervised. That is, unlike previous models of phonotactic learning (e.g., Pierrehumbert, 2003; Hayes & Wilson, 2008), the model does not receive any feedback from segmented utterances or word forms in the lexicon. The learner induces phonotactic constraints from its immediate language environment, which consists of unsegmented speech. The model thereby provides a computational account of phonotactic learning during the very first stages of lexical acquisition. In fact, through the induction of phonotactics from unsegmented speech, the learner is able to bootstrap into word learning. Alternatively, one might argue that infants rely on segmentation cues other than phonotactics to learn their first words, and then derive phonotactics from a proto-lexicon. Such a view raises the new question of how such other segmentation cues would be learned. In general, knowledge of words cannot be a prerequisite for the learning of segmentation cues, since words are the *result* of segmentation (e.g., Brent & Cartwright, 1996; Swingley, 2005). If these cues are to be used to bootstrap into word learning, then such cues need to be learned from unsegmented speech input. It was therefore argued that at least some knowledge of phonotactics comes before the infant starts to build up a vocabulary of words. This knowledge is learned from continuous speech using a combination of statistical



learning and generalization. An interesting open issue is what the effect of the lexicon will be on the child's acquired phonotactic knowledge when her vocabulary reaches a substantial size. Specifically, one might wonder what types of phonotactic constraints would become available to the learner if the learner would induce phonotactics from the lexicon. This issue will be further explored in Chapter 4.

While the model presented does not rule out the possibility that human learners acquire phonotactics from isolated word forms (as predicted by the Lexicon-Based Learning hypothesis, see Chapter 1), the model provides support for the Speech-Based Learning hypothesis by showing that, at least in theory, it is possible to learn phonotactic constraints from continuous speech. Importantly, this possibility has not been demonstrated previously. Specifically, the learning of feature-based generalizations from an unsegmented speech stream has not previously been explored. STAGE shows that, when combined with a statistical learning mechanism, feature-based generalizations can be induced from continuous speech. The next chapter proceeds to provide empirical evidence in favor of the Speech-Based Learning hypothesis. A series of computer simulations will show that the learning of phonotactic constraints from continuous speech is not just a theoretical possibility. Rather, the induced constraints turn out to be very useful for segmentation. That is, feature-based generalizations, learned from continuous speech, improve the segmentation performance of the learner, and thus potentially help the learner to form a mental lexicon.



# 3

## SIMULATIONS OF SEGMENTATION USING PHONOTACTICS<sup>1</sup>

---

In this chapter, a series of computer simulations is presented which tests the performance of STAGE in detecting word boundaries in transcriptions of continuous speech. It is hypothesized that STAGE, which induces phonotactic constraints with various degrees of generality, outperforms purely statistical approaches to speech segmentation, such as transitional probability. While it is an open issue whether infant language learners use phonotactic generalizations in speech segmentation, better performance by the model in these simulations would demonstrate a potential role for such generalizations. Specifically, if feature-based generalizations are useful for segmentation, then such generalizations potentially contribute to the development of the lexicon.

### 3.1 INTRODUCTION

The goal of this chapter is to simulate the segmentation of continuous speech using phonotactics. These computer simulations allow for a comparison between models that make different assumptions about the learning mechanisms that are used for phonotactic learning. The crucial comparison is between models with and without feature-based generalization. While probabilistic cues for segmentation have been used in many models (e.g., Brent, 1999a; Cairns et al., 1997; Daland, 2009; Rytting, 2004), none of these models have explored the potential relevance of more abstract phonotactic constraints for speech segmentation. In fact, most studies assume that such constraints are learned from the lexicon, and thus can only develop *after* the learner has acquired the skills necessary to segment speech (e.g., Hayes & Wilson, 2008). The model presented in the previous chapter (STAGE) shows that it is possible to induce phonotactic constraints from continuous speech using the mechanisms of statistical learning and generalization. Here, it is examined how the phonotactic generalizations, created by the model, affect the segmentation of continuous speech. If the generalizations improve the segmentation performance of the learner, then this is an indication that phonotactic generalizations may also

---

<sup>1</sup> Sections of this chapter were published in: 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.

have a role in segmentation by human learners, assuming that learners will optimize their segmentation skills (e.g., Cutler et al., 1986).

Several psycholinguistic studies have shown that adult listeners use abstract phonological constraints for segmentation (e.g., Suomi et al., 1997; Warner et al., 2005). However, these studies have not addressed the issue of how such constraints are acquired. The simulations in this chapter demonstrate a first unified account of the learning of abstract constraints, and the segmentation of continuous speech using these constraints. While the simulations are not a test for the psychological plausibility of the model, the implication of successful simulations would be that the approach can be considered as a *potential* account of human segmentation behavior. (A direct link between the computational model and human segmentation data will be made in Chapter 4.) It should be noted that the simulations do not aim at obtaining perfect segmentations, but rather address the effect of the learning of abstract, natural class constraints compared to learning without generalization. A more complete model of speech segmentation would involve the integration of multiple cues (see e.g., Christiansen et al., 1998, for a segmentation model that combines phonotactics and stress). For the current purposes, we focus on the contribution of phonotactics to speech segmentation.

The experiments complement previous computational studies in that the segmentation simulations involve fairly accurate representations of spoken language. That is, while segmentation studies typically use orthographic transcriptions of child directed speech that are transformed into canonical transcriptions using a phonemic dictionary, the spoken utterances used in this chapter have been transcribed to a level which includes variations that are typically found in the pronunciation of natural speech, such as acoustic reductions and assimilations. Experiment 1 compares STAGE to statistical learning using segmentation models that process biphones in isolation (i.e., without inspecting the context in which they occur; Section 3.2). The effect of varying thresholds for these models is investigated in Experiment 2 (Section 3.3). In Experiment 3, the learning models are compared using a setting in which segmentation decisions for biphones are affected by context (Section 3.4). Finally, Experiment 4 addresses the development of STAGE as a function of input quantity (Section 3.5). The chapter ends with a discussion of the findings (Section 3.6).

### 3.2 EXPERIMENT 1: BIPHONES IN ISOLATION

The goal of the first experiment is to assess whether infants would benefit from phonotactic generalizations in speech segmentation. This question is addressed in a computer simulation of speech segmentation, allowing for

a comparison between models that vary in their assumptions about infant learning capacities. The crucial comparison is between models that are solely based on biphone probabilities, and STAGE, which relies on both statistical learning and generalization.

### 3.2.1 Method

#### Materials

The models are tested on their ability to detect word boundaries in broad phonetic transcriptions of the Spoken Dutch Corpus (*Corpus Gesproken Nederlands*, CGN). To create representations of continuous speech, all word boundaries were removed from the transcribed utterances. The ‘core’ corpus, about 10% of the total corpus, contains a fairly large sample (78,080 utterances, 660,424 words) of high quality transcriptions of spoken Dutch. These broad phonetic transcriptions are the result of automatic transcription procedures, which were subsequently checked and corrected manually (Goddijn & Binnenpoorte, 2003). Due to this procedure, variations in the pronunciation of natural speech are preserved in the transcriptions to a large extent. For example, the word *natuurlijk* (‘naturally’) occurs in various phonemic realizations. The Spoken Dutch Corpus contains the following realizations for *natuurlijk*. (Frequency of occurrence is given in parentheses. Only realizations that occur at least 10 times in the corpus are displayed.)

- (3.1) nətylək (86), natylək (80), nətyrlək (70), nətyk (68), ntyk (57), natyrlək (56), natyrlək (55), natyk (54), tyk (43), natylək (40), nətyləg (29), nətyg (28), natyk (23), natyləg (20), tyg (19), ntyg (18), nətyrləg (17), natyg (16), natyg (13), nətyrləg (12), ntylək (12), nətyrk (10), natyrləg (10).

This example illustrates some of the variability that is found in pronunciations of natural speech. In fact, the canonical transcription /nətyrlək/ is not the most frequent realization of *natuurlijk* in natural speech. The combination of its size and level of transcription accuracy makes the Spoken Dutch Corpus fairly representative of spoken Dutch. In contrast, previous computational segmentation studies typically used orthographic transcriptions of child-directed speech that were transformed into canonical transcriptions using a phonemic dictionary. Different realizations of words are lost in such a transcription procedure. The current study complements previous modeling efforts by investigating the performance of segmentation models in a setting that includes natural variability in the pronunciation of connected speech.

*Procedure*

The models are tested on novel data (i.e., data that was not used to train the model). To further increase the generalizability of the results, simulations are based on 10-fold cross-validation (see e.g., Mitchell, 1997). Each utterance in the corpus is randomly assigned to one out of ten disjunct sets, such that each set contains approximately 10% of the data points (i.e., biphones) in the corpus. With this random partition, ten simulations are done for each model. In each simulation one of the ten sets (i.e., 10% of the corpus) is used as test set and the remaining nine sets (90% of the corpus) are used as training set. This procedure gives a more reliable estimate of a model's performance than a single randomly chosen test set.

The models are trained on unsegmented utterances in the training set. The models are then given the task of predicting word boundaries in the test set. Output of the models thus consists of a hypothesized segmentation of the test set. The models are evaluated on their ability to detect word boundaries in the test utterances. The following metrics are used to evaluate segmentation performance:

Hit rate (H):

$$H = \frac{\textit{TruePositives}}{\textit{TruePositives} + \textit{FalseNegatives}} \quad (3.2)$$

False alarm rate (F):

$$F = \frac{\textit{FalsePositives}}{\textit{FalsePositives} + \textit{TrueNegatives}} \quad (3.3)$$

*d*-prime (*d'*):

$$d' = z(H) - z(F) \quad (3.4)$$

The hit rate measures the number of word boundaries that the model actually detects. The false alarm rate measures the number of boundaries that are incorrectly placed. The learner should maximize the number of hits, while minimizing the number of false alarms. The *d'* score (see e.g., MacMillan & Creelman, 2005) reflects how well the model distinguishes hits from false alarms. The following 'dumb' segmentation strategies would therefore each result in a *d'* score of zero: a) inserting a boundary into every biphone, b) not inserting any boundaries, and c) randomly inserting boundaries. These metrics have some advantages over evaluation metrics that are commonly used in the field of Information Retrieval (IR), i.e. recall, precision, and F-score, since such metrics do not necessarily assign low scores to random models.

Specifically, random models (in which no learning takes place) will obtain high precision scores whenever there are many potential boundaries to be found. (See Fawcett, 2006, for a related discussion.)

Note that the corpus transcriptions do not specify the exact location of a word boundary in cases of cross-word boundary phonological processes, such as assimilations, deletions, degeminations and glide insertions. For example, *kan nog* ('can still') is often transcribed as /kənɔx/. In such cases, it is unclear whether a segment (in this case, /n/) should go with the word to the left or to the right. These phonemes are interpreted as belonging to the onset of the following word, rather than to the coda of the preceding word.

#### *Segmentation models*

Five models are compared with respect to their abilities to successfully detect word boundaries: a random baseline; a statistical learning model based on transitional probabilities (TP); a statistical learning model based on observed/expected ratios (O/E); STAGE's Frequency-Driven Constraint Induction (FDCI; i.e., excluding Single-Feature Abstraction); and the complete STAGE model (i.e., including Single-Feature Abstraction). No learning takes place in the random baseline, and, hence, boundaries are inserted at random. More specifically, for each biphone in the test utterance a random decision is made whether to keep the biphone intact or to break up the biphone through the insertion of a word boundary. The two statistical models (TP, O/E) serve to illustrate the segmentation performance of a learner that does not induce constraints, nor construct generalizations. These models use segment-based probabilities directly to predict word boundaries in the speech stream. In addition to evaluating the complete model, the performance of STAGE's Frequency-Driven Constraint Induction (FDCI) is evaluated separately. This is done to provide a clearer picture of the added value of generalization in the model. That is, we are interested in the contribution of both the statistically induced constraints, and the abstract, natural class-based constraints to the segmentation performance of the model.

In the current experiment, phonotactic knowledge is applied to the segmentation of biphones in isolation (i.e., not considering the context of the biphone). For the statistical learning models, a threshold-based segmentation strategy is used (Cairns et al., 1997; Rytting, 2004; Swingley, 2005). A segmentation threshold can be derived from the properties of the statistical distribution. Since a biphone occurs either more or less often than expected, the threshold for observed/expected ratios is set at  $O/E = 1.0$ . That is, if observed is less than expected, a boundary is inserted. The TP segmentation threshold is less straightforward. For every segment  $x$ , transitional probability defines a distribution over possible successors  $y$  for *this segment*. Transitional probability

thus defines a collection of multiple statistical distributions, each requiring their own segmentation thresholds. TP thresholds are defined in the same way as O/E ratio: The threshold is the value that would be expected if all segments were equally likely to co-occur. For example, if a segment has three possible successors, then each successor is expected to have a probability of  $1/3$ . If the TP of a biphone is lower than this expected value, a boundary is inserted between the two segments.<sup>2</sup>

The statistical models insert boundaries whenever observed is smaller than expected, regardless of the size of this deviation. STAGE makes different assumptions about thresholds: A markedness constraint is induced when observed is *substantially* smaller than expected; a contiguity constraint is induced when observed is *substantially* larger than expected. In Chapter 2, it was argued that generalization over such statistically induced phonotactic constraints would be valuable to the learner, since the generalizations are likely to affect the segmentation of statistically neutral biphones in a positive way. It thus makes sense to compare STAGE to the threshold-based statistical models. If phonotactic generalizations improve the segmentation performance of the learner, then STAGE should have a better segmentation performance than a statistical model that simply considers all biphone values, and inserts boundaries based on a single segmentation threshold. As a first test for STAGE we set  $t_M = 0.5$  ('observed is less than half of expected') and  $t_C = 2.0$  ('observed is more than twice expected'). The induced constraints are interpreted in the OT segmentation model (Figure 2.1), which takes  $xy$ -sequences (i.e., biphones) as input, and returns either  $xy$  or  $x.y$ , depending on which candidate is optimal.

### 3.2.2 Results and discussion

Table 3.1 shows the hit rate, false alarm rate, and  $d'$  scores for each model. The table contains the estimated means (obtained through 10-fold cross-validation), as well as the 95% confidence intervals for those means.

The random baseline does not distinguish hits from false alarms at all, which results in a  $d'$  score of approximately zero. While the random baseline inserts the largest amount of correct boundaries, it also makes the largest amount of errors. The  $d'$  score thus reflects that this model has not learned anything. The two statistical models (TP, O/E) detect a fair amount of word boundaries (about one third of all boundaries in the test set), while keeping the false alarm rate relatively low. The learning performance is illustrated by the  $d'$  values, which show a large increase compared to the random baseline.

<sup>2</sup> Since all segment combinations tend to occur in continuous speech, the TP segmentation thresholds can be approximated by a single threshold:  $TP = \frac{1}{|X|}$ , where  $|X|$  is the size of the segment inventory.



3.2 EXPERIMENT 1: BIPHONES IN ISOLATION

Table 3.1: Simulation results for Experiment 1 (biphones in isolation).

Model		Hit rate		
		Mean	95% CI	
Learning	Segmentation		Lower	Upper
-	random	0.4994	0.4979	0.5009
TP	thresholds	0.3126	0.3102	0.3149
O/E	thresholds	0.3724	0.3699	0.3750
FDCI	OT	0.4774	0.4763	0.4786
StAGE	OT	0.4454	0.4375	0.4533

Model		False alarm rate		
		Mean	95% CI	
Learning	Segmentation		Lower	Upper
-	random	0.5004	0.4992	0.5016
TP	thresholds	0.1069	0.1060	0.1077
O/E	thresholds	0.1372	0.1354	0.1390
FDCI	OT	0.2701	0.2691	0.2710
StAGE	OT	0.1324	0.1267	0.1382

Model		d'		
		Mean	95% CI	
Learning	Segmentation		Lower	Upper
-	random	-0.0024	-0.0062	0.0015
TP	thresholds	0.7547	0.7489	0.7606
O/E	thresholds	0.7678	0.7592	0.7765
FDCI	OT	0.5560	0.5532	0.5588
StAGE	OT	0.9785	0.9695	0.9874

*Note.* The displayed scores are the means obtained through 10-fold cross-validation, along with the 95% confidence interval (CI). TP = transitional probability, O/E = observed/expected ratio, FDCI = Frequency-Driven Constraint Induction, StAGE = Statistical learning and Generalization.

These scores also show that the formula used to implement statistical learning (either TP or O/E) does not have a great impact on segmentation performance. O/E has a higher hit rate than TP, but also has a higher false alarm rate.

If FDCI is applied to the O/E ratios, the performance of the learner worsens. This is not surprising: FDCI reduces the scope of the phonotactic learner to the edges of the statistical distribution. That is, boundaries are inserted if  $O/E < 0.5$ , and boundaries are *not* inserted if  $O/E > 2.0$ . For the remaining biphones (with  $0.5 \leq O/E \leq 2.0$ ; the ‘grey’ area), the learner has no other option but to insert boundaries at random. Therefore, the scores for FDCI are closer to the random baseline. A look at the performance of the complete model reveals that STAGE outperforms both the random baseline and the two statistical models in distinguishing hits from false alarms. Through generalization over the statistically induced constraints, the learner has widened the scope of its phonotactic knowledge. The result is that statistically neutral biphones are not segmented at random, nor are they segmented on the basis of their unreliable O/E ratios (which are by definition either higher or lower than 1.0). In contrast, those biphones are affected by phonotactic generalizations, which say that they should either be segmented or not due to their phonological similarity to biphones in either the markedness or contiguity category. This strategy results in the best segmentation performance. Compared to its purely statistical counterpart (O/E), STAGE has both a higher hit rate and a lower false alarm rate (although the difference between false alarm rates is marginal). This results in a  $d'$  score that is substantially and significantly higher than those of the statistical models (which is reflected in the large difference in means, and non-overlapping confidence intervals).

These results show that STAGE, which employs both statistical learning and generalization, is better at detecting word boundaries in continuous speech. These findings provide evidence for a potential role for phonotactic generalizations in speech segmentation. That is, if infants were to construct generalizations on the basis of statistically learned biphone constraints, they would benefit from such generalizations in the segmentation of continuous speech.

While the segmentation thresholds for the statistical learners make a mathematically sensible distinction between high and low probability biphones, and there is evidence that biphone probabilities directly affect speech segmentation by infants (Mattys & Jusczyk, 2001b), there is currently no psycholinguistic evidence supporting any exact value for these thresholds. Moreover, the exact values of the constraint induction thresholds employed by STAGE ( $t_M = 0.5$ ;  $t_C = 2.0$ ) are rather arbitrary. It is therefore important to consider a wider range of possible threshold values. Experiment 2 looks at

the effects of varying thresholds for both the statistical learning models and STAGE.

### 3.3 EXPERIMENT 2: THRESHOLDS

Experiment 2 asks to what extent the results of Experiment 1 can be attributed to the specific threshold configurations that were used. Specifically, the current experiment aims at determining whether STAGE’s superior performance was due to a single successful threshold configuration, or whether STAGE in general outperforms statistical learners, regardless of the specific thresholds that are used in the model. To this end a Receiver Operating Characteristic (ROC) analysis is done (e.g., MacMillan & Creelman, 2005; Fawcett, 2006; Cairns et al., 1997). Such an analysis is useful for visualizing the performance of a classifier (such as a threshold-based segmentation model), since it portrays the complete performance in a single curve. An ROC curve is obtained by plotting hit rates as a function of false alarm rates over the complete range of possible threshold values. Such a curve thus neutralizes the effect of using a single, specific threshold in a model and gives a more general picture of the model’s performance as the threshold is varied.

#### 3.3.1 Method

##### *Materials and procedure*

The materials are identical to those of Experiment 1. The procedure is identical to Experiment 1, with the exception that only the first of the ten cross-validation sets is used.

##### *Segmentation models*

The crucial comparison will again be between purely statistical learning models (TP, O/E) and STAGE. Rather than defining a single threshold for the statistical learners, all relevant threshold values are considered. Thresholds are derived from the statistical distribution after the models have processed the training set.<sup>3</sup> A simulation on the test set is conducted for each threshold value using the same segmentation principle as in the previous experiment: If the probability of a biphone is lower than the current threshold, a boundary is inserted.

For STAGE the situation is slightly more complex, since STAGE uses two induction thresholds ( $t_M, t_C$ ). Testing all combinations of values for the two

<sup>3</sup> Since there are 1,541 different biphones in the training set, each with a different probability, there are 1,541 relevant threshold values.

thresholds is not feasible. A range of thresholds for STAGE is tested, based on the assumption that  $t_M$  should be smaller than 1.0, and  $t_C$  should be larger than 1.0. A baseline configuration can thus be formulated:  $t_M = 1.0$ ;  $t_C = 1.0$ . In this case, all biphones with  $O/E < 1.0$  result in the induction of a markedness constraint, and all biphones with  $O/E > 1.0$  result in the induction of a contiguity constraint. As a consequence, the baseline configuration has no ‘neutral probability’ category. Such a category is introduced by pushing the thresholds away from 1.0 towards the low- and high-probability edges of the statistical distribution. Varying the induction thresholds causes changes in the amount of specific constraints that are induced by the learner, and affects the generalizations that are based on those constraints. For the induction of markedness constraints the values  $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$  are used for  $t_M$ . Similarly, the values  $\{1.0, 1.11, 1.25, 1.43, 1.67, 2.0, 2.5, 3.33, 5.0, 10.0\}$  are used for  $t_C$  (using logarithmic steps). This results in a total of  $10 \times 10 = 100$  different configurations. Note that the configuration from Experiment 1 ( $t_M = 0.5$ ;  $t_C = 2.0$ ) is exactly in the middle. Thresholds are considered that are both less and more conservative than this configuration.

### 3.3.2 Results and discussion

The resulting ROC graph is shown in Figure 3.1. Random performance in an ROC graph is illustrated by the diagonal line. For each point on this line, the hit rate is equal to the false alarm rate, and the corresponding  $d'$  value is zero.  $d'$  increases as the hit rate increases and/or the false alarm rate decreases. Perfect performance would be found in the upper left corner of the ROC space (i.e., where the hit rate is 1, and the false alarm rate is 0). In general, the closer a model’s scores are to this point, the better its performance is (due to high hit rate, low false alarm rate, or both).

Since STAGE uses O/E ratios for the induction of constraints, it is particularly interesting to look at the difference between the O/E line and the different configurations of STAGE (represented as circles in the graph). The TP performance is included for completeness. It should be noted, however, that TP slightly outperforms O/E for a substantial part of the ROC graph. The graph shows that most of the 100 configurations of STAGE lie above the performance line of the statistical learning models (most notably O/E). This should be interpreted as follows: If we make a comparison between STAGE and statistical learning, based on model configurations with identical false alarm rates, STAGE has a higher hit rate (and therefore a higher  $d'$ ). Conversely, for configurations with identical hit rates, STAGE has a lower false alarm rate (and therefore a higher  $d'$ ). This confirms the superior performance of STAGE that was found in Experiment 1. STAGE tends to outperform statistical models,

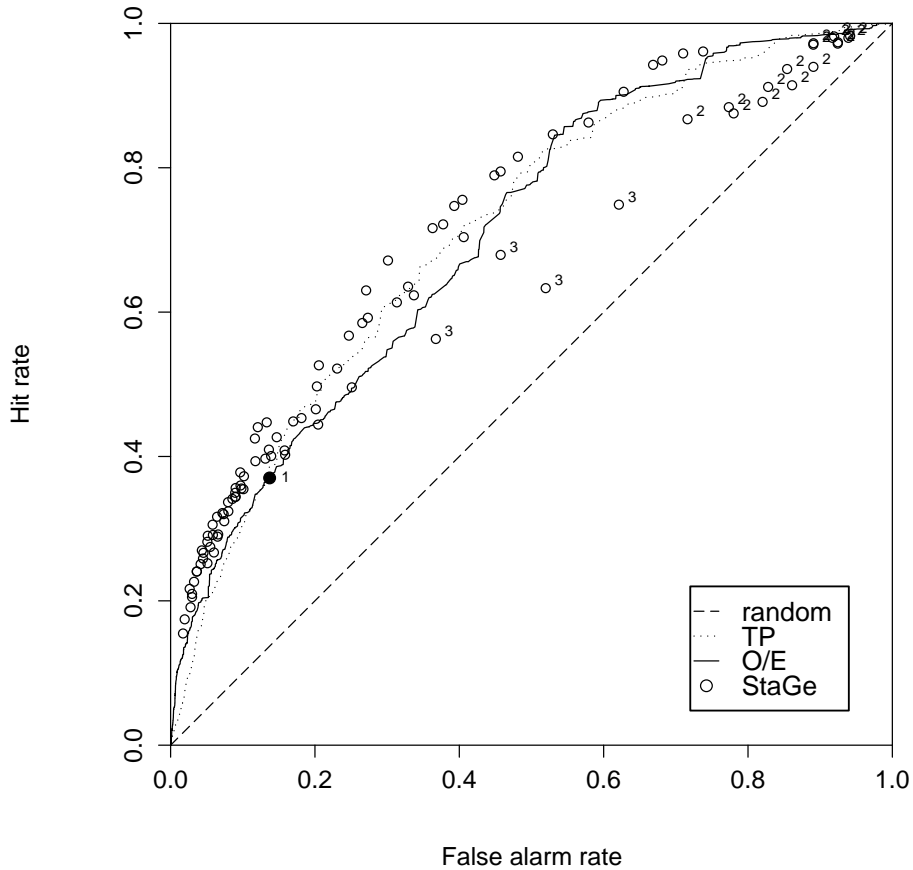


Figure 3.1: An ROC (Receiver Operating Characteristic) graph showing the performance of segmentation models over various thresholds. TP = transitional probability, O/E = observed / expected ratio, StaGe = Statistical learning and Generalization. The STA<sub>GE</sub> baseline configuration ( $t_M, t_C = 1.0$ ) is the solid circle, marked with '1'. Extreme contiguity thresholds ( $t_C = 5.0$ , or  $t_C = 10.0$ ) are marked with '2', whereas configurations with both extreme markedness and extreme contiguity thresholds ( $t_M = 0.1$ , or  $t_M = 0.2$ ;  $t_C = 5.0$ , or  $t_C = 10.0$ ) are indicated with '3'.

regardless of the specific thresholds that are used (with some exceptions, which are discussed below).

While the configuration from Experiment 1 ( $t_M = 0.5; t_C = 2.0$ ) retrieved about 44% of the word boundaries at a false alarm rate of 13% (resulting in a  $d'$  of 0.98), the results in the current experiment show that this performance can be changed without great loss of  $d'$ . For example, the model can be made less conservative, boosting the hit rate to 67%, when using the configuration  $t_M = 0.4, t_C = 2.5$ . In this case, the false alarm rate is 30%. While both the hit rate and false alarm rate are higher, the configuration yields a  $d'$  that is comparable to the original configuration:  $d' = 0.97$ . Similarly, the model can be made more conservative (e.g.,  $t_M = 0.3, t_C = 1.43$ , hit rate: 0.32, false alarm rate: 0.07,  $d'$ : 1.00).

Interestingly, the baseline configuration (the solid circle marked with “1”;  $t_M = 1.0; t_C = 1.0$ ) is positioned exactly on the curve of the O/E statistical learning model. In fact, the baseline configuration has a performance that is nearly identical to the statistical O/E model from Experiment 1 (with segmentation threshold O/E = 1.0). In this case there appears to be neither a positive nor a negative influence of constructing generalizations. This is not surprising: Due to the high number of specific constraints in the baseline model, and since specific constraints are typically ranked high in the constraint set, the statistical values tend to overrule the complementary role of phonological similarity. The result is that the model behaves similarly to the purely statistical (O/E) model. Consider, for example, the sequence /tx/ with  $\frac{O(tx)}{E(tx)} = 1.0451$ . In the baseline configuration, this sequence is kept intact, since the learner induces a highly-ranked constraint CONTIG-IO(tx). The statistical O/E model makes the same prediction: No boundary is inserted, due to an O/E ratio that is slightly higher than 1.0. In contrast, a model with  $t_M = 0.5; t_C = 2.0$  ignores such biphones because of their neutral probability. As a consequence, the sequence /tx/ is affected by a phonotactic generalization,  $*x \in \{t,d\}; y \in \{k,x\}$ , which favors segmentation of the sequence. By pushing the induction thresholds to the edges of the statistical distribution, a smaller role is attributed to probability-based segmentation, and a larger role is attributed to phonological similarity to biphones at the edges of the statistical distribution. This is indeed helpful: Inspection of the segmented version of the corpus reveals 4,418 occurrences (85.6%) of /t.x/, against only 746 occurrences (14.4%) of /tx/.

Figure 3.1 also shows that there are cases in which STAGE performs worse than the statistical learning models. In these cases, the model uses thresholds which are too extreme. Specifically, the model fails for configurations in which the threshold for contiguity constraints is high ( $t_C = 5.0$  or  $t_C = 10.0$ , marked with “2” in the graph). In such cases there are too few contiguity constraints to

### 3.4 EXPERIMENT 3: BIPHONES IN CONTEXT

provide counter pressure against markedness constraints. The model therefore inserts too many boundaries, resulting in a high number of errors. Finally, the worst scores are obtained for configurations that, in addition to an extreme contiguity threshold, employ an extreme markedness threshold ( $t_M = 0.1$  or  $t_M = 0.2$ ;  $t_C = 5.0$  or  $t_C = 10.0$ , marked with “3” in the graph). There are not enough constraints for successful segmentation in this case, and the model’s performance gets closer to random performance.

The current analysis shows that the superior performance of STAGE, compared to purely statistical models, is stable for a wide range of thresholds. This is an important finding, since there is currently no human data supporting any specific value for these thresholds. The findings here indicate that the learner would benefit from generalization for any combination of thresholds, except when both thresholds are set to 1.0 (in which case generalization has no effect), or when extreme thresholds are used (in which case there are too few constraints).

A possible criticism of these experiments is that statistical learning was not implemented in accordance with the original proposal by Saffran, Newport, and Aslin (1996). In Experiments 1-2, statistical thresholds were used for the direct application of probabilistic phonotactic cues to the speech segmentation problem. In contrast, the work on statistical learning by Saffran, Newport, and Aslin (1996) proposes that word boundaries are inserted at *troughs* in transitional probability. In Experiment 3, a series of simulations is conducted which is closer to the original proposal. Models are used which benefit from context in the segmentation of continuous speech.

### 3.4 EXPERIMENT 3: BIPHONES IN CONTEXT

In Experiment 3, the same learning models are tested as in Experiment 1. However, the models use a different segmentation strategy. Rather than considering biphones in isolation, the learner uses the immediate context of the biphone. That is, for each biphone  $xy$ , the learner also inspects its neighboring biphones  $wx$  and  $yz$ .

#### 3.4.1 Method

##### *Materials and procedure*

The materials and procedure are identical to those of Experiment 1. The same training and test sets are used.

*Segmentation models*

For the statistical models (TP and O/E), the trough-based segmentation strategy is implemented as described in Brent (1999a) (and which is a formalization of the original proposal by Saffran, Newport, and Aslin (1996)). Whenever the statistical value (either TP or O/E) of the biphone under consideration ( $xy$ ) is lower than the statistical values of its adjacent neighbors, i.e. one biphone to the left ( $wx$ ) and one to the right ( $yz$ ), a boundary is inserted into the biphone  $xy$ . Note that trough-based segmentation is ‘threshold-free’, since it only considers relative values of biphones.

Again, a simulation is included using Frequency-Driven Constraint Induction (FDCI) (i.e., without applying generalization) to show how much of STAGE’s performance is due to statistical constraint induction, and how much is due to feature-based abstraction. For FDCI and the complete version of STAGE the original threshold configuration with thresholds  $t_M = 0.5$  and  $t_C = 2.0$  is used. In this experiment we ask whether this configuration is better at detecting word boundaries in continuous speech than the trough-based segmentation models. The input to the OT segmentation model is the same as for the trough-based model, namely  $wxyz$  sequences. Segmentation candidates for these sequences are  $wxyz$ ,  $w.xyz$ ,  $wx.yz$ , and  $wxy.z$  (see Figure 2.1). The model inserts a word boundary into a biphone whenever  $wx.yz$  is optimal.<sup>4</sup>

3.4.2 *Results and discussion*

Table 3.2 shows the estimated means, and 95% confidence intervals, of the hit rates, false alarm rates, and  $d'$  scores for each model.

In this experiment the random baseline performs slightly above chance, due to the bias of not inserting boundaries at utterance-initial and utterance-final biphones. In general, using context has a positive impact on segmentation: The performance of all models has increased compared to the performance found in Experiment 1, where biphones were considered in isolation. As in Experiment 1, FDCI by itself is not able to account for the superior performance. By adding Single-Feature Abstraction to the frequency-driven induction of constraints, the model achieves a performance that is better than that of both statistical models (as measured by  $d'$ ). While the difference in  $d'$  is smaller than in Experiment 1, the difference is significant (due to non-overlapping confidence intervals). As in Experiment 1, the formula used to implement

<sup>4</sup> The segmentation models tested here do not consider the initial and final biphones of an utterance as potential boundary positions, since the current segmentation setting requires neighboring biphones on both sides. No boundaries are therefore inserted in these biphones. The size of this bias is reflected in the random baseline, which uses the same  $wxyz$ -window, but makes random decisions with respect to the segmentation of  $xy$ .



### 3.4 EXPERIMENT 3: BIPHONES IN CONTEXT

Table 3.2: Simulation results for Experiment 3 (biphones embedded in context).

<b>Model</b>		<b>Hit rate</b>		
		Mean	95% CI	
Learning	Segmentation		Lower	Upper
-	random	0.4900	0.4885	0.4915
TP	troughs	0.6109	0.6093	0.6125
O/E	troughs	0.5943	0.5930	0.5955
FDCI	OT	0.3700	0.3684	0.3716
StAGE	OT	0.4135	0.4062	0.4207

<b>Model</b>		<b>False alarm rate</b>		
		Mean	95% CI	
Learning	Segmentation		Lower	Upper
-	random	0.4580	0.4575	0.4586
TP	troughs	0.2242	0.2235	0.2249
O/E	troughs	0.2143	0.2138	0.2149
FDCI	OT	0.1478	0.1471	0.1484
StAGE	OT	0.0913	0.0882	0.0945

<b>Model</b>		<b>d'</b>		
		Mean	95% CI	
Learning	Segmentation		Lower	Upper
-	random	0.0803	0.0765	0.0840
TP	troughs	1.0399	1.0346	1.0452
O/E	troughs	1.0301	1.0258	1.0344
FDCI	OT	0.7142	0.7096	0.7188
StAGE	OT	1.1142	1.1081	1.1203

*Note.* The displayed scores are the means obtained through 10-fold cross-validation, along with the 95% confidence interval (CI). TP = transitional probability, O/E = observed/expected ratio, FDCI = Frequency-Driven Constraint Induction, StAGE = Statistical learning and Generalization.

statistical learning (TP vs. O/E) does not seem to have a substantial impact on the segmentation results. Note that in this experiment the statistical models have a higher hit rate than STAGE. However, this coincides with a much higher false alarm rate. In contrast, STAGE is more conservative: It places fewer, but more reliable, word boundaries than the models based on statistical learning. The net result, as measured by  $d'$ , is that STAGE is better at distinguishing sequences with and without word boundaries.

Although the infant should eventually learn to detect all word boundaries, the relatively small amount of hits detected by the model does not necessarily pose a large problem for the learning infant for two reasons. First, phonotactics is merely one out of several segmentation cues. Some of the boundaries that are not detected by the model might therefore still be detected by other segmentation cues. Second, the high  $d'$  score of the model is mainly the result of a low false alarm rate. The model thus makes relatively few errors. Such an undersegmentation strategy may in fact result in accurate proto-words, i.e. chunks of speech which are larger than words, but to which meaning can easily be attributed (e.g. 'thisis.thedoggy'). In contrast, oversegmentation (e.g. 'this.is.the.do.ggy') results in inaccurate lexical entries to which no meaning can be attributed. The tendency towards undersegmentation, rather than oversegmentation, is supported by developmental studies (e.g., Peters, 1983). See Appendix C for a selection of marked-up utterances from the Spoken Dutch Corpus which exemplify this undersegmentation behavior.

To get an impression of how robust the results in the current experiment are, a range of threshold values was tested for STAGE. Because of the computational cost involved in running this type of simulation, a smaller range of thresholds is considered than in Experiment 2, using only the first of our ten cross-validation sets. A total of 9 different configurations were tested (3 thresholds for each constraint category:  $t_M = 0.4, 0.5, 0.6$ ;  $t_C = 1.67, 2.0, 2.5$ ). Out of these 9 configurations, one configuration ( $t_M = 0.6$ ,  $t_C = 1.67$ ) performed worse than the statistical learning models (hit rate: 0.3712; false alarm rate: 0.1013;  $d'$ : 0.9455). The best performance was obtained using  $t_M = 0.4$  and  $t_C = 2.5$  (hit rate: 0.5849; false alarm rate: 0.1506;  $d'$ : 1.2483). It thus appears that, in the current setting, better performance can be obtained by pushing the thresholds further towards the low- and high-probability edges.

The results of Experiment 3 are similar to the results of Experiments 1 and 2, and therefore provide additional support for the hypothesis that learners benefit from generalizations in the segmentation of continuous speech. Regardless of whether the learner employs a segmentation strategy that considers biphones in isolation, or whether the learner exploits the context of neighboring biphones, in both cases STAGE outperforms models that rely solely on biphone probabilities. These findings show that the combined

### 3.5 EXPERIMENT 4: INPUT QUANTITY

strengths of statistical learning and generalization provide the learner with more reliable cues for detecting word boundaries in continuous speech than statistical learning alone (i.e. without generalization).

In Experiments 1–3 STAGE was tested at the end point of learning, i.e. after the model had processed the complete training set. Experiment 4 looks at how the segmentation performance of the model develops as a function of the amount of input that has been processed by the model.

#### 3.5 EXPERIMENT 4: INPUT QUANTITY

The current experiment serves to illustrate developmental properties of the model. That is, given the mechanisms of statistical learning and generalization, how does the model’s segmentation behavior change as more input is given to the learner? It should be stressed that the model’s trajectory should not be taken literally as a time course of infant phonotactic learning. Several unresolved issues (discussed below) complicate such a comparison. Nevertheless, the experiment allows us to better understand STAGE’s learning behavior. In particular, it is a valid question to ask whether the model has reached stable segmentation performance after processing the training set. In addition, the order in which different constraints are learned by the model will be described.

##### 3.5.1 *Modeling the development of phonotactic learning*

The current model works on the assumption that the segment inventory and feature specifications have been established prior to phonotactic learning. It is, however, likely that the processes of segmental acquisition and phonotactic acquisition will, at least partially, overlap during development. Since STAGE makes no prediction regarding the development of the speech sounds themselves (segments, features), and since there exists no corpus documenting such a development, a static, adult-like segment inventory will be assumed here.

STAGE models infant phonotactic learning as a combined effort of statistical learning and generalization. Both mechanisms have been shown to be available to 9-month-old infants (biphone probabilities: Mattys & Jusczyk, 2001b; similarity-based generalization: Saffran & Thiessen, 2003; Cristià & Seidl, 2008). STAGE can therefore be thought of as modeling phonotactic learning in infants around this age. Unfortunately, developmental data regarding the *exact* ages (or input quantities) at which each of these mechanisms become active in phonotactic learning are currently lacking. In the current experiment both mechanisms will be assumed to be used from the start. STAGE could

in principle, however, start with statistical learning and only start making generalizations after the model has accumulated a certain critical amount of input data.

Another issue with respect to modeling development concerns the model's use of memory (Brent, 1999b). The current implementation of the model assumes a perfect memory: All biphones that are encountered in the input are stored, and are used for constraint induction. Hence, phonotactic constraints are derived from accumulated statistical information about biphone occurrences. While the model's processing of the input is incremental, the perfect memory assumption obviously is a simplification of the learning problem.

Finally, the current set of simulations rely on static, manually-set thresholds for the induction of markedness and contiguity constraints. STAGE is used in the same form as in Experiment 1 (segmentation of biphones in isolation;  $t_M = 0.5$ ,  $t_C = 2.0$ ). The difference with the previous experiments is that the performance of the model is tested at various intermediate steps, rather than only testing the model at the end point of learning.

### 3.5.2 *Method*

#### *Materials and procedure*

The materials are identical to those of Experiment 1. Only the first of the ten cross-validation sets is used. The procedure is different: Rather than presenting the complete training set to the model at once, the model is presented with input in a stepwise fashion. The total training set (= 100%) contains about 2 million data points (biphones). Starting with an empty training set, the set is filled with training utterances which are added to the set in random order. The model is trained repeatedly after having processed specified percentages of the training set, using logarithmic steps. For each intermediate training set, the model's performance (hit rate, false alarm rate) on the test set is measured. As a consequence, the model's performance progresses from random segmentation (at 0%) to the performance reported in Experiment 1 (100%).

### 3.5.3 *Results and discussion*

The developmental trajectory of STAGE is shown in Figure 3.2, where the hit rate and false alarm rate are plotted as a function of the input quantity (measured as the number of biphones on a  $\log_{10}$  scale). The model's segmentation performance appears to become stable after processing  $\pm 15,000$  biphone tokens ( $\log_{10} \approx 4.2$ ), although the false alarm rate still decreases slightly after

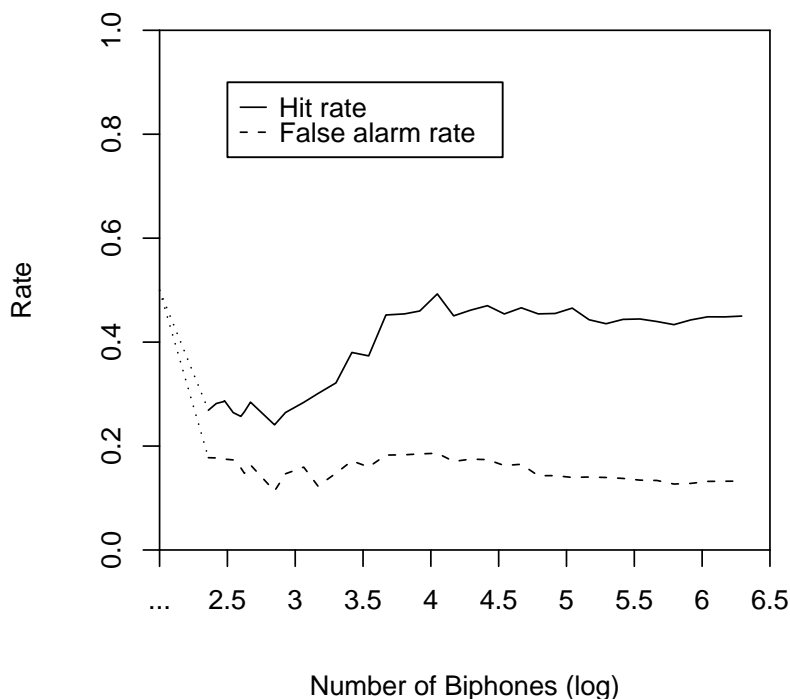


Figure 3.2: The development of STAGE, measuring segmentation performance on the test set as a function of input quantity in the training set. Initially, the learner segments at random (hit rate and false alarm rate are 0.5). The learner starts by inducing contiguity constraints (reducing the number of boundaries that are posited) and induces markedness constraints only after a substantial amount of input has been processed.

this point. Starting from random segmentation in the initial state, the model shows effects of undersegmentation after processing only a minimal amount of input: Both the hit rate and the false alarm rate drop substantially. Interestingly, the false alarm rate stays low throughout the whole trajectory. The hit rate starts increasing after a substantial amount of input ( $\pm 650$  biphones;

$\log_{10} \approx 2.8$ ) has been processed, and becomes relatively stable at  $\pm 15,000$  biphones. Learning where *not* to put boundaries thus precedes the insertion of word boundaries.

To explain the segmentation behavior of the model we consider the constraints that are induced at the various developmental stages. The first constraints to emerge in the model are contiguity constraints. This is caused by overrepresentations in the statistical distribution at this point: Since only a small amount of the total number biphones has been processed, all biphones that do occur in this smaller set are likely to have high O/E ratios. The model tends to induce contiguity generalizations that affect CV and VC biphones (e.g.,  $\text{CONTIG-IO}(x \in \{t,d,s,z\}; y \in \{a,\emptyset\})$ ), and thereby prevents insertion of boundaries into such biphones. The segmentation of CC and VV sequences is left to random segmentation (and a relatively small number of specific contiguity constraints). Among the high-ranked specific constraints are several Dutch function words (e.g.,  $\text{CONTIG-IO}(m)$ , ‘in’;  $\text{CONTIG-IO}(d\emptyset)$ , ‘the’), as well as transitions between such words (e.g.,  $\text{CONTIG-IO}(nd)$ ). As a consequence, function words tend to be glued together (e.g.,  $/md\emptyset/$ ). This type of undersegmentation continues to exist throughout the learning trajectory, and appears to be a general property of the model.

As the distribution becomes more refined, statistical underrepresentations start appearing, and more biphones fall into the markedness category. The growing number of markedness constraints causes the hit rate to increase. Some of the first specific markedness constraints are  $*\emptyset\emptyset$ ,  $*tn$ ,  $*nn$ ,  $*e\emptyset$ , and  $*mt$ . Generalizations start appearing after the model has processed about 1,500 biphones ( $\log_{10} \approx 3.2$ ), such as  $*x \in \{\emptyset\}; y \in \{l,\varepsilon,i,\emptyset\}$ , and  $*x \in \{n\}; y \in \{l,r\}$ . The early-acquired markedness constraints seem to affect other types of biphones (CC, VV) than the early-acquired contiguity constraints (CV, VC). While the model ultimately learns a mixture of markedness and contiguity constraints, affecting all types of biphones, this distinction is a general property of the model.

Taking the modeling simplifications for granted, some of the findings here are supported by developmental studies (e.g., undersegmentation: Peters, 1983). Conversely, new findings that follow from the computational model could provide a basis for future experimental testing in developmental research.

### 3.6 GENERAL DISCUSSION

This chapter investigated the potential role of phonotactic generalizations in speech segmentation. It was hypothesized that learners would benefit from constructing phonotactic generalizations in the segmentation of continuous

speech. This hypothesis was confirmed in a series of computer simulations, which demonstrated that STAGE, which acknowledges a role both for statistical learning and for generalization, and which uses a modified version of OT to regulate interactions between constraints, was better at detecting word boundaries in continuous speech data than models that rely solely on biphone probabilities. Specifically, the generalizations seem to positively affect the segmentation of biphones whose phonotactic probability cannot reliably be classified as being either high or low. The superior performance of STAGE was found regardless of the segmentation window that was used (i.e., biphones in isolation or in context). An analysis of the developmental trajectory of the model indicates that the model learns contiguity constraints before markedness constraints, leading to strong initial undersegmentation.

STAGE relies on the edges of the statistical distribution, rather than on the whole distribution. The model employs statistical thresholds to filter out biphones that cannot be reliably classified as being of either high or low probability. Successful generalization crucially depends on this categorization: Experiment 2 showed that generalization has no effect if all biphones are taken into account in the generalization process. In addition, the experiment shows that generalization fails when the thresholds are set to extreme values. STAGE thus provides an explicit description of how generalization relies on statistical learning: statistical learning provides a basis for generalization. This basis is constructed by the model through the use of thresholds on the values that are obtained through statistical learning.

Although it is not known whether infants actually do learn phonotactic generalizations from continuous speech, and use such generalizations in speech segmentation, the current chapter provides indirect support for such a strategy. The simulations show that infants would benefit from such an approach in the segmentation of continuous speech. Of course, it remains to be determined by experimental testing whether infants actually exploit the benefits of both statistical learning and generalization in a way that is predicted by the model. Nevertheless, the plausibility of STAGE as a model of infant phonotactic learning is based on psycholinguistic evidence for the learning mechanisms that it combines. Infants' sensitivity to the co-occurrence probabilities of segment pairs has been demonstrated (Jusczyk et al., 1994; Mattys & Jusczyk, 2001b; White et al., 2008). In addition, infants' capacity to abstract over linguistic input in order to construct phonotactic generalizations has been demonstrated (Saffran & Thiessen, 2003; Chambers et al., 2003). A recent series of artificial grammar learning experiments by Finley and Badecker (2009) provides further evidence for the role of feature-based generalizations in phonological learning. There is thus evidence for both statistical learning and feature-based generalization in phonotactic learning.

In addition, STAGE is compatible with available evidence about infants' representational units. The generalization algorithm implements phonological features to express the similarity between segments. Although evidence for the psychological reality of such abstract phonological features is limited, infants have been shown to be sensitive to dimensions of acoustic similarity (Jusczyk, Goodman, & Baumann, 1999; Saffran & Thiessen, 2003; White et al., 2008). Furthermore, several studies suggest that abstract phonological features may constrain infant phonotactic learning (Cristià & Seidl, 2008; Seidl & Buckley, 2005).

Given these findings, STAGE seems to make reasonable assumptions about the mechanisms and representations that are involved in phonotactic learning by infants. However, the model is not committed to these representations *per se*. The statistical learning component of the model, Frequency-Driven Constraint Induction, could be applied to other units, such as syllables or allophones. The generalization mechanism, Single-Feature Abstraction, could, in principle, work with different types of features. It remains to be seen how differences in assumptions about the representational units that are processed by the model would affect the speech segmentation performance of the model.

In sum, the mechanisms used by STAGE have received much attention in the psycholinguistic literature, and appear to be available to infant language learners. STAGE provides a computational account of how statistical learning and generalization might interact in the induction of phonotactics from continuous speech. The combined strengths of statistical learning and generalization provide the learner with a more reliable cue for detecting word boundaries in continuous speech than statistical learning alone. The computational study in this chapter thus demonstrates a potential role for phonotactic generalizations in speech segmentation. In the next chapter, two open issues are addressed. First, the issue of to what extent the generalizations learned by STAGE reflect known phonological constraints is addressed. Second, it is investigated whether the model can account for human segmentation behavior, thereby providing the first piece of evidence for the psychological plausibility of the model.



# 4

## MODELING OCP-PLACE AND ITS EFFECT ON SEGMENTATION

---

The previous chapters have shown that the computational mechanisms of statistical learning and generalization can account for the learning of phonotactic constraints from continuous speech. Furthermore, phonotactic generalizations improve the performance of the learner in computer simulations of speech segmentation. The success of the model in these simulations raises two issues for further investigation. While the model improves segmentation (as compared to models that rely solely on segment co-occurrence probabilities), it is not clear whether the model should also be taken as a possible account for the induction of *linguistic* constraints. In order to be a valuable addition to earlier linguistic models of constraint induction (e.g., Albright, 2009; Hayes & Wilson, 2008), the model should be able to account for phonotactic constraints that have been proposed in theoretical linguistics. The first issue is thus the following: To what extent can the model induce constraints which have been proposed in theoretical phonology? This is an important issue for the following reason. In the previous chapters, it was argued that generalization mechanisms such as the one implemented in STAGE may provide a crucial link between statistical approaches to language acquisition, and traditional linguistic approaches which assume more abstract representations. That is, adding a generalization mechanism to the statistical learning of phonotactic constraints leads to emerging (rather than innate) abstract phonological constraints (see also, Albright, 2009; Albright & Hayes, 2003). If the model is to make any claim with respect to the acquisition of linguistic constraints, it is important to see whether STAGE has the potential to model well-studied phonological phenomena.

The second issue concerns the psychological plausibility of the model. While the model is successful in corpus simulations, it is an open issue to what extent the model can account for human segmentation behavior. Two predictions of the model require empirical validation with human segmentation data. First, STAGE induces a mixture of specific (segment-based) and abstract (feature-based) constraints. Do human listeners segment speech in a similar fashion? That is, do they use specific constraints, abstract constraints, or both? Second, STAGE is built on the assumption of bottom-up learning from the speech stream, rather than top-down learning from the lexicon. Do human listeners rely on phonotactic knowledge that is derived from continu-

ous speech? Or do phonotactic constraints for speech segmentation originate from the lexicon?

In this chapter, these two issues are addressed in conjunction by investigating to what extent *STAGE* can provide a learnability account of an abstract phonotactic constraint which has been shown to affect speech segmentation. The chapter connects to the work of Boll-Avetisyan and Kager (2008) who show that *OCP-PLACE* (a phonological constraint which states that sequences of consonants sharing place of articulation should be avoided) affects the segmentation of continuous artificial languages by Dutch listeners. This leads to two specific questions: (i) Can *STAGE* induce constraints which resemble *OCP-PLACE*? (ii) Can *STAGE* account for the effect that *OCP-PLACE* has on the segmentation of continuous speech by Dutch listeners?

The setup of the chapter is as follows. Section 4.1 gives an overview of *OCP-PLACE*, and its effect on segmentation. In Section 4.2, the setup for simulations of *OCP-PLACE* will be discussed, including the approach to predicting human segmentation data. The chapter will proceed in Section 4.3 with examining the output of the model, and the predictions that follow from the model with respect to word boundaries. The output of the model will be used as a predictor for human segmentation data in Sections 4.4 and 4.5. The implications of the findings will be discussed in Section 4.6, emphasizing the contribution of the chapter to supporting the Speech-Based Learning hypothesis.

#### 4.1 INTRODUCTION

Studies of phonological systems support the view that phonotactic constraints are abstract in the sense of referring to natural classes, defined by phonological features. The Obligatory Contour Principle (OCP) is a general principle which restricts the co-occurrence of elements that share phonological properties (e.g., Goldsmith, 1976; Leben, 1973; McCarthy, 1986, 1988). For example, the constraint *OCP-PLACE* (McCarthy, 1988) states that sequences of consonants that share place of articulation ('homorganic' consonants) should be avoided. In many languages, this constraint has the particular gradient effect that pairs of labials (across vowels) tend to be underattested in the lexicon (Arabic: Frisch et al., 2004; English: Berkley, 2000; Muna: Coetzee & Pater, 2008; Japanese: Kawahara, Ono, & Sudo, 2006; Dutch: Kager & Shatzman, 2007).

Different views exist regarding the status of *OCP-PLACE* in phonological theory. McCarthy (1988) proposes that the OCP is one of three universal phonological primitives (along with feature spreading and delinking). The analysis assumes that the OCP holds on all feature-based tiers. Evidence comes from word roots in Arabic, which avoid not only identical consonants

(e.g., *\*smm*), but also non-identical consonants sharing place of articulation (e.g., *\*kbm*). The occurrence of consecutive labials is restricted, since the OCP forbids adjacent elements on the *labial* tier.

A different view is taken by Boersma (1998) who argues that the OCP is not an autosegmental primitive, but rather is the result of functional principles. In the functional interpretation, OCP effects emerge from an interaction of constraints against perceptual confusion (i.e., adjacent identical elements are hard to perceive), and constraints against repetition of articulatory gestures. The consequence of this approach is that the OCP is regarded not as an innate phonological device. Rather, OCP is reduced to more fundamental articulatory and perceptual constraints.

A third view is that OCP-PLACE is a constraint that is the result of abstraction over word forms (or roots) in the lexicon. In an analysis of OCP-PLACE in Arabic, Frisch et al. (2004) define a gradient constraint, whose degree of violation (as measured by the observed/expected ratio) is a function of the similarity between consonant pairs in terms of shared natural classes. Highly similar homorganic consonants are strongly underrepresented in the lexicon of Arabic roots, whereas relatively dissimilar pairs are underrepresented to a lesser degree. Frisch et al. distinguish a functional, diachronic perspective from a formal, synchronic perspective. In their view, similarity avoidance shapes the structure of the lexicon of the language. During language acquisition, individual speakers learn an abstract phonotactic constraint, OCP-PLACE, from the set of root forms in this lexicon. In the view of Frisch et al. (2004), abstract phonotactic constraints are the result of generalization over statistical patterns in the lexicon, and are thus not directly functionally grounded (contrary to the view of Boersma, 1998; Hayes, 1999).

Frisch et al. (2004) propose a similarity metric as a predictor for gradient OCP-PLACE effects in the lexicon. In their account, the degree of underattestation in the lexicon is determined by the number of natural classes that are shared by two consonants ( $\text{similarity} = \# \text{ shared natural classes} / (\# \text{ shared natural classes} + \# \text{ non-shared natural classes})$ ). While the metric of Frisch et al. is successful in predicting gradient OCP effects in the lexicon of Arabic roots, they leave open the question of how individual speakers acquire the abstract OCP-PLACE constraint from the lexicon.

OCP-PLACE does not reflect a universal distribution, regulating consonant occurrences in all languages. Languages differ in their manifestation of OCP effects, in particular with respect to the strength of OCP-PLACE for sequences agreeing on other feature dimensions. Coetzee and Pater (2008) compare the avoidance of homorganic consonants in Muna and Arabic. In both languages the degree of attestedness of homorganic consonants is affected by similarity on other featural dimensions. The languages differ, however, in the relative

strengths of OCP-PLACE effects for different natural classes. While in Arabic similarity avoidance is strongest for homorganic consonants that also agree in sonorancy, Muna shows a more balanced OCP-PLACE effect for agreement in voicing, sonorancy, and stricture. This observation is problematic for the similarity-based account of Frisch et al. (2004). As Coetzee and Pater point out, the similarity metric predicts only a limited amount of cross-linguistic variation in similarity avoidance. Specifically, only differences in the size of a natural class would lead to a cross-linguistic difference. Frisch et al. (2004) point out that OCP-PLACE is weakest for coronals, due to the large number of coronals in the segment inventory of Arabic. In Muna, however, the number of labials, coronals, and dorsals varies only slightly. Nevertheless, Coetzee and Pater report a substantial difference in the degree of underattestation between these classes, dorsals being the most underattested, followed by labials, and then coronals. The similarity metric thus is not able to account for this variation, indicating that the learner needs to learn about the language-specific co-occurrence patterns through exposure to such patterns during learning.

This limitation of the similarity metric for predicting cross-linguistic variation is supported by analyses of similarity avoidance in Dutch. In Dutch, there appears to be no significant correlation at all between the similarity metric and the degree of underattestedness in the lexicon (Kager, Boll-Avetisyan, & Chen, 2009). The consequence is that language learners cannot rely on a general mechanism for similarity avoidance. Instead, language-specific OCP effects will have to be acquired from experience with language data (Boersma, 1998; Coetzee & Pater, 2008; Frisch et al., 2004; Frisch & Zawaydeh, 2001).

Coetzee and Pater propose a universal set of multiple OCP-PLACE constraints (pharyngeal, dorsal, coronal, labial) relativized for agreement on subsidiary features (sonorancy, stricture, voice, emphatic, prenasalization). Their account is based on weighted constraints in Harmonic Grammar (Legendre et al., 1990; Pater, 2009). The model obtains language-specific gradient well-formedness scores for consonant sequences through language-specific weighting of the universal constraints. The learner is presented with sequences of non-identical homorganic consonants, which are distributed according to their frequency of occurrence in the lexicon. Constraint weights are adjusted whenever the learner's own output deviates from the observed learning datum. Weights are adjusted by decreasing the weights of constraints that are violated by the learning datum, and increasing the weights of constraints that are violated by the learner's own output. After training the model is used to create acceptability scores for consonant sequences by assessing the Harmony score of a form relative to its most harmonic competitor. The acceptability scores are then used to predict the O/E ratios of sequences in the lexicon. While

this approach is feature-based, and able to capture cross-linguistic gradient wellformedness, it is based on the assumption of a universal constraint set that is given to the learner.

What is going to be presented in this chapter is an account of OCP-PLACE effects using *induced* constraints. In Chapter 2, a computational model was proposed for the induction of phonotactic constraints. The model works on the assumption that the learner constructs feature-based generalizations on the basis of statistically learned co-occurrence patterns. This view is similar to the theoretical proposal of Frisch et al. (2004) that individual learners acquire OCP-PLACE through abstraction over statistical patterns in the lexicon. It is thus worthwhile to investigate whether STAGE can provide a formal account of the induction of OCP-PLACE. Another reason for pursuing the induction of OCP-PLACE using STAGE is that OCP-PLACE has been found to affect speech processing in general, and speech segmentation in particular. Since STAGE was developed as a model to account for the induction of phonotactic cues for segmentation, the model potentially provides a unified account of the learnability of OCP-PLACE, and its effect on segmentation. Note that STAGE has been proposed as a bottom-up model ('learning from continuous speech'), rather than a top-down ('learning from the lexicon') model. The model can, however, be trained on co-occurrence patterns from either source, and this chapter will therefore explore both speech-based and lexicon-based learning. Before outlining the proposal for the induction of OCP-PLACE using STAGE, evidence for the psychological reality of OCP-PLACE will be briefly discussed, focusing on the role of OCP-PLACE in speech segmentation.

#### 4.1.1 *OCP effects in speech segmentation*

In addition to underattestation of OCP-violating words (such as /smaf/) in the lexicons of various languages, OCP has been shown to affect listeners' behavior in a variety of tasks. For example, Berent and Shimron (1997) found that the avoidance of root-initial geminates (i.e., repeated consonants at the beginning of a root) in Hebrew root morphemes affects the rating of nonwords by native Hebrew speakers. In accordance with the structure of Hebrew roots, root-initial gemination was judged to be unacceptable. Similar results were obtained in a study of wellformedness judgments by native speakers of Arabic. Novel roots containing a violation of OCP-PLACE were judged to be less word-like than novel roots which did not violate OCP-PLACE (Frisch & Zawaydeh, 2001). The effect of OCP-PLACE was found after controlling for possible effects of analogy to existing words, supporting the psychological reality of an abstract phonotactic constraint. OCP-PLACE has also been found to affect

speech perception by English listeners, particularly in phoneme identification tasks (Coetzee, 2005).

There is also evidence that OCP-PLACE has an effect on speech segmentation. The cue provided by OCP-PLACE to segmentation relies on the following observation. If consonants sharing place of articulation are avoided within words in a language, then the occurrence of such consonant pairs in the speech stream would indicate the presence of a word boundary to the listener. Interpreting such consonants as belonging to a single word would constitute a violation of OCP-PLACE. In contrast, breaking up a sequence of homorganic consonants through the insertion of a word boundary would avoid violation of OCP-PLACE, resulting in two distinct words with a higher degree of phonotactic wellformedness.

In a word spotting task, Kager and Shatzman (submitted) found that Dutch listeners are faster at detecting labial-initial words when preceded by a labial-initial syllable (e.g., *bak* 'bin' in *foebak*) than when preceded by a coronal-initial syllable (e.g., *soebak*). Supposedly, the phonotactic wellformedness of *soebak* slows down the detection of the embedded target *bak*. In contrast, the violation of OCP-PLACE in *foebak* results in a phonotactically enforced word boundary *foe.bak* which aligns with the onset of the target word *bak*. Such a boundary would trigger initiating a new attempt at lexical access, which would speed up the detection of the embedded target word (see e.g., Cutler & Norris, 1988). Complementing earlier studies which show that listeners attend to feature discontinuity in segmentation (e.g., a violation of vowel harmony in Finnish; Suomi et al., 1997), Kager and Shatzman provide evidence that listeners also use feature continuity (as reflected in a violation of OCP-PLACE) as a cue for finding word boundaries in the speech stream.

Additional evidence for OCP effects in speech segmentation comes from artificial language learning experiments. Boll-Avetisyan and Kager (2008) tested whether human learners have indeed internalized an abstract OCP-PLACE constraint, and use it as a cue for segmentation. The artificial languages to which Dutch participants were exposed consisted of a continuous stream of CV syllables where the consonant was either a labial (P) or a coronal (T). The stream contained sequences of two labials followed by one coronal (e.g., ... TPPTPPT...). The syllables in the language were concatenated in such a way that the resulting stream contained no statistical cues to word boundaries. The language could thus be segmented in three logically possible ways: TPP, PPT, or PTP. In the test phase, participants were presented with a two-alternative forced-choice task, in which they had to determine which words were part of the language they had just heard. If participants use OCP-PLACE, they should have a significant preference for PTP segmentations (which respect the constraint) over PPT and TPP segmentations (which violate

it). Indeed, participants showed a preference for PTP words over PPT words and TPP words, indicating that participants used OCP-PLACE to segment the speech stream. The study by Boll-Avetisyan and Kager (2008) shows that OCP-PLACE can be used to extract novel words from the speech stream, indicating a potential role for OCP-PLACE in word learning.

In sum, there is evidence that OCP-PLACE is a feature-based phonotactic constraint which affects speech processing in a variety of languages. Due to the language-specific nature of OCP effects, learners need to induce OCP-PLACE from experience with input data. The observation that learning OCP-PLACE requires abstracting over statistical patterns in the input, combined with the evidence that OCP-PLACE affects speech segmentation, makes the induction of OCP-PLACE a well-suited test case for the bottom-up speech-based learning approach. In this chapter, we will look at whether STAGE is able to account for the induction of OCP-PLACE. Importantly, STAGE uses no explicit mechanism for similarity avoidance and assumes that constraints are induced without referring to the lexicon as a basis for phonotactic learning. The theoretical part of this chapter looks at whether the constraints induced by STAGE reflect OCP-PLACE. The empirical part of this chapter will be concerned with investigating whether STAGE is able to account for OCP effects in speech segmentation. Human segmentation data will be used, taken from the study by Boll-Avetisyan and Kager (2008). The ability of STAGE to account for these data will be investigated. The main topics of interest are comparing models that vary in the assumptions they make about the abstractness of OCP-PLACE (consonant probabilities, STAGE, a single abstract OCP-PLACE; Experiment 1), and comparing models that vary in the assumptions they make about the input (continuous speech, word types, word tokens; Experiment 2).

#### 4.1.2 *Modeling OCP-PLACE with STAGE*

The interpretation of OCP-PLACE in this chapter is somewhat different from the view taken in earlier work. While OCP-PLACE has traditionally been thought of as a constraint holding at the level of morphemes (such as root morphemes in Arabic), the view adopted here will be one of a sublexical constraint which is not restricted to apply within morpheme boundaries, but rather operates on consonant sequences in fluent speech. This view traces back to the Speech-Based Learning (SBL) hypothesis (proposed in Chapter 1), which states that infants induce phonotactic constraints from continuous speech, and subsequently use these constraints for the detection of word boundaries in continuous speech. In this view, phonotactic constraints operate on continuous speech input in order to facilitate the development of the mental lexicon. The current chapter asks whether the same approach, which was based on

phonotactic learning by infants, can provide a learnability account of abstract phonotactic constraints that have been proposed in theoretical phonology. The question is thus whether OCP-PLACE is learnable in a bottom-up fashion from a corpus of transcribed continuous speech. If OCP-PLACE can be induced from continuous speech, it may act as a valuable sublexical cue during word learning, since the occurrence of consonants sharing place of articulation in the speech stream would signal the presence of a word boundary to the learner.

The empirical part of the chapter is focused on the effects of OCP on segmentation. In contrast, earlier studies have been concerned with either predicting OCP effects in the lexicon (Coetzee & Pater, 2008; Frisch et al., 2004), or effects of OCP constraints on human wellformedness judgments (e.g., Frisch & Zawaydeh, 2001). A complete account of OCP-PLACE has to be able to explain OCP effects for wellformedness judgments, as well as OCP effects for segmentation. The analysis presented in this chapter therefore provides only a partial account of OCP-PLACE, complementing earlier empirical studies by examining OCP effects for segmentation.<sup>1</sup> One critical aspect that has become clear from earlier studies is that a model of OCP-PLACE needs to be able to account for effects of *gradience*, explaining different degrees of similarity avoidance.

As explained in Chapter 2, the constraints induced by STAGE are evaluated through strict domination. As a consequence, the model will always produce the same output for a given input sequence (in contrast to models that use stochastic evaluation, e.g., Boersma & Hayes, 2001), and in that sense the model does not produce gradient effects. As will become clear later on in this chapter, however, the model does produce gradient *segmentation* effects. This is not caused by the model's constraint evaluation mechanism, but rather by the model's use of context. In short, the model favors segmentation of labial-labial sequences in most contexts, but not all. The result is a gradient effect on segmentation, avoiding the occurrence of labial-labial sequences in most cases, but allowing a violation of OCP-PLACE in cases where this would avoid the violation of a higher ranked constraint. In Section 4.2, it will be shown how these predictions can be used to model gradient preferences of participants for experimental items in segmentation experiments.

Some general limitations for modeling OCP-PLACE effects with constraint induction models were observed by Hayes and Wilson (2008). Hayes and Wilson argue that since the notion of OCP relies heavily on similarity avoidance, modeling the learning of a constraint such as OCP-PLACE would require the implementation of a metric assessing the similarity between consonants in a sequence. In addition, it would require taking the distance between

---

<sup>1</sup> Some suggestions for deriving wellformedness predictions from the model are given in Chapter 6.



consonants into account, since OCP effects are stronger for consonant pairs at close distance than at greater distance.

STAGE faces the same problems as those mentioned by Hayes and Wilson (2008). STAGE uses no explicit mechanism or bias for similarity avoidance to learn phonotactic constraints. That is, the model does not assess the feature difference between  $c_1$  and  $c_2$  in a sequence (e.g.,  $c_1 V c_2$ ). Also, the model has no way of representing gradient distance effects. Taking the latter limitation for granted, the focus will be on modeling local OCP effects (i.e., using consonant pairs with only intervening vowels). I will follow the view of Frisch et al. (2004) that similarity avoidance need not be encoded in the synchronic grammar. Instead, OCP-PLACE may be learnable through abstraction over statistical regularities in the input.

The observation that learning OCP-PLACE requires abstracting over a language-specific distribution makes the induction of OCP-PLACE an interesting test case for STAGE. STAGE was designed to perform exactly such abstractions. STAGE's Frequency-Driven Constraint Induction is used by the learner to detect statistical patterns in the input (i.e., sequences which are highly under- or overrepresented). Single-Feature Abstraction is subsequently used by the learner as a mechanism to build abstract, feature-based constraints on the basis of the statistically-induced constraints on specific segment sequences. It is thus interesting to ask whether the formal learning model that was proposed in the previous chapters, when applied to sequences of non-adjacent consonants, could result in the induction of OCP-PLACE.

The application of STAGE to the induction of OCP-PLACE raises several issues worth investigating. The first issue concerns the level of abstractness of OCP-PLACE. STAGE makes an interesting prediction with respect to this issue. STAGE implements two learning mechanisms, segment-based statistical learning and feature-based generalization (see Chapter 2). Since the generalization mechanism operates on segment-specific constraints, and all feature-based generalizations are added to the constraint set, this approach results in a set of constraints which vary in terms of their generality. STAGE thus predicts both effects of specific consonant probabilities, as well as of general feature-based regularities in the data (see also, Albright, 2009). In Experiment 1 (Section 4.4), it is examined to what extent models that vary in their level of representation can account for OCP effects on speech segmentation. That is, the mixed constraint set produced by STAGE is compared to a statistical model based purely on consonant probabilities, and to a model that implements OCP-PLACE as a single, categorical feature-based constraint. The experiment thus complements the simulations in Chapter 3, where it was shown that adding generalization to statistical learning improves the segmentation performance of the learner. In this chapter, a different angle is taken by examining the effects of having

specific constraints in addition to a constraint that is commonly assumed to be an abstract feature-based constraint. The examination of the roles of both specific and abstract constraints in phonotactic learning and speech segmentation is thereby continued.

The second issue deals with the input that is used by the learner for the induction of OCP-PLACE. STAGE was built on the idea of speech-based learning. STAGE assumes that constraints are learned in a bottom-up fashion from the speech stream, rather than from the lexicon. This raises the question of whether OCP-PLACE can be learned from a language-specific distribution of consonants in continuous speech, or whether the induction of OCP-PLACE requires a lexicon. Since STAGE can in principle be applied to any statistical distribution of biphones, regardless of whether the distribution is based on sequences in continuous speech or sequences in the lexicon, both alternatives will be explored in this chapter. Specifically, in Experiment 2 (Section 4.5) the continuous speech-based learner will be compared to a lexicon-based version of STAGE, employing either type or token probability distributions.

A final issue that deserves attention is the way in which the input is presented to the learner. In Chapters 2 and 3, STAGE was presented with adjacent segments for the induction of phonotactics. However, since OCP-PLACE is a constraint restricting the occurrence of consonant pairs *across vowels*, the constraint appears to be beyond the scope of the biphone-based model. Similar problems were faced by Hayes and Wilson (2008) for the modeling of Shona vowel harmony. Adopting the view that phonological processes often target either exclusively vowels or exclusively consonants, Hayes and Wilson chose to apply their induction model on a vowel projection. That is, the model was presented with sequences of vowels in the data, ignoring intervening consonants. This allowed the model to express the non-local process in local terms. The same line of reasoning can be applied for the induction of OCP-PLACE. STAGE can be presented with sequences of consonants, ignoring intervening vowels, thereby reducing non-adjacent consonants to adjacent segments on the consonantal tier.

A potential problem of this approach, however, is that such a projection does not distinguish between sequences of consonants that are truly adjacent (i.e., consonant clusters), and sequences of consonants that are separated by vocalic material. The consonant projection may be reducing the data too much, since adjacent consonants and non-adjacent consonants are often affected by different phonological processes. For example, while non-adjacent consonants in Dutch show effects of place *dissimilation* (OCP-PLACE, Kager & Shatzman, 2007), nasal-obstruent clusters in Dutch show effects of place *assimilation* (Booij, 1995). Nasal place assimilation in Dutch transforms /pɪmpas/ ('ATM card') into /pɪmpas/, containing two adjacent labials (and thus violating

OCP-PLACE). In a consonant projection, the distinction between adjacent and non-adjacent processes is lost, which potentially hinders their learnability. The approach taken here is to present the learner with non-adjacent consonants exclusively. Specifically, the model is trained on C(V)C sequences (ignoring intervening vowels). By presenting the model with consonant sequences that occur across a vowel, the non-local dependency is expressed in local (bigram) terms. The model can be straightforwardly applied to these representations.

A modification to the way in which our model processes the input naturally leads to the question of whether such a modification is justified. Extensive modifications to computational models can easily result in the criticism that any phenomenon can be modeled with the right set of modifications. As described above, the choice is supported by considerations of phonological processes. In addition, two pieces of evidence from psycholinguistic studies seem to justify the processing assumption. The first piece of evidence comes from research on statistical learning. Several studies have shown that human learners are able to track transitional probabilities of consonants across intervening vowels in a stream of CV sequences in artificial continuous speech (Bonatti et al., 2005; Newport & Aslin, 2004). The statistical learning component of STAGE is thus supported for the proposed modification of a C(V)C processing window. These studies show that learners are, at least, capable of learning dependencies across vowels.<sup>2</sup> The question is whether adding STAGE’s generalization component to the statistical learning of non-adjacent consonant dependencies would result in a constraint that resembles OCP-PLACE.

The second piece of evidence comes from research on feature-based generalization. Finley and Badecker (2009) show that human learners are able to pick up patterns of vowel harmony from a set of words from an artificial language. Since vowel harmony is a process requiring feature continuity in vowels *across consonants*, the study by Finley and Badecker shows that learners can perform feature-based generalization on non-adjacent segments. Interestingly, also vowel harmony has been shown to effect speech segmentation (Suomi et al., 1997; Vroomen et al., 1998). It thus appears that abstract constraints on non-adjacent segments are learnable, and have an effect on speech processing. The basic operations performed by STAGE, segment-based statistical learning and feature-based generalization, are supported by studies on the learning of non-adjacent patterns by human learners. What remains

<sup>2</sup> The findings of Newport and Aslin (2004) and Bonatti et al. (2005) can also be explained by learning on a consonantal tier. However, the languages in these experiments only contained non-adjacent consonants. Showing that learning takes place on a consonantal tier would require showing that learners treat adjacent and non-adjacent consonants equally (since their distinction is lost on a consonantal tier). A more parsimonious interpretation of the existing evidence is that learners keep track of non-adjacent consonants in the speech stream.

unknown is whether feature-based generalizations about non-adjacent consonants can be learned from continuous speech input, and what the effect of such generalizations is on speech segmentation.

#### 4.2 METHODOLOGY

The simulations of the induction of OCP-PLACE will be highly similar to the induction of biphone constraints in the previous chapters. However, the evaluation of the resulting constraint set will be radically different. While in Chapter 3 learning models were evaluated with respect to their ability to accurately predict the locations of word boundaries in a test set consisting of novel unsegmented corpus utterances, the current set of simulations is concerned with the ability of computational learning models to accurately predict human segmentation behavior in an artificial language. That is, the models are trained on native language (L<sub>1</sub>) input, and tested on a novel, artificial language (AL). This section describes methodological issues involved in setting up such a simulation, and discusses in detail how the computational models are used to predict human data.

##### 4.2.1 *Materials*

The simulations involve the induction of Dutch phonotactics affecting non-adjacent consonants. The data consist of broad phonetic transcriptions of the Spoken Dutch Corpus (*Corpus Gesproken Nederlands*, CGN; Goddijn & Binnenpoorte, 2003). Representations of continuous speech were made by removing all word boundaries within utterances. The complete corpus is used as a training set (78,080 utterances, 660,424 words). The test set consists of the artificial language created by Boll-Avetisyan and Kager (2008). In this language, three positional slots are repeatedly filled with different CV syllables to yield a continuous speech stream. Each positional slot is filled with a syllable from a fixed set for that position. The first and second positions contain syllables starting with a labial (Position 1: /pa/, /bi/, /mo/, Position 2: /po/, /be/, /ma/). The third position is filled with syllables starting with a coronal (Position 3: /tu/, /do/, /ne/). Due to the CV structure of the syllables, consonants are always separated from each other by an intervening vowel. For example:

(4.1) ...papotumomanemopotupabetubiponemobenebimadomoponebi...

The language was carefully constructed in order to ensure that participants in the study by Boll-Avetisyan and Kager relied on L<sub>1</sub> phonotactic constraints on non-adjacent consonants during segmentation of the artificial

speech stream. In particular, the language was controlled for experimentally induced statistical biases and for L1 phonotactics not of interest to their study. Importantly, the language was controlled for statistical cues to word boundaries. Since each position is filled with one out of three syllables, transitional probabilities between syllables were 0.33 throughout the speech stream, and participants could thus not rely on syllable probabilities for segmentation (contrary to earlier studies by Saffran, Newport, & Aslin, 1996, and others). In addition, the language was controlled for possible L1 cues for segmentation, such as biphone probabilities, positional syllable frequencies, segmental duration, and intonation.

#### 4.2.2 Procedure

Simulations were conducted using the STAGE Software Package (see Appendix A). The learning model is trained on non-adjacent consonants (i.e., on consonant pairs that occur with intervening vowels). Input to the learner consists of transcribed utterances of continuous speech. STAGE builds up a statistical distribution of O/E ratios for non-adjacent consonants. From this distribution the learner induces segment-specific constraints (Frequency-Driven Constraint Induction) and feature-based generalizations on the basis of the induced specific constraints (Single-Feature Abstraction). Thresholds are set at  $t_M = 0.5$  for the induction of markedness constraints, and  $t_C = 2.0$  for the induction of contiguity constraints. Constraints are ranked using the statistical measure of Expected frequency. (See Chapter 2 for details on the types of constraints and the learning procedure used by STAGE.) The constraint induction procedure thus produces a ranked constraint set, containing both specific and abstract constraints on non-adjacent consonants.

The induced constraint set is used to predict word boundaries in the transcribed continuous artificial language. Vowels are expected not to affect the segmentation of the language (L1 biphone probabilities had been controlled for in the construction of the language), and are removed from the transcription. The remaining sequences of consonants in the artificial language are submitted to the OT segmentation model (see Chapter 2), which uses the constraint set to predict optimal segmentations of chunks of input.

The learner inspects consonant sequences in context (*wxyz*-sequences), using the segmentation procedure that was used in Experiment 3 in Chapter 3. That is, neighboring consonant sequences (*wx*, *yz*) play a role in determining whether the current sequence under inspection (*xy*) should be segmented. All sequences of four consonants in the language are thus evaluated and considered for segmentation. Given the fact that the artificial language contains sequences of two labials (P) followed by a coronal (T), the segmentation model

is confronted with three different types of sequences: PP(.)TP, PT(.)PP, and TP(.)PT, where ‘(.)’ indicates a potential word boundary. The candidate set for a sequence  $wxyz$  consists of the following possible segmentations:  $\{w.xyz, wx.yz, wxy.z, wxyz\}$ . A boundary is inserted into a sequence  $xy$  only if  $wx.yz$  is returned as the optimal segmentation. For example, an input TPPT has the possible segmentations  $\{T.PPT, TP.PT, TPP.T, TPPT\}$ , and will be segmented only if candidate TP.PT is assessed as incurring the least serious constraint violations (as compared to the other candidates) in the induced constraint set.

In sum, the output of the segmentation procedure consists of a hypothesized segmentation of the continuous artificial language. The hypothesized word boundaries are based on violations of induced constraints by consonant sequences in the speech stream. It should be noted that the model can in principle perfectly mimic a segmentation based on a categorical OCP-PLACE. In order to do so the model should always insert boundaries in TPPT sequences, since the insertion of a boundary into PP avoids the violation of OCP-PLACE (i.e., avoids the occurrence of a sequence of two labials).

#### 4.2.3 *Evaluation*

In the evaluation of the hypothesized segmentation of the artificial language, the output of the computer simulation will be matched to human segmentation data. Specifically, the simulation output is used as a predictor for the human judgments obtained by Boll-Avetisyan and Kager (2008). In their study, participants listened to the artificial language (described above) for 10 minutes. After this familiarization phase participants were given a test phase which was intended to assess how participants had segmented the artificial speech stream. The specific interest was in determining whether participants had used OCP-PLACE as a cue to finding word boundaries. This was done by setting up comparisons between logically possible segmentations of the speech stream. That is, participants could have heard PPT, PTP, or TPP words, depending on where they inserted boundaries in the speech stream. Participants were given a two-alternative forced-choice task in which they had to indicate upon each trial which out of two words belonged to the language they had just heard. In the experiment, 14 participants were assigned to each of three different conditions: PTP-PPT, PTP-TTP, PPT-TTP. If participants relied on OCP-PLACE, then they should have a preference for PTP words over PPT and TPP words. In order to explore whether STAGE can accurately predict participants’ preferences for experimental items, predictions of the model will be matched with data from the PTP-PPT condition (which was the condition that yielded the biggest learning effect in Boll-Avetisyan and Kager’s study).

Table 4.1: Human judgments on experimental items.

PTP items		PPT items	
Item	Score	Item	Score
madomo	0.8095	mobedo	0.5476
ponebi	0.7381	pabene	0.5476
ponemo	0.7381	papone	0.5000
podomo	0.6905	mobetu	0.4524
madobi	0.5714	papodo	0.4524
madopa	0.5714	pabedo	0.4048
ponepa	0.5714	pamado	0.4048
podobi	0.5476	pamatu	0.4048
potumo	0.5476	papotu	0.3810
podopa	0.4762	pabetu	0.3571
potubi	0.4524	pamane	0.3333
potupa	0.2381	mobene	0.2619

*Note.* Data from Boll-Avetisyan and Kager (2008).

Wellformedness scores are calculated for each experimental item by averaging judgments across participants. For example, if an item is always selected in the forced-choice trials, the item would get a score of 1. If an item is never chosen, the score is 0. Intermediate values are simply proportions of how often an item has been selected in the forced-choice task. For example, if an item is selected in 70% of the trials, it gets a score of 0.7. A score of 0.5 indicates chance-level performance, in which case there is neither a preference nor dispreference for an item. Values above 0.5 indicate that an item is generally preferred, whereas values below 0.5 indicate that an item is generally dispreferred. Table 4.1 shows the items taken from Boll-Avetisyan and Kager (2008) and their corresponding wellformedness scores.

The crucial part in the evaluation is to match the hypothesized segmentation of the artificial language with the item scores in Table 4.1. Recall that the output produced by the simulations consists of a hypothesized segmentation of the artificial language. Item score predictors are obtained by simply counting how often each item occurs in the model’s segmentation output. That is, whenever an item matches exactly with a stretch of speech between two hypothesized word boundaries (with no intervening word boundaries), the frequency count of the item is updated by 1. We thus use item frequencies in the model’s segmentation output as a predictor for the human judgments on

those items. For example, the segmentation output of a model could be the following:

(4.2) pa.potu.momanemo.potupa.betu.bipo.nemobe.nebi.madomo.ponebi

In this case, the item frequencies of /potupa/, /madomo/, and /ponebi/ are increased. Note that the strict criterion of an exact word match implicitly penalizes models that tend to over- or undersegment, since hypothesized words that are larger or smaller than the actual item do not contribute to the frequency counts of the item.

A measure of how well a model explains human segmentation behavior is obtained by using the item frequencies as a predictor of the wellformedness scores in statistical analyses based on linear regression. Before we look at how well different models are able to account for the data, we simulate the induction of OCP-PLACE using STAGE. The next section describes the contents of the induced constraint set and the predictions it makes with respect to word boundaries.

#### 4.3 THE INDUCTION OF OCP-PLACE

Before we look into how the model compares to other models and other input data in explaining human segmentation behavior, we look at the output of the learning procedure when training STAGE as described above. We first look at the constraints that are induced by STAGE, and then proceed to the predictions the model makes with respect to the locations of word boundaries in the artificial language.

##### *The output of the induction procedure*

STAGE's Frequency-Driven Constraint Induction creates specific constraints from the statistical distribution. Specifically, underattested sequences ( $O/E < 0.5$ ) trigger the induction of markedness constraints, whereas overattested sequences ( $O/E > 2.0$ ) cause the induction of contiguity constraints. If STAGE is able to induce a constraint set that encodes an abstract OCP-PLACE constraint, then the specific constraints should provide the basis for such a generalization. That is, specific underrepresentations of labial-labial sequences in the data should be picked up by the learner and be generalized (through Single-Feature Abstraction) into a more abstract constraint affecting labials as a natural class. Table 4.2 shows all specific constraints induced by the model that affect labial-labial sequences across a vowel (PVP).

The first thing to note is that all specific PVP constraints are markedness constraints. That is, there are no specific contiguity constraints that aim



Table 4.2: Specific \*PVP constraints (sequences with O/E &lt; 0.5) with corresponding ranking values.

CONSTRAINT	RANKING
*[m]V[f]	1495.3318
*[b]V[m]	1480.8816
*[m]V[p]	1225.0285
*[m]V[b]	1159.8271
*[m]V[v]	996.4387
*[f]V[p]	812.7798
*[f]V[v]	661.1154
*[v]V[f]	579.2754
*[v]V[p]	474.5628
*[p]V[v]	288.5850

to preserve PVP sequences. The second thing to note is that not all PVP sequences are underattested to such a degree that they trigger the induction of a specific markedness constraint. Since the natural class of labials includes 5 segments (p, b, f, v, m)<sup>3</sup> there are  $5 \times 5 = 25$  different PVP sequences. Out of these 25 potential constraints, 10 (= 40%) specific constraints arise as a result of Frequency-Driven Constraint Induction. The remaining 15 PVP sequences have O/E values that lie between the values of the two induction thresholds ( $0.5 \leq O/E \leq 2.0$ ). Interestingly, the set of specific constraints only contains markedness constraints against *non-identical* homorganic consonants. This is in line with earlier studies that suggest that identical homorganic consonants (e.g., mVm) may display different OCP effects than non-identical homorganic consonants (e.g., Frisch & Zawaydeh, 2001).

The next question of interest is whether the 10 specific markedness constraints provide a sufficient basis for the construction of an abstract OCP-PLACE constraint. Generalizations that are based on the specific PVP constraints (using Single-Feature Abstraction) are shown in Table 4.3. The constraint set illustrates several properties of STAGE's generalization mechanism. As discussed in Chapter 2, the model induces a multitude of constraints with varying degrees of generality. The ranking of the constraints, which is based on averaged expected frequencies, results in a hierarchy in which specific constraints (and generalizations with strong statistical support by specific constraints) are ranked higher than very broad generalizations, which are

<sup>3</sup> /w/ is labio-dorsal, and is therefore not included in the set of labials proper.

Table 4.3: Specific and abstract constraints affecting labial-labial sequences exclusively.

CONSTRAINT	RANKING
*[m]V[f]	1495.3318
*[b]V[m]	1480.8816
*[m]V[p,f]	1360.1802
*[m]V[f,v]	1245.8852
*[m]V[p]	1225.0285
*[m]V[p,b,f,v]	1219.1565
*[m]V[p,b]	1192.4278
*[m]V[b]	1159.8271
*[m]V[b,v]	1078.1329
*[m]V[v]	996.4387
*[f]V[p]	812.7798
*[f]V[v]	661.1154
*[f,v]V[p]	643.6713
*[v]V[f]	579.2754
*[v]V[p,f]	526.9191
*[p,f]V[v]	474.8502
*[v]V[p]	474.5628
*[f,v]V[p,f]	466.6545
*[f,v]V[p,b,f,v]	315.9667
*[p]V[v]	288.5850
*[p,b,f,v]V[p,b,f,v]	176.0199

supported by only a limited number of specific constraints. As a result of the absence of specific constraints targeting identical consonants, the constraint set does not encode a strong tendency to avoid identical labials. Such sequences are only affected by relatively low-ranked generalizations.

It should be noted that due to the feature set that is used, /m/ does not group with the other labials. This is due to the inclusion of the feature +/- sonorant in the feature set (see Appendix B). The feature is, however, not required to provide each segment with a unique feature bundle. As a consequence, obstruents and sonorants differ by at least two features (sonorant + some other feature). Since STAGE abstracts over single-feature differences, generalizations cannot group obstruents and sonorants together. This can be seen in the most general constraint,  $*[p,b,f,v]V[p,b,f,v]$ , which matches closely with OCP-PLACE (which can be written as  $*[p,b,f,v,m]V[p,b,f,v,m]$ ),

except for the absence of /m/ in the constraint. Note, however, that the relatively large number of specific constraints affecting /m/ causes a tendency to avoid sequences of /m/ followed by a labial. That is, while not all labials are grouped together in a single OCP-PLACE constraint, the net effect of the constraint set is that PVP sequences are avoided. There are, however, four PVP sequences that are not affected by the constraints in Table 4.3: pVm, fVm, vVm, and mVm. Inspection of the complete constraint set reveals that the learner remains neutral with respect to these sequences.<sup>4</sup> In sum, while not all PVP sequences are underrepresented in the data, the constraint set displays general tendencies to avoid PVP sequences due to Single-Feature Abstraction over specific \*PVP constraints.

An interesting property of Single-Feature Abstraction is that OCP effects sometimes arise from constraints that do not specifically target homorganic consonants. OCP effects can be obtained indirectly, as a result of generalization over natural classes that include (but are not limited to) labials. The following hypothetical example illustrates how such effects can occur. When Frequency-Driven Constraint Induction produces three specific constraints \*[p]V[d], \*[t]V[d], \*[t]V[b], a generalization will be constructed stating that voiceless anterior plosives may not be followed by voiced anterior plosives: \*[p,t]V[b,d]. This generalization disfavors the labial-labial sequence pVb. In this case, pVb is avoided even though there is no constraint specifically targeting the avoidance of labial-labial sequences. The OCP effect in this case is due to abstract constraints affecting other natural classes which include labials. In fact, there is quite a large number of constraints which affect PVP sequences, but not exclusively. Examples of such constraints can be found in Table 4.4. These constraints arise due to the fact that the learner picks up various segment-based constraints (both OCP and non-OCP) which all contribute to the construction of feature-based generalizations.

This observation raises the question of whether OCP effects that have been demonstrated in previous studies (both lexicon-based analyses and experimental judgments) should even be attributed to OCP-PLACE, or whether they should be attributed to alternative constraints affecting sequences of natural classes including labial consonants, but not classes of labials exclusively. Below the question of to what extent OCP-like behavior should be attributed to OCP-PLACE is addressed for the case of speech segmentation.

In sum, even though there is no explicit similarity avoidance mechanism, and not all PVP sequences are underrepresented in the data, STAGE arrives

<sup>4</sup> Inspection of O/E values for these sequences shows that a slightly less conservative induction threshold would have resulted in inclusion of specific constraints for the sequences in the constraint set; pVm, mVm: O/E < 0.6; fVm, vVm: O/E < 0.7. This shows that the sequences are underrepresented in the data, albeit not as strongly as the other sequences. See Chapter 3 (Experiment 2) for a discussion of the consequences of changing threshold values.

Table 4.4: Abstract constraints affecting a variety of natural classes that include labials.

CONSTRAINT	RANKING
*[f]V[p,t]	2256.3592
*[f,v]V[p,t]	1786.8968
*[v]V[p,t]	1317.4345
*[f,v]V[p,t,f,s]	1126.9261
*[v]V[p,t,f,s]	1125.6727
*[f,v,s,z]V[p,t]	965.3505
*[v]V[f,s]	933.9110
*[f,v]V[p,t,k,f,s,f,x]	833.5898
*[v,z]V[p,t]	802.5214
*[v]V[p,t,k,f,s,f,x]	787.2657
*[f]V[v,z]	752.6590
*[v]V[f,s,f,x]	718.9978
*[f,v]V[p,b,t,d,f,v,s,z]	657.5455
*[v,z]V[p,t,f,s]	634.7384
*[f,v,s,z]V[p,t,f,s]	599.4141
*[f,v]V[f,s,f,x]	565.3337
*[f]V[C]	560.1140
*[v,z]V[p]	524.8896
*[f]V[v,z,ʒ,ʁ,h]	507.3987
*[f,v,s,z]V[p]	465.6398
*[f]V[f,v,s,z,ʒ,ʁ,x,ʁ,h]	464.8525
*[f,v]V[C]	452.2714
etc.	...

Note. V = vowel, C = obstruents = [p,b,t,d,k,g,f,v,s,z,ʒ,ʁ,x,ʁ,h,ɕ].

at a general, feature-based tendency to avoid PVP sequences. It achieves this through generalization over underrepresented sequences. These generalizations have different levels of generality and need not affect labials exclusively. Similarity avoidance can thus be encoded in the constraint set through generalization over statistical underrepresentations.

#### 4.3.1 Predictions of the constraint set with respect to word boundaries

The ranking of the constraints determines which constraints are actually used by the learner during segmentation. Constraints relevant for segmentation

Table 4.5: Constraints used in segmentation of the artificial language.

CONSTRAINT	RANKING	FREQUENCY	
		Count	Percentage
*[b]V[m]	1480.8816	27	(11.1%)
*[m]V[p,f]	1360.1801	27	(11.1%)
*[m]V[p,b,f,v]	1219.1565	27	(11.1%)
*[C]V[p,t]	376.2584	98	(40.3%)
*[p,b,f,v]V[p,b,t,d,f,v,s,z]	337.7910	54	(22.2%)
*[p,f]V[C]	295.7494	12	(4.9%)
*[C]V[t,s,j]	288.4389	8	(3.3%)
*[p,b,f,v]V[t,d,s,z,ʃ,ʒ,ʝ]	287.5739	6	(2.5%)
*[C]V[p,b,t,d]	229.1519	4	(1.6%)

*Note.* The frequency of use (FREQUENCY) indicates how many *wxyz*-sequences in the language (out of a total of 243) are affected by the constraint during evaluation. V = vowel, C = obstruents = [p,b,t,d,k,g,f,v,s,z,ʃ,ʒ,ʝ].

of the artificial language were extracted by considering OT tableaux for all PPTP, PTPP, and TPPT sequences that were presented to the model. The term ‘relevant’ is used here to indicate that the constraint caused a reduction in the size of the candidate set. Table 4.5 shows that, while STAGE creates a large number of generalizations, only 9 constraints were relevant for segmentation of the artificial language. Due to OT’s strict domination, the vast majority of lower-ranked constraints do not play a role in speech segmentation. While this is a general property of the model, the effect is rather extreme here due to the simple structure of the artificial language, which only contains 6 different consonants. It is therefore not surprising that only a limited number of constraints are relevant in this case. The table also indicates the frequency of use for each constraint, which is a count of how many sequences were affected by the constraint during segmentation. This gives insight into the relative importance of the different constraints.

The relevant constraints appear to be of two types. At the top of the hierarchy in Table 4.5 there are situated three high-ranked constraints affecting labials exclusively. These constraints represent more specific versions of OCP-PLACE, affecting labials, but not all labials. The high ranking values of these constraints ensure that the model often behaves like OCP-PLACE, and favors segmentation of PVP sequences (see Figure 4.1 for an example). In addition, there are six low-ranked constraints affecting natural classes that

MODELING OCP-PLACE AND ITS EFFECT ON SEGMENTATION

<i>Input:</i>	mapodomo	*[m]V[p,f] ( <i>r</i> = 1360.18)
☞	ma.podomo	
	mapo.domo	*
	mapodo.mo	*
	mapodomo	*

Figure 4.1: An OT tableau showing how the model creates OCP-like segmentation behavior. The optimal location of a word boundary is between labials /m/ and /p/ due to a highly ranked constraint disfavoring mVp and mVf sequences.

<i>Input:</i>	bipotubi	*[C]V[p,t] ( <i>r</i> = 376.26)	*[p,f]V[C] ( <i>r</i> = 295.75)
	bi.potubi	*	*
☞	bipo.tubi	*	
	bipotu.bi	**	*
	bipotubi	**	*

Figure 4.2: An OT tableau showing how the model creates segmentation behavior that is different from OCP-based segmentation. The optimal location of a word boundary is between labial /p/ and coronal /t/ due to a highly ranked constraint favoring /p/ (and /f/) in final position (i.e., before a word boundary).

include, but are not limited to, labials. For example, \*[C]V[p,t] militates against obstruents followed by voiceless anterior plosives. While such constraints can produce OCP-like behavior (since they affect sequences of labials), they can also produce segmentations that deviate from the predictions of OCP-PLACE (see Figure 4.2). Since these constraints affect a large number of *wxyz*-sequences in the artificial language, the constraints provide a source for segmentation behavior that is different from OCP-based segmentation.

The constraints in Figure 4.2 are particularly interesting for two reasons. The first issue concerns *alignment*. The two constraints in Figure 4.2 affect a maximally large natural class (given the limitations imposed by the feature set in use) on one side of the vowel and only a small natural class on the other side. Such ‘single-sided generalizations’ produce the effect of an alignment constraint. That is, the constraints are markedness constraints that disfavor

the occurrence of some specific consonants when either followed or preceded by whatever other consonant (i.e., regardless of the quality of that neighboring consonant). These constraints therefore always favor the insertion of a word boundary right after or before such a consonant. In other words, the constraint says that such a particular consonant should be word-final or word-initial, respectively. For example, the constraint  $*[p,f]V[C]$  states that a word boundary should be inserted after /p/ or /f/, regardless of which obstruent follows. That is, whenever the learner encounters /p/ or /f/ in the speech stream, he/she should treat the consonant as word-final. Conversely, the constraint  $*[C]V[p,t]$  states that /p/ and /t/ should be word-initial. It thus appears that purely sequential markedness constraints are able to mimic the effect of alignment constraints. The model is thereby able to capture positional effects without explicitly encoding word edges in the structure of the constraints. The alignment effect simply emerges as a result of feature-based generalizations over sequential constraints.

The second issue concerns *gradience*. The constraints in the tableau in Figure 4.2 have a preference for /p/-initial words, unless /p/ is followed by /t/ in the speech stream (in which case it is preferred to have /t/ in word-initial position, and /p/ in word-final position). The tableau illustrates that, while the constraint set encodes an overall preference for PTP words, there are exceptions to this tendency, in which case the model decides on a different segmentation. This decision is based on the context in which a sequence occurs. That is, the model decides not to insert a boundary into bVp in *bipotubi* because bVp occurs here in the context of a following /t/. According to the constraint set, the occurrence of pVt is worse than the occurrence bVp, and the model therefore has difficulty segmenting *potubi*. The model will only segment *potubi* from the speech stream if it is preceded by a consonant which would be ruled out by a higher ranked constraint than the constraints shown in the tableau. Context thus has the effect that different PTP words are extracted from the speech stream with different frequencies. As a consequence, the model produces gradient effects in segmentation. In the next section, it will be assessed whether the model also produces the *right* gradient effects.

In order to get insight into the model's gradient segmentation behavior, a quantitative analysis was conducted which classifies the boundaries inserted by the model in terms of natural class sequences. Table 4.6 shows how often the model inserted boundaries into PPTP, PTPP, and TPPT sequences. Overall, about 30% of the sequences were broken up by a boundary. This shows that the model on average inserts a boundary after about three syllables. This roughly corresponds to the structure of the artificial language, which consists of a three-syllable pattern. The tableau in Figure 4.2, however, suggests that the model does not exclusively segment TP.PT sequences. Inspection of the

Table 4.6: Absolute count and percentage of word boundaries inserted into the different natural class sequences in the artificial language by *STAGE*.

Sequence	Number of occurrences	Number of boundaries	Percentage
PP(.)TP	899	183.50	20.41%
PT(.)PP	899	118.25	13.15%
TP(.)PT	899	510.00	56.73%
Total:	2697	811.75	30.10%

*Note.* ‘(.)’ indicates a boundary location. Calculations are based on tableau predictions. Fractions can occur when a tableau is uncertain (i.e., when the tableau produces multiple optimal candidates), in which case the number of remaining candidates determines the probability of each remaining candidate being selected as winner.

number of boundaries inserted in each type of sequence indeed reveals that boundaries are inserted into all types of sequences. The model does, however, insert boundaries into TPPT sequences far more often (56.73%) than into PPTP (20.41%) or PTPP (13.15%) sequences. These data show that the model has a strong (but not absolute) preference for PTP words, which satisfy OCP-PLACE.

In sum, the model is inclined to segment words that respect OCP-PLACE, but also every now and then favors a different segmentation. This raises the question which model best reflects human segmentation behavior: a model that always favors PTP words, or a model that generally favors PTP words, but also produces some PPT or TPP words. Experiment 1 addresses this issue by comparing models that make different assumptions about the nature of OCP-PLACE. Several models will be tested on their ability to predict human judgments on the experimental items. Specifically, *STAGE* is compared to both a categorical interpretation of OCP-PLACE, and to a gradient interpretation of OCP-PLACE. It should be noted that the empirical studies in this chapter are only of modest size. That is, they are based on an experiment involving human judgments on 24 different items (averaged over 14 participants). Nevertheless, the analyses highlight several interesting differences between models of OCP-PLACE, and their effects on segmentation.



## 4.4 EXPERIMENT 1: MODEL COMPARISONS

The first study looks at how well STAGE is able to account for human segmentation data. STAGE induces feature-based regularities resembling OCP-PLACE, but does not show a categorical ban on PVP sequences. In order to see which segmentation strategy provides a better account of the human data, STAGE is compared to a categorical feature-based interpretation of OCP-PLACE. That is, an interpretation of OCP-PLACE which categorically rules out all sequences of consonants which have the same place of articulation. There is, however, a second interpretation of OCP-PLACE that is worth investigating. Several studies have argued that OCP-PLACE is a gradient constraint, affecting different sequences of homorganic consonants to different degrees. The strength of this gradient constraint is typically measured as the degree of under-attestation (as measured by O/E ratio) in the lexicon (Pierrehumbert, 1993; Frisch et al., 2004; Coetzee & Pater, 2008). The second comparison is thus between STAGE and a segmentation model based on O/E ratios. Specifically, STAGE will be compared to O/E ratios calculated over C(V)C sequences in continuous speech. This is done because the resulting gradient model is basically identical to the statistical learning model used in Chapter 3. As a result, the experiment sets up the same comparison as in Chapter 3, namely between a model based solely on segment probabilities, and a model based on segment probabilities plus feature-based generalization. STAGE uses O/E ratios in continuous speech as a basis for the induction of phonotactic constraints. A comparison between O/E ratios and STAGE thus assesses whether feature-based generalization improves the segmentation performance of the learner, as compared to the purely segment-based baseline model. A crucial difference with the simulations in Chapter 3 is that ‘segmentation performance’ here means the model’s fit to human data, rather than the model’s ability to predict word boundaries in corpus transcriptions.

In sum, STAGE is compared to two opposite extremes: A categorical, feature-based constraint on the one hand, and a gradient, purely statistical consonant distribution on the other hand. STAGE lies between these two extremes: It is based on consonant co-occurrence probabilities, and uses these probabilities to induce abstract feature-based constraints. Since the model retains consonant-specific constraints, the model is able to capture both segment-specific exceptions, as well as feature-based regularities.

4.4.1 *Segmentation models*

A comparison is set up between three different models: STAGE, OCP-PLACE<sub>cat</sub>, and O/E ratio. The learning mechanisms and segmentation behavior of

STAGE have been discussed in the previous section, and will therefore not be repeated here. OCP-PLACE<sub>cat</sub> is the categorical interpretation of OCP-PLACE. For the current set of simulations this constraint is simply given to the learner. The constraint categorically rules out all PVP sequences. Applying OCP-PLACE<sub>cat</sub> to the segmentation of the artificial language therefore results in a consistent PTP.PTP.PT... segmentation. Consequently, the model has an absolute preference for PTP words.

Since we are interested in the added value of feature-based generalization in explaining human data, the gradient OCP-PLACE model is based on O/E ratios calculated over continuous utterances in the Spoken Dutch Corpus. STAGE is based on these O/E values, but adds Frequency-Driven Constraint Induction and Single-Feature Abstraction in order to induce phonotactic constraints. Both STAGE and the O/E model make use of context (neighboring sequences) when making decisions about segmentation. The O/E model is thus implemented in the same way as the trough-based statistical learner from Chapter 3: The model inserts a boundary into a sequence *wxyz* whenever the O/E ratio of *xy* is lower than the O/E ratio of *wx* and *yz*. The difference between the O/E model and STAGE is thus that STAGE uses Frequency-Driven Constraint Induction and Single-Feature Abstraction for the induction of phonotactic constraints, whereas the O/E model uses consonant probabilities directly for trough-based segmentation.

#### 4.4.2 Linear regression analyses

All three models create a different segmentation of the artificial language. As explained in the Methodology section, the frequencies of test items in the model's output are counted, and used as a predictor for the human judgments on those items. Table 4.7 shows the test items, along with their wellformedness scores (taken from Table 4.1), and frequency of occurrence which was counted separately for each model. The upper half of the table contains scores and frequencies for PTP words, the lower half the values for PPT words. The counts clearly show that all models dislike PPT words: These items almost never occur in the models' segmentation outputs. While OCP-PLACE<sub>cat</sub> has an absolute ban against PPT words, the other models occasionally (but rarely) pick up such a word. The models seem to differ more radically in how they judge PTP words. While OCP-PLACE<sub>cat</sub> assigns high frequency scores to all PTP words, the other models give high frequencies to some PTP words, but low frequencies to other PTP words.<sup>5</sup> In fact, STAGE gives high frequencies to only 6 of the 12 PTP words, 4 of which belong to the most highly rated test

<sup>5</sup> The frequencies of PTP words in the categorical model vary slightly due to the randomized construction of the language.

Table 4.7: Human judgments and model output frequencies for experimental items (Experiment 1).

Item	Score	STAGE	OCP-PLACE <sub>cat</sub>	O/E ratio
madomo	0.8095	16	39	39
ponebi	0.7381	18	34	21
ponemo	0.7381	26	36	20
podomo	0.6905	26	38	17
madobi	0.5714	4	32	30
madopa	0.5714	3	25	3
ponepa	0.5714	16	35	19
podobi	0.5476	24	38	17
potumo	0.5476	4	33	23
podopa	0.4762	8	40	4
potubi	0.4524	3	37	20
potupa	0.2381	2	33	14
mobedo	0.5476	0	0	0
pabene	0.5476	2	0	0
papone	0.5000	0	0	0
mobetu	0.4524	0	0	0
papodo	0.4524	0	0	0
pabedo	0.4048	0	0	0
pamado	0.4048	1	0	0
pamatu	0.4048	1	0	0
papotu	0.3810	0	0	0
pabetu	0.3571	0	0	2
pamane	0.3333	1	0	0
mobene	0.2619	0	0	0

items by the human participants. The overall frequencies of STAGE are thus in general lower than for the other models.<sup>6</sup>

The results of the linear regression analyses are shown in Table 4.8. Stepwise regression analyses, in which predictors are added one at a time, show that STAGE explains additional variance over each of the other models (and not vice versa). That is, after taking out the variance explained by OCP-PLACE<sub>cat</sub> or O/E ratio, STAGE is still a significant predictor of the remaining variance. STAGE is thus the best predictor of the human data.

<sup>6</sup> A discussion of STAGE's general tendency to produce undersegmentations is given in Chapter 3.

Table 4.8: Results of linear regression analyses (Experiment 1).

Model	adj. $R^2$	Significance level
STAGE	0.5111	$p < 0.001$
OCP-PLACE <sub>cat</sub>	0.2917	$p < 0.01$
O/E ratio	0.3969	$p < 0.001$

*Note.* Stepwise analyses show that STAGE explains additional variance over other models:

STAGE***	+ OCP-PLACE <sub>cat</sub>	( <i>n.s.</i> )	OCP-PLACE <sub>cat</sub> **	+ STAGE**
STAGE***	+ O/E ratio	( <i>n.s.</i> )	O/E ratio***	+ STAGE**

#### 4.4.3 Discussion

The results show that STAGE is able to explain a substantial amount of the variance in the human data (about 50%). Importantly, STAGE outperforms both the categorical and the gradient interpretation of OCP-PLACE. These findings indicate that, at least for the current experiment, STAGE is a better model of human speech segmentation than a model based on a categorical OCP-PLACE, or a model based on consonant probabilities alone. In particular, the results support the view that speech segmentation by human learners is affected by both specific and abstract phonotactic constraints. STAGE produces both categorical OCP effects (PTP > PPT) and gradient OCP effects (e.g., ponemo > potubi). The mix of specific and abstract constraints that was induced by STAGE provides the best fit to human data.

The success of STAGE in accounting for the segmentation of an artificial language by human participants was obtained by a speech-based learner: The model was trained on utterances of unsegmented speech. This approach assumes that the learner does not rely on the lexicon for the induction of phonotactic constraints for speech segmentation. While the approach is a plausible model of infant phonotactic learning, since infants are only at the beginning stages of developing a mental lexicon, it should be noted that the participants in the study by Boll-Avetisyan and Kager (2008) were adults. It may thus very well be the case that the participants in segmenting the artificial speech stream relied on phonotactics that was derived from the lexicon, rather than from continuous speech. Experiment 2 therefore addresses the input issue: Which type of input provides the best account of explaining the human segmentation data? This is done by training three different versions of STAGE: a learner based on continuous speech input, a learner based on

word types in the lexicon, and a learner based on word tokens in the lexicon. In addition to investigating the most plausible input source for artificial language segmentation, the experiment allows us to investigate the effect that the various sorts of input have on the type of constraints that are induced by STAGE.

#### 4.5 EXPERIMENT 2: INPUT COMPARISONS

The second study addresses the input issue in phonotactic learning. Previous proposals for the learning of phonotactic constraints argue that constraints are the result of abstractions over statistical patterns in the lexicon (Frisch et al., 2004; Hayes & Wilson, 2008). In contrast, STAGE abstracts over statistical patterns in continuous speech. This lexicon-free approach is able to produce a variety of phonotactic constraints which have proven to be useful in speech segmentation. This raises the question which of the two sources of input is most valuable for the induction of phonotactic constraints for speech segmentation, the lexicon or continuous speech? Furthermore, the simulations in this chapter have shown that STAGE, when trained on continuous speech, induces constraints that resemble, but do not exactly match, OCP-PLACE. It is worth investigating whether a different type of input would result in a constraint set that matches OCP-PLACE more closely. The input issue is addressed by training STAGE on different input representations. By keeping all other properties of the model fixed, a clear picture of the effect of changing the input can be obtained.

##### 4.5.1 Segmentation models

Transcribed utterances in the Spoken Dutch Corpus (Goddijn & Binnenpoorte, 2003) were used to train three different versions of STAGE. The first version is the continuous speech-based learner that was also used in Experiment 1, which will be referred to as STAGE<sub>cont</sub>. The second and third versions of STAGE are trained on word types (STAGE<sub>type</sub>) and word tokens (STAGE<sub>token</sub>), respectively. In order to allow for a pure analysis of the structure of the input, the same underlying corpus was used. That is, a lexicon of word forms was constructed from the segmented corpus. The word forms were not reduced to canonical transcriptions, thereby preserving variations due to assimilations and reductions. Any differences between the models' outputs are therefore due to the structure of the input, and not due to differences in transcriptions.

The differences are found along two dimensions. First, the learner is either presented with between-word sequences, or not. Second, the learner either weighs sequences according to a word's frequency of occurrence, or not. The

Table 4.9: Models with different input structures.

Model	Input	BWS	Frequency weighting
STAGE <sub>cont</sub>	Continuous utterances	yes	yes
STAGE <sub>type</sub>	Word types	no	no
STAGE <sub>token</sub>	Word tokens	no	yes

Note. BWS = between-word sequences

different forms are best illustrated with a toy example. Consider an utterance *abc abc* in the corpus. This is a repetition of a single word *abc*. The utterance would be presented to STAGE<sub>cont</sub> as an unsegmented utterance *abcabc*. There are three different sequences in this utterance: *ab*, *bc*, and *ca*. The word-internal sequences each occur twice. In addition, there is one occurrence of the between-word sequence *ca*. The learner thus includes between-word sequences in building up a statistical distribution. Furthermore, the structure of the unsegmented speech stream is such that sequence frequencies are weighted according to the frequency of occurrence of words in the stream. Note that the learner is blind with respect to whether a sequence is a within-word or a between-word sequence. The learner simply counts occurrences of sequences in the input.

The other learners, STAGE<sub>token</sub> and STAGE<sub>type</sub>, are presented with a lexicon of isolated words which is derived from the same underlying corpus. The mini-corpus *abcabc* has the following lexicon: *abc* (2). That is, it has the word *abc*, which occurs twice. Crucially, between-word sequences are not represented in the lexicon. Such sequences do not affect the statistical distribution that is built up by the lexicon-based learners. These models only process the word-internal sequences *ab* and *bc*. The difference between type-based and token-based learning is whether or not to include the frequency of occurrence of the word. The token-based learner assigns a frequency of 2 to the sequences, whereas the type-based learner counts the sequences as a single occurrence (i.e., within a single word). Note that, since the token-based learner takes word frequencies into account, the only difference with the continuous speech-based learner is the absence of the between word sequence *ca*. The differences are summarized in Table 4.9.

The effect of the input manipulation is that each model is based on a different statistical distribution of C(V)C sequences. This results in a different constraint set for each model, and a different hypothesized segmentation of the artificial language. The question is which of the input representations

provides the best account of the human item preferences, and which constraint set most closely resembles OCP-PLACE. Note that a different distribution may require a shift of the induction thresholds that are used by STAGE. That is, the thresholds used so far ( $t_M = 0.5, t_C = 2.0$ ) worked well for the speech-based learner, but might be less suitable for the others.

In order to establish that a difference in performance between the models is due to the input structure, and not to a specific threshold configuration used, two different types of analyses were performed. The first analysis evaluates the models based on an identical threshold configuration ( $t_M = 0.5, t_C = 2.0$ ). In the second analysis, to rule out the potential advantage or disadvantage of a particular threshold configuration, the models are evaluated on a best-fit basis. Multiple simulations were run for each model. Each simulation was based on a different set of thresholds. As discussed in Experiment 2 of Chapter 3, a different set of induction thresholds causes changes in the amount of specific markedness and contiguity constraints that are induced by the learner, which subsequently affects the generalizations that are based on those constraints.

The same thresholds are considered as in Experiment 2 of Chapter 3:  $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$  as possible values for  $t_M$  (markedness constraints), and  $\{1.0, 1.11, 1.25, 1.43, 1.67, 2.0, 2.5, 3.33, 5.0, 10.0\}$  as values for  $t_C$  (contiguity constraints). Combining these thresholds exhaustively results in a total of  $10 \times 10 = 100$  different configurations. Note that the ‘standard’ configuration ( $t_M = 0.5; t_C = 2.0$ ) is exactly in the middle. The simulations will thus produce models that are both less and more conservative than this configuration. For the comparison of  $\text{STAGE}_{cont}$ ,  $\text{STAGE}_{type}$ , and  $\text{STAGE}_{token}$ , the configuration of each model which produces the best fit to the human data will be used. The result is that the models are compared on both a fixed set of thresholds ( $t_M = 0.5; t_C = 2.0$ ; Analysis 1) and a varying set of thresholds (best fit; Analysis 2).

#### 4.5.2 Linear regression analyses

As in Experiment 1, the frequencies of the test items in the segmentation outputs of the different models are used as a predictor of the human judgments on those items. The output frequencies are given in Table 4.10. The frequency counts in Analysis 1 (fixed thresholds) show that the  $\text{STAGE}_{type}$  and  $\text{STAGE}_{token}$  yield very different results.  $\text{STAGE}_{type}$  has a clear preference for PTP words, assigning high frequencies to all PTP words, and near-to-zero frequencies to PPT words. It thereby displays segmentation behavior very similar to  $\text{OCP-PLACE}_{cat}$  (see Table 4.7). In contrast,  $\text{STAGE}_{token}$  has higher frequencies for PPT words than for PTP words, and does not seem to produce any OCP-like behavior.

Table 4.10: Human judgments and model output frequencies for experimental items (Experiment 2).

Item	Score	STAGE - Analysis 1 (fixed thresholds)			STAGE - Analysis 2 (best-fit thresholds)		
		<i>cont</i>	<i>type</i>	<i>token</i>	<i>cont</i>	<i>type</i>	<i>token</i>
madomo	0.8095	15	39	0	17	39	15
ponebi	0.7381	18	29	4	18	18	4
ponemo	0.7381	23	36	5	23	36	8
podomo	0.6905	29	38	0	27	38	13
madobi	0.5714	3	29	5	7	13	5
madopa	0.5714	2	16	0	2	13	0
ponepa	0.5714	15	30	2	15	12	0
podobi	0.5476	24	32	2	24	19	4
potumo	0.5476	4	33	1	4	4	4
podopa	0.4762	8	27	0	8	24	0
potubi	0.4524	3	35	3	3	2	0
potupa	0.2381	2	22	4	2	5	0
mobedo	0.5476	0	0	20	0	0	0
pabene	0.5476	1	0	7	0	0	6
papone	0.5000	0	0	6	0	0	4
mobetu	0.4524	0	0	17	0	0	0
papodo	0.4524	0	0	16	0	0	0
pabedo	0.4048	0	1	30	0	0	0
pamado	0.4048	2	0	17	2	0	0
pamatu	0.4048	0	0	6	2	0	0
papotu	0.3810	0	0	14	0	0	0
pabetu	0.3571	0	3	21	0	0	0
pamane	0.3333	0	0	3	2	0	6
mobene	0.2619	0	0	3	0	0	0

Note. Thresholds used in Analysis 1:  $STAGE_{cont,type,token}$ :  $t_M = 0.5, t_C = 2.0$ . Best-fit thresholds used in Analysis 2:  $STAGE_{cont}$ :  $t_M = 0.5, t_C = 3.33$ ;  $STAGE_{type}$ :  $t_M = 0.4, t_C = 1.43$ ;  $STAGE_{token}$ :  $t_M = 0.3, t_C = 1.11$ .

Regression analyses for  $STAGE_{cont}$ ,  $STAGE_{type}$ , and  $STAGE_{token}$  are given in Table 4.11. The results for Analysis 1 shows that  $STAGE_{cont}$  and  $STAGE_{type}$  are significant predictors of the human wellformedness scores, but  $STAGE_{token}$  is not. Stepwise regression analyses indicate that both  $STAGE_{cont}$  and  $STAGE_{type}$  explain additional variance over  $STAGE_{token}$ . In addition, stepwise analyses indicate that  $STAGE_{cont}$  explains additional variance over  $STAGE_{type}$ : When



Table 4.11: Results of linear regression analyses (Experiment 2).

Input	STAGE - Analysis 1 (fixed thresholds)		STAGE - Analysis 2 (best-fit thresholds)	
	adj. $R^2$	Significance level	adj. $R^2$	Significance level
<i>cont</i>	0.4790	$p < 0.001$	0.5026	$p < 0.001$
<i>type</i>	0.3826	$p < 0.001$	0.5721	$p < 0.001$
<i>token</i>	0.0798	<i>n.s.</i>	0.4596	$p < 0.001$

Note. Stepwise analyses:

**Analysis 1:**

<i>cont</i> *	+ <i>type</i>	( <i>n.s.</i> )	<i>type</i> ( <i>n.s.</i> )	+ <i>cont</i> *
<i>cont</i> **	+ <i>token</i>	( <i>n.s.</i> )	<i>token</i> ( <i>n.s.</i> )	+ <i>cont</i> **
<i>type</i> **	+ <i>token</i>	( <i>n.s.</i> )	<i>token</i> ( <i>n.s.</i> )	+ <i>type</i> **

**Analysis 2:**

<i>cont</i> ( <i>n.s.</i> )	+ <i>type</i>	( <i>n.s.</i> )	<i>type</i> ( <i>n.s.</i> )	+ <i>cont</i> ( <i>n.s.</i> )
<i>cont</i> *	+ <i>token</i> *	( <i>n.s.</i> )	<i>token</i> *	+ <i>cont</i> *
<i>type</i> **	+ <i>token</i>	( <i>n.s.</i> )	<i>token</i> ( <i>n.s.</i> )	+ <i>type</i> **

STAGE<sub>type</sub> is entered first into the regression model, then STAGE<sub>cont</sub> is still a significant predictor ( $p < 0.05$ ). The reverse is not true. In sum, Analysis 1 shows that STAGE<sub>cont</sub> again is the best predictor of the human data from the artificial language learning experiment by Boll-Avetisyan and Kager (2008).

Inspection of the constraint sets used by STAGE<sub>type</sub> and STAGE<sub>token</sub> during segmentation reveals the source of the difference in performance between the two models. The set of constraints used by STAGE<sub>type</sub> closely matches OCP-PLACE: The majority of the constraints restrict the occurrence of PVP sequences (see Table 4.12). In addition, there are some relatively low ranked constraints disfavoring dVp and tVp, resulting in somewhat lower occurrence frequencies for items violating these constraints (such as *madopa*, *potupa*). The constraint set induced by STAGE<sub>token</sub> shows a completely different picture (see Table 4.13). Here the most high-ranked constraints include constraints against TVP sequences (most notably against dVP). This explains why the model has a preference for PPT segmentations: It often inserts a boundary between coronals and labials, resulting in PT.PPT.PPT... segmentations. Interestingly, the model also induces a contiguity constraint CONTIG-IO([n]V[m]), which aims to preserve nVm sequences. This constraint has the effect that, while in general PPT words are preferred, words that end with an /n/-initial syllable (e.g., *pabene*) are dispreferred. The contiguity constraint militates against the insertion of a boundary after such syllables. Since few boundaries are posited

Table 4.12: Constraints used in segmentation of the artificial language by  $\text{STAGE}_{type}$  ( $t_M = 0.5, t_C = 2.0$ ).

CONSTRAINT	RANKING	FREQUENCY	
		Count	Percentage
*[m]V[m]	308.8524	27	(11.1%)
*[b]V[m]	289.7448	27	(11.1%)
*[p,b]V[m]	249.5673	27	(11.1%)
*[m]V[p,f]	180.4863	27	(11.1%)
*[b]V[p]	161.8902	27	(11.1%)
*[m]V[p,b,f,v]	155.1712	27	(11.1%)
*[p,b,f,v]V[p]	117.6980	27	(11.1%)
*[b,d,v,z]V[p]	92.2151	8	(3.3%)
*[p,b,t,d,f,v,s,z]V[p]	69.5577	8	(3.3%)
*[p,b,f,v]V[p,t,k,f,s,j,x]	64.1567	10	(4.1%)
*[b,d,v,z]V[p,b]	53.4554	19	(7.8%)
*[p,b,f,v]V[p,b,f,v]	43.3952	15	(6.2%)

*Note.* The frequency of use (FREQUENCY) indicates how many *wxyz*-sequences in the language (out of a total of 243) are affected by the constraint.

after /n/, there are few words ending with an /n/-initial syllable in the model's segmentation output.

In sum, it seems that, while  $\text{STAGE}_{type}$  most closely resembles the predictions of a categorical OCP-PLACE constraint, it does not match the gradient preferences of participants in the experiment as well as  $\text{STAGE}_{cont}$  does. The latter obtains a better match to the data by distinguishing between the well-formedness of different PTP words. In addition, the token-based learner fails in accounting for the human data. One question that comes up is *why* the token-based performs so much worse than the type-based learner.

As mentioned earlier, the difference between the models may be caused by the fact that the statistical distributions are different, while at the same time the induction thresholds are the same for all models. It could be the case that OCP-PLACE is learnable by a token-based model if the statistical distribution is simply cut off at different points. In addition, we may get better  $\text{STAGE}_{cont}$  and  $\text{STAGE}_{type}$  models simply by using different induction thresholds. Analysis 2 is therefore concerned with models that have been optimized for their threshold configuration. The best fit to the human data was found for the following configurations:  $\text{STAGE}_{cont}$ :  $t_M = 0.5, t_C = 3.33$ ;

Table 4.13: Constraints used in segmentation of the artificial language by  $\text{STAGE}_{\text{token}}$  ( $t_M = 0.5, t_C = 2.0$ ).

CONSTRAINT	RANKING	FREQUENCY	
		Count	Percentage
*[d]V[m]	2406.2062	27	(11.1%)
*[b,d]V[m]	1635.3616	26	(10.7%)
*[m]V[m]	1559.1889	20	(8.2%)
CONTIG-IO([n]V[m])	1224.5688	19	(7.8%)
*[d]V[p,f]	1218.6504	25	(10.3%)
*[d]V[p,b]	1049.2083	19	(7.8%)
*[p,b,t,d]V[m]	925.8681	36	(14.8%)
*[t,d]V[b]	803.8574	18	(7.4%)
*[m]V[p,f]	789.6689	10	(4.1%)
*[t,d]V[p,b]	695.9537	17	(7.0%)
*[b,d,v,z]V[p,t,k]	684.5911	12	(4.9%)
*[m]V[p,b]	679.8727	8	(3.3%)
*[p,b]V[m]	648.6330	1	(0.4%)
*[m,n]V[b]	533.5175	8	(3.3%)
*[p,b,t,d,f,v,s,z]V[p,t,k]	524.2290	8	(3.3%)
*[p,b,f,v]V[t,d,s,z]	496.0740	2	(0.8%)
*[m,n]V[p,b]	457.2829	6	(2.5%)
*[b,d,v,z]V[p,b]	362.2443	1	(0.4%)
*[p,b,t,d,f,v,s,z]V[p,b,t,d]	338.8520	1	(0.4%)

*Note.* The frequency of use (FREQUENCY) indicates how many  $wxyz$ -sequences in the language (out of a total of 243) are affected by the constraint.

$\text{STAGE}_{\text{type}}$ :  $t_M = 0.4, t_C = 1.43$ ;  $\text{STAGE}_{\text{token}}$ :  $t_M = 0.3, t_C = 1.11$ . The output frequencies of the models are shown in Table 4.10 (Analysis 2).

Interestingly, the configuration of  $\text{STAGE}_{\text{cont}}$  was already optimal: The output frequencies are nearly identical to those in the earlier runs of the model. It should be noted that the small deviations found for the different runs of  $\text{STAGE}_{\text{cont}}$  are not due to a different constraint set, but rather to a small number of sequences for which there is no single optimal segmentation candidate. In such cases a winner is chosen at random. For this reason the exact frequency of a test item may vary slightly between different runs. Also note that the value for  $t_C$  is irrelevant here, since only markedness constraints are used for segmentation of the artificial language. The constraint sets used

Table 4.14: Constraints used in segmentation of the artificial language by  $\text{STAGE}_{type}$  ( $t_M = 0.4$ ,  $t_C = 1.43$ ; best fit).

CONSTRAINT	RANKING	FREQUENCY	
		Count	Percentage
CONTIG-IO([m]V[n])	938.8947	27	(11.1%)
*[m]V[m]	308.8524	27	(11.1%)
CONTIG-IO([n]V[m])	299.5556	20	(8.2%)
*[b]V[m]	289.7448	27	(11.1%)
CONTIG-IO([m]V[d])	285.7460	13	(5.3%)
*[p,b]V[m]	249.5673	27	(11.1%)
*[m]V[p,f]	180.4863	27	(11.1%)
*[m]V[p,b,f,v]	155.1712	27	(11.1%)
*[p,b,f,v]V[p,t,k,f,s,j,x]	36.3611	70	(28.8%)
CONTIG-IO([t,d,s,z]V[C])	32.2084	20	(8.2%)
CONTIG-IO([C]V[p,b,f,v])	24.5104	54	(22.2%)
*[b,v]V[C]	23.5896	14	(5.8%)
CONTIG-IO([C]V[p,b,t,d])	23.1961	2	(0.8%)

*Note.* The frequency of use (FREQUENCY) indicates how many *wxyz*-sequences in the language (out of a total of 243) are affected by the constraint. V = vowel, C = obstruents = [p,b,t,d,k,g,f,v,s,z,j,ʒ,x,ʏ,h,ç].

by  $\text{STAGE}_{cont}$  in Experiments 1 and 2 (both Analysis 1 and 2) are thus the same.

The type-based learner ( $\text{STAGE}_{type}$ ) also resembles its counterpart from Analysis 1, but assigns lower frequencies to several PTP words. The model therefore makes some clear distinctions between the predicted wellformedness of different PTP words. The token-based learner's optimal performance was obtained by making the model extremely conservative. It assigns low frequencies (and many zero-frequencies) to most words, with somewhat higher frequencies for *madomo* and *podomo*. The results of the linear regression analyses for these models are shown in Table 4.11 (Analysis 2). The performance of the models in terms of their optimized  $R^2$  are relatively similar. Step-wise analyses indicate that the only conclusion that can be drawn is that  $\text{STAGE}_{type}$  is a better predictor than  $\text{STAGE}_{token}$ . This result, together with the low frequencies in the model's output, leads to the conclusion that the token-based learner again fails to accurately predict the human wellformedness scores. The token-based learner does not match with the human data as well as the other models do. Furthermore, the model's constraint set does

Table 4.15: Constraints used in segmentation of the artificial language by  $\text{STAGE}_{\text{token}}$  ( $t_M = 0.3, t_C = 1.11$ ; best fit).

CONSTRAINT	RANKING	FREQUENCY	
		Count	Percentage
CONTIG-IO([m]V[n])	6511.2649	27	(11.1%)
CONTIG-IO([b,d]V[t])	5347.2605	27	(11.1%)
CONTIG-IO([m]V[t])	5098.1932	27	(11.1%)
CONTIG-IO([b,d]V[t,d])	3653.2264	27	(11.1%)
CONTIG-IO([m]V[t,d])	3483.0647	27	(11.1%)
CONTIG-IO([p,b,t,d]V[t,s])	2100.9127	27	(11.1%)
CONTIG-IO([p,t,k,f,s,f,x]V[n])	1833.5053	27	(11.1%)
CONTIG-IO([p,b,t,d]V[t,d,s,z])	1389.6902	27	(11.1%)
CONTIG-IO([b,d]V[p,b,t,d,f,v,s,z])	1363.1418	92	(37.9%)
CONTIG-IO([n]V[m])	1224.5688	27	(11.1%)
CONTIG-IO([C]V[n])	1206.8609	27	(11.1%)
CONTIG-IO([p,b,t,d]V[p,t,f,s])	1063.6762	47	(19.3%)
CONTIG-IO([p,t,k,f,s,f,x]V[m,n])	1048.7304	53	(21.8%)
*[b]V[m]	864.5170	27	(11.1%)
*[m]V[p,f]	789.6689	27	(11.1%)
CONTIG-IO([p,b,t,d]V[p,b,t,d,f,v,s,z])	710.5020	45	(18.5%)
*[m]V[p,b]	679.8727	27	(11.1%)
CONTIG-IO([C]V[m,n])	661.1707	12	(4.9%)
*[m,n]V[b]	533.5175	18	(7.4%)
*[m,n]V[p,b]	457.2829	24	(9.9%)

*Note.* The frequency of use (FREQUENCY) indicates how many  $wxyz$ -sequences in the language (out of a total of 243) are affected by the constraint. V = vowel, C = obstruents = [p,b,t,d,k,g,f,v,s,z,f,ʒ,x,ɣ,h,ʧ].

not resemble OCP-PLACE. The conservativeness of the model is illustrated by the constraint set that was used by the model: The model relied on a large number of contiguity constraints (see Table 4.15). Since contiguity constraints aim to preserve sequences, the result is a hypothesized segmentation that contains very few word boundaries.

There is no significant difference between the performances of  $\text{STAGE}_{\text{cont}}$  and  $\text{STAGE}_{\text{type}}$ . Inspection of their constraint sets does reveal some differences, however. Whereas  $\text{STAGE}_{\text{cont}}$  exclusively uses markedness constraints (Table 4.5),  $\text{STAGE}_{\text{type}}$  has both markedness constraints against PVP, and contiguity constraints preserving various sequences (Table 4.14). The net effect is that

this model favors PTP words, but with some exceptions, such as words starting with pVt.  $\text{STAGE}_{cont}$  also has low frequencies for words starting with pVt, but in addition also has low frequencies for words starting with mVd. These models, like the human participants, thus show a general OCP-PLACE effect, resulting in a general preference for PTP words. At the same time, however, both humans and the models show effects of gradience: Not all PTP words are judged equally wellformed. It appears that such finer-grained distinctions make these models a better model of the human segmentation data.

#### 4.5.3 Discussion

The results of Experiment 2 show that a continuous-speech-based learner has a better fit than a type-based learner (Analysis 1) or a fit comparable to a type-based learner (Analysis 2) in predicting human preferences in the artificial language segmentation experiment by Boll-Avetisyan and Kager (2008). The success of both the type-based and the continuous speech-based model can be attributed to the mixture of abstract and specific constraints. This allows the models to capture both general tendencies and fine-grained preferences in the human data. Interestingly, a token-based learner fails, regardless of the specific threshold configuration that is used.

In order to explain the failure of the token-based learner, a more detailed examination of the statistical distribution is required. The difference between the type-based and the token-based learner is whether or not the learner is sensitive to word frequencies. A closer look at the lexicon (based on the *Corpus Gesproken Nederlands*) that was used by the token-based learner shows that the failure can be attributed to the phonotactic structure of the most highly frequent words in the lexicon. The most highly frequent words containing CVC sequences are function words that consist of TVT sequences (e.g., /dat/ ‘that’, /nit/ ‘not’, /dan/ ‘then’), or PVT sequences (e.g., /fan/ ‘of’, /mɛt/ ‘with’, /mar/ ‘but’). The use of token frequencies results in an inflation of the difference in the degree of attestedness between sequences that occur in function words, and sequences which do not occur in function words. Sequences in TVT function words (such as /dat/ and /dan/) are overrepresented, which in effect makes sequences which do not form function words (e.g., TVP sequences such as /dap/ and /dam/) relatively underrepresented. Depending on the thresholds that are used, this could lead to the induction of markedness constraints against TP sequences and/or the induction of contiguity constraints favoring TT sequences. Similarly, inflation of the frequencies of PVT function words (/fan/, /mɛt/) causes underattestation of sequences of PP consonants due to relatively low occurrence frequencies of such sequences (/fam/, /mɛp/). In this case, the frequency boost leads to markedness constraints against PP

sequences, and/or contiguity constraints favoring PT sequences. Examples of these types of constraints are indeed found in the constraint sets that are used by the token-based learner (e.g., \*TP and \*PP constraints in Table 4.13; CONTIG-IO(PT) and \*PP constraints in Table 4.15).

The high frequencies of CVC function words also has an effect on the ranking of the constraints. Specifically, the high-frequency function words involve coronals, which leads to an increase in the individual segment frequencies of coronals, and thus has the effect that the expected frequencies of sequences involving coronals increases. The consequence is that constraints affecting coronals will in general be ranked higher than sequences affecting labials exclusively. This again can be seen in the constraint sets that are used by the token-based learner during segmentation. Constraints affecting coronals tend to be ranked higher than constraints affecting labials (\*TP  $\gg$  \*PP, in Table 4.13; CONTIG-IO(PT)  $\gg$  CONTIG-IO(PP), \*PP, in Table 4.15).

While the phonotactic structure of CVC function words can account for the difference between a type-based and a token-based learner, it does not explain why the continuous speech-based learner performs so well. Since the speech-based learner counts all occurrences of sequences in the speech stream, it is in fact also a token-based learner. The crucial difference, however, is that the speech-based learner includes between-word sequences in the statistical distribution. Occurrences of between-word sequences apparently neutralize the inflation of the distribution that was found in the token-based learner. Specifically, they have the effect of boosting the frequency counts of sequences that do not occur in function words. That is, while sequences such as dVp or tVp do not occur in highly frequent function words, they do occur frequently between words. In fact, the frequency neutralization effect is possibly, at least in part, due to function words that consist of 2 segments, occurring in sequence with labial-initial words. For example, dVp benefits in terms of frequency from the fact that /də/ is a highly-frequent function word, and that there are many Dutch words starting with /p/. Since 2-segment function words typically do not start with labials, the attestedness of PVP sequences is affected to a lesser degree. As a consequence, there is still evidence for OCP-PLACE in the speech stream, and OCP-PLACE is thus learnable from continuous speech. Further research on different languages would in the end have to show to what extent these results are specific to Dutch, and to what extent the learning approach presented here provides a general account for the induction of phonotactic constraints, such as OCP-PLACE.

## 4.6 GENERAL DISCUSSION

The current chapter aimed to provide a constraint induction account of the learning of OCP-PLACE. The speech-based learning approach that was developed in the previous chapters was extended to the induction of abstract, feature-based constraints on non-adjacent consonants (i.e., consonants with intervening vowels). The model, STAGE, uses no explicit mechanism for similarity avoidance and assumes that constraints are induced without referring to the lexicon as a basis for phonotactic learning. The main theoretical issue was whether STAGE, when applied to non-adjacent consonants in transcriptions of continuous speech utterances, could induce constraints that resemble OCP-PLACE. The empirical part of this chapter was concerned with investigating whether STAGE is able to account for the OCP effects on speech segmentation that were found in a study by Boll-Avetisyan and Kager (2008) using artificial language learning experiments.

Experiment 1 compared STAGE to both categorical and gradient interpretations of OCP-PLACE. It was found that the output produced by STAGE better resembles human preferences for experimental items than a categorical feature-based OCP-PLACE constraint banning all homorganic consonants. In addition, it was found that STAGE also performs better than a gradient model based on consonant probabilities. Since STAGE induces both general, feature-based constraints and finer-grained segment-based constraints, the results support the view that segmentation involves both specific and abstract phonotactic constraints. In Experiment 2, the question was addressed which type of input structure best accounts for the induction of OCP-PLACE, and its effect on segmentation. The speech-based version of STAGE performed better than (Analysis 1) or comparable to (Analysis 2) a version of STAGE that was trained on word types. Both models outperformed a version of the model that was trained on word tokens.

The results in this chapter provide the first support for the psychological plausibility of STAGE. While the simulations in Chapter 3 were evaluated on the ability to accurately predict the locations of word boundaries in a corpus of unsegmented speech, the simulations in this chapter focused on the ability of STAGE to accurately predict human segmentation behavior. The empirical finding that neither categorical nor gradient interpretations of OCP-PLACE, and also neither type-based nor token-based versions were better predictors of the human data than STAGE trained on continuous speech, combined with the fact that the mechanisms used by STAGE are based on learning mechanisms that have been shown to be available to human language learners, indicates that STAGE has some potential as a model of human phonotactic learning and speech segmentation.



In Chapter 3, it was found that adding feature-based generalizations to the statistical learning of phonotactic constraints improves the segmentation performance of the learner. Importantly, Experiment 1 of the current chapter demonstrates the added value of feature-based generalization in accounting for human segmentation data. It was found that generalization improves the fit to human data, as compared to models that rely on pure consonant distributions. Since the lexicon-free approach also outperforms a single, pre-defined OCP-PLACE, the results suggest that the success of the approach can be attributed to the mix of specific and abstract constraints. Also in Experiment 2, the success of both the continuous speech-based model and the type-based model can be attributed to the mixture of specific and general constraints.

These results complement the findings of the previous chapter: Feature-based generalizations improve the segmentation performance of the learner in both corpus simulations and simulations of human data. The findings of Chapter 3 and Chapter 4 thus provide converging evidence that the segmentation of continuous speech using phonotactics involves both specific and abstract phonotactic constraints. The success of STAGE as a model of the induction of phonotactic constraints for speech segmentation can thus be attributed to the learning mechanisms it employs: segment-based statistical learning and feature-based generalization. The resulting mix of specific and more general constraints has proven to be successful in both types of empirical studies that have been conducted, corpus simulations and simulations of human data.

The second topic of interest concerns the comparison of models that vary in the assumptions they make about the input (continuous speech, word types, word tokens; Experiment 2). This issue is closely related to the investigation of the speech-based learning (SBL) hypothesis, as outlined in Chapter 1. The SBL hypothesis states that phonotactics is induced from continuous speech input, and subsequently facilitates the formation of the mental lexicon through the detection of word boundaries in continuous speech. In contrast, the lexicon-based learning (LBL) approach assumes that phonotactics is learned from the lexicon, and thus is the result of, rather than a prerequisite for, word learning. In line with the SBL hypotheses, it was attempted to induce OCP-PLACE from continuous speech input. STAGE, when applied to C(V)C sequences in unsegmented utterances, induces a constraint set which contains constraints with varying levels of generality. The set contains constraints that resemble but do not exactly match OCP-PLACE. Some of the constraints are more specific, some are more general and affect a different natural class not exclusively consisting of labial sequences. The fact that the constraint set to a large degree mimics the behavior of OCP-PLACE, and even outperforms categorical and gradient interpretations of OCP-PLACE in simulating human data, provides support for the SBL approach. Specifically, it shows that neither a similarity avoid-

ance mechanism nor a lexicon are absolute requirements for the induction of OCP-PLACE. Our account of the learning of OCP-PLACE therefore partly corresponds to the proposal by Frisch et al. (2004). Our account shares with theirs that similarity avoidance need not be encoded in the synchronic grammar, but rather may be learned indirectly through abstractions over statistical patterns in the input. However, our account adds to theirs the possibility that OCP-PLACE need not be the result of abstractions over patterns in the *lexicon*. Our simulations show that it is also conceivable that OCP-PLACE is learned through abstractions over statistical patterns found in the learner's direct speech input, without referring to internal representations of word forms in the mental lexicon.

A strong test for the SBL approach is the comparison between different input structures in Experiment 2. This directly contrasts speech-based learners with lexicon-based learners. The comparison was intended to highlight qualitative differences between input consisting of continuous speech, word types, and word tokens. While the statistical distribution produced by word tokens is qualitatively different from distributions that are derived from both continuous speech and word types, the distributions of continuous speech and word types are relatively similar. Both models produced constraint sets that resulted in similar segmentation behavior. On the basis of the current set of results, no conclusion can be drawn with respect to which of these models provides a better account of the induction of OCP-PLACE, and of its effect on human speech segmentation. The type-based learner and the continuous speech-based learner, however, make radically different assumptions with respect to the learner's learning capacities. The type-based (or any lexicon-based) learner assumes that the learner has already built up a lexicon, and derives phonotactic constraints from patterns in the lexicon. This requires the *a priori* ability to segment speech before phonotactic learning can take place. In contrast, the continuous speech-based learner induces constraints from unsegmented input, and uses phonotactics in the construction of the lexicon. While it is still conceivable that a type-based learner was responsible for the phonotactic knowledge used by adult participants in the study by Boll-Avetisyan and Kager (2008), such a model is not a very plausible account of phonotactic learning by infants, who are only at the beginning stages of developing a mental lexicon.

The speech-based learner is compatible with findings from studies on infant phonotactic learning, and at the same time is able to account for adult segmentation data. This view thus provides a unified, lexicon-free approach to phonotactic learning in which adults and infants rely on the same phonotactic learning and segmentation mechanisms. More research is needed to find cases where the two models make clearly different predictions, which can be tested

empirically. For now, the conclusion is that speech-based learning is able to account for the induction of phonotactic constraints, and for their usage in speech segmentation.

A possible direction for future research on the input issue would be a comparison of STAGE to other lexicon-based learning models. In the current study, all models were based on the same learning mechanisms. This allowed for a comparison that zooms in on the role of differently structured input in phonotactic learning and speech segmentation. While STAGE as a general model is based on the idea of abstractions over statistical patterns in the input, and can thus be applied to both segmented and unsegmented input, it is not known how STAGE would perform when compared to other lexicon-based learning models. An interesting case would for example be to train the Hayes and Wilson (2008) learner on the Dutch lexicon and see to what extent the model induces constraints that resemble OCP-PLACE. While the model by Hayes and Wilson was not designed to make predictions about speech segmentation, it could still be trained on the same data as STAGE was. Since their model produces constraints that are accompanied by numerical weights, one could even imagine feeding the constraints and weights to the OT segmentation model and inspecting its predictions with respect to the locations of word boundaries. Conversely, STAGE could be adjusted to predict the wellformedness judgments on which the Hayes and Wilson model was evaluated.

The finding that a model based on type frequencies better accounts for human phonotactic learning than a model based on token frequencies is in line with a substantial body of research that advocates word types as the main source for phonotactic learning (Pierrehumbert, 2003; Hay, Pierrehumbert, & Beckman, 2004; Albright, 2009; Richtsmeier, Gerken, & Ohala, 2009; Hamann & Ernestus, submitted). While effects of token frequency have been found (e.g., Bailey & Hahn, 2001), type frequencies have been shown to be essential for phonotactic learning (Richtsmeier et al., 2009; Hamann & Ernestus, submitted). The current study provides additional evidence in favor of type frequencies as compared to token frequencies in phonotactic learning. While type and token frequencies are correlated, and the advantage of type frequencies over token frequencies has typically been reported as being of modest size (e.g., Albright & Hayes, 2003; Hayes & Wilson, 2008; Albright, 2009), the simulations described in this chapter indicate that the use of type or token frequencies can lead to qualitatively different results. At least for the learning of constraints on non-adjacent consonants, different statistical distributions can lead to different generalizations built upon those distributions (such as the \*TP constraints induced by the token-based learner).

In addition, our study investigates a third potential source of phonotactic learning, which has not been considered in earlier studies. While a learner based on token frequencies fails, our learner based on consonant sequence frequencies in continuous speech has a comparable performance to a type-based learner in simulating the induction of OCP-PLACE, and in accounting for its effect on speech segmentation by human participants. It would be interesting to push the speech-based learning hypothesis further and see to what extent the approach is able to account for the induction of phonotactics in general. This would crucially involve setting up a comparison where a type-based learner and a continuous speech-based learner would make different predictions with respect to phonotactic learning.

A final note on the different input structures concerns a more principled distinction between type and token frequencies. In the current study, we assumed that the distinction between type and token frequencies was merely quantitative. Token frequencies result in a boost of the frequencies of sequences that occur in highly frequent words. In this view, since STAGE is simply counting occurrences across the lexicon, there is no principled distinction between a sequence occurring 2 times in 80 different words and the same sequence occurring 4 times in 40 different words. Both result in a total of 160 occurrences across the lexicon and thus would lead to equivalent results in our model. A recent study by Hamann and Ernestus (submitted), however, showed that the former improved adults' learning of phonotactic restrictions, whereas the latter did not. This suggests a more principled distinction between type and token frequencies than is commonly assumed. Hamann, Apoussidou, and Boersma (to appear) propose that this difference can be accounted for by incorporating a semantic level into the model, along the lines of Apoussidou (2007).

The constraint set induced by STAGE revealed an interesting property of the abstraction mechanism of the model. Part of the success of the continuous speech-based learner was due to the induction of constraints that behave as alignment constraints during segmentation (e.g., \*[C]V[p,t], see Table 4.5). These constraints are in fact markedness constraints where one consonant position is free and the other restricted. For example, a constraint \*Xp states that whenever /p/ occurs, it should be preceded by a word boundary, regardless of the preceding consonant. In other words, /p/ should be word-initial. Not inserted a boundary before /p/ (and thus not treating /p/ as a word-initial consonant) would constitute a violation of the markedness constraint. Such single-sided generalizations are an automatic consequence of abstraction over statistically induced segment-specific constraints. The model thus constructs constraints that resemble alignment constraints without any specific bias or new constraint type. The effect of a new type of constraint (alignment) is sim-

ply due to the statistical patterns in the input, combined with the feature-based generalization mechanism.

To conclude, the current chapter provides evidence that `STAGE` is able to induce constraints that resemble `OCP-PLACE`. Positive results were found for versions of the model that were trained either on continuous speech, or on type frequencies. Both models induced general `OCP-PLACE` constraints, augmented with constraints favoring or restricting specific consonant co-occurrences. This allowed the models to capture both general tendencies and fine-grained preferences in the human segmentation data. The chapter thus provides additional evidence for the role of feature-based generalizations in speech segmentation, and further highlights the potential usefulness of continuous speech as a source for phonotactic learning.



# 5

## THE INDUCTION OF NOVEL PHONOTACTICS BY HUMAN LEARNERS

---

The remainder of this dissertation is concerned with extending the psycholinguistic evidence for the Speech-Based Learning hypothesis. Computer simulations in the previous chapters have shown that phonotactics can be learned from continuous speech through a combination of statistical learning and generalization. This approach was shown to be effective in simulations of speech segmentation using unsegmented corpora (Chapter 3), and could account for human data concerning the segmentation of continuous artificial languages (Chapter 4). The current chapter aims at providing direct evidence that human learners can learn phonotactics from continuous speech, by setting up a series of phonotactic learning experiments in which adult participants are tested on their ability to induce novel phonotactic constraints from a continuous stream of speech from an artificial language. The speech streams are constructed in such a way that it is unlikely that participants would derive phonotactics from statistically learned word forms. In addition, this chapter aims at providing evidence for an assumption that was made in Chapter 4, namely that learners have the ability to ignore intervening vowels for the induction of constraints on non-adjacent consonants. This processing assumption was made to allow for the induction of OCP-PLACE by STAGE. The current chapter examines to what extent human learners can induce novel phonotactics of this kind as a result from exposure to continuous speech input.

In Section 5.1, an overview is given of earlier psycholinguistic work on phonotactic learning, statistical learning, and feature-based generalization. Sections 5.2 and 5.3 describe two artificial language learning experiments which focus on providing new evidence for the statistical learning of constraints on non-adjacent consonants. Section 5.4 describes the third experiment which examines whether learners induce feature-based generalizations from continuous speech. Finally, the implications of the findings as well as suggestions for future work are given in Section 5.5.

### 5.1 INTRODUCTION

Experimental studies addressing the induction of phonotactics by human learners show that a variety of constraints can be induced, and are learned by both adults and infants from exposure to language data conforming to

these constraints. These studies provide insight into the types of phonotactic constraints that can be learned by human learners, and into the level of abstraction at which these constraints are represented.

Onishi et al. (2002) show that adult learners can learn phonotactic regularities from a set of training words after only several minutes of exposure. The study examined the learning of first-order phonotactic constraints, in which specific consonants were confined to either word-initial or word-final position (e.g., /bæp/, not \*/pæb/), and second-order constraints, in which consonant-vowel sequences were linked (e.g., /bæp/ or /pɪb/, but not \*/pæb/ or \*/bɪp/). During familiarization participants heard CVC words that showed the phonotactic restriction. In the test phase, participants were faster at repeating novel test words which obeyed the experiment-induced constraints, than test words which violated those constraints. These effects were found both for the learning of first-order constraints, and for the learning of second-order constraints.

In a follow-up study by Chambers, Onishi, and Fisher (2010), adult participants were trained on CVC words with phonotactic constraints on initial and final consonants, and were subsequently tested on words containing vowels not heard during familiarization. Learners responded more rapidly to legal words with a novel vowel than to illegal words with that same vowel. Their results show that learners are able to abstract across vowel contexts, and are able to learn proper first-order constraints. The study thereby provides evidence that consonants and vowels are processed independently in phonotactic learning (see also, Newport & Aslin, 2004; Bonatti et al., 2005). Phonotactic generalizations in which consonants are restricted to onset or coda position have been shown to be learnable by 16.5-month-old infants. Infants listen longer to unstudied illegal syllables than to unstudied legal syllables (Chambers et al., 2003). It thus appears that both adults and infants can induce constraints that link consonants to specific positions.

A study by Saffran and Thiessen (2003) indicates that infants more easily learn phonotactic patterns when positions are assigned to segments from the same natural class, than when assigned to unrelated segments. During familiarization, 9-month-old infants heard AVBAVB words, where A and B were natural classes of voiceless stops (p, t, k) and voiced stops (b, d, g), separated by vowels (V). Infants were able to distinguish novel test items that followed the voicing pattern to which they had been familiarized from test items that followed the opposite voicing pattern. In contrast, when the positions contained featurally mixed classes (p, d, k in one position; b, t, g in the other position) infants did not learn the pattern to which they had been exposed. These findings suggest that phonotactic constraints on natural classes may be easier to learn than constraints on arbitrary segment classes.



One aspect that all of the above-mentioned studies have in common is that participants are tested on their ability to generalize the experimentally induced phonotactic constraint to novel words (i.e., to items that did not occur in the familiarization phase). That is, in order to assess whether participants have abstracted the phonotactic structure of the words, as opposed to memorizing training items as a whole, participants are tested on novel items conforming to the same phonotactic regularities as the training words. Such a test phase rules out the possibility that participants were simply learning words, rather than learning phonotactics. In order to test whether participants represent phonotactic constraints at the level of the phonological feature, an even stronger test for generalization is needed. Since feature-based phonotactic constraints affect natural classes of segments, a proper test for feature-based abstraction would be to familiarize participants with items that contain segments from a subset of the natural class, and test them on items that contain a different subset of the natural class. In other words, test items should be novel words that contain novel segments, which were not heard during familiarization.

Seidl and Buckley (2005) exposed 9-month-old infants to training words that followed a pattern at the level of natural classes. In their first experiment, two different artificial languages were constructed. One followed a natural (i.e., phonetically grounded) pattern in which oral stops were restricted to word-initial position, and fricatives and affricates were restricted to intervocalic position (e.g., /pasat/). The other language consisted of the reverse positional structure, resulting in an arbitrary (i.e., not grounded) pattern of word-initial fricatives and affricates, followed by intervocalic stops (e.g., /sapat/). They found that both the natural and the arbitrary patterns were learned by infants after only a few minutes of exposure. Similar results were found in their second experiment for the learning of consonant-vowel sequences sharing place features (natural pattern) and sequences not sharing place of articulation (arbitrary pattern). The results are consistent with the view that infants can learn a phonotactic pattern, either natural or arbitrary, when the pattern affects a natural class of segments. However, while the test words contained some novel consonants from the same natural class, in addition to familiar consonants heard during training, the effect of these novel segments was not analyzed independently from words with familiar consonants. The results could therefore have been driven solely by the occurrence of familiar consonants in the test items.

In a follow-up study, Cristià and Seidl (2008) used test words in which the restricted position was made up exclusively of consonants that had not occurred during familiarization. In their study, 7-month-old infants were exposed to CVC words from an artificial language that either had onsets with segments that formed a natural class (plosives + nasals, which share

the specification ‘–continuant’) or onsets with segments that did not form a natural class (fricatives + nasals, which do not form a coherent class). During the test phase, infants listened to words that contained novel plosives and fricatives. When trained on the natural class, infants distinguished between words that did or did not conform to the phonotactic structure of the training language. In contrast, infants trained on the incoherent class were not able to distinguish legal from illegal test items. In a control experiment, they showed that the difference in learning difficulty was not due to an inherent difference between plosives and fricatives. These results provide evidence that infants represent phonotactic generalizations at the level of phonological features, and support the view that infants learn phonotactic constraints which are more general than the specific segments that were used during training.

Similar results were obtained in other studies of phonological learning by infants. White et al. (2008) show that infants can learn stop and fricative voicing alternations. Infants were presented with distributional information that could be used to infer that two segments were related by a phonological process (and hence refer to a single underlying phoneme). While the responses of 8.5-month-old infants were found to be driven by transitional probabilities (and not by the phonological alternation), 12-month-old infants showed the capacity to generalize two segments in complementary distribution into a single phonemic category. In addition, Maye et al. (2008) show that exposing 8-month-old infants to a bimodal distribution of Voice Onset Time (VOT) for one place of articulation (e.g., *da/ta* contrast for dentals) facilitates the discrimination of a voicing contrast for a different place of articulation (e.g., *ga/ka* contrast for velars). These findings are consistent with the view that infants construct generalizations at the level of an abstract feature *voice* as a result of exposure to a specific training items.

Further evidence for feature-based generalization comes from a study by Finley and Badecker (2009) who show that adult learners are able to learn patterns of vowel harmony which generalize to novel vowels in test items. Participants were trained on alternations that displayed front/back vowel harmony. The subset used for training consisted of either low or mid vowels. The test items contained members of the vowel subset (low or mid) not used during training. In control conditions, participants heard stems only, and thus received no evidence of vowel harmony in alternations. They found that participants trained on vowel harmony patterns were able to generalize to both novel stem vowels and novel suffix vowels.

In sum, these studies provide evidence that phonotactic constraints do not only affect specific segments, but also natural classes of segments. Patterns of segments that form natural classes are easier to learn than patterns that form arbitrary classes. In addition, based on exposure to patterns of specific

segments from a natural class, learners abstract over the feature-based similarity between these segments and induce constraints that go beyond the training segments, affecting the natural class as a whole. Similar findings have been reported in studies on phonotactic learning through speech *production*. There is evidence that adult learners can induce experiment-wide phonotactics on specific segments as a result of pronouncing experimental items (e.g., Dell, Reed, Adams, & Meyer, 2000; Goldrick & Larson, 2008). In addition to constraints on specific segments, there is evidence that phonological features affect phonotactic learning in speech production (Goldrick, 2004).

While the learning of novel phonotactics from isolated word forms has been studied extensively, the learning of phonotactics from continuous speech has been addressed only indirectly. Specifically, such studies focus on the role of segment probabilities in word segmentation (Newport & Aslin, 2004; Bonatti et al., 2005; Toro et al., 2008). Newport and Aslin (2004) show that adult learners are able to track consonant-to-consonant transitional probabilities as well as vowel-to-vowel transitional probabilities, in a continuous stream of artificial speech consisting of CV syllables. Since low probabilities are typically associated with word boundaries in segmentation (Saffran, Newport, & Aslin, 1996), the ability to track phonotactic probabilities in the speech stream provides learners with a cue for word segmentation. During the test phase, participants had to choose between ‘words’, consisting of high-probability sequences, and ‘part-words’, containing a low-probability sequence, thereby straddling a word boundary as defined by the statistical structure of the speech stream. Languages in which consonant probabilities were manipulated and languages in which vowel probabilities were manipulated were learnable, as shown by a significant preference for words over part-words. These results indicate that segment co-occurrence probabilities assist speech segmentation. Moreover, the study provides evidence that co-occurrence probabilities can be learned when segments are not immediately adjacent, but are separated by intervening vowels or consonants in the speech stream.

In a similar study, Bonatti et al. (2005) found a different pattern of results. When using artificial languages that were slightly more complex than those used by Newport and Aslin (in terms of the number of word frames, and immediate frame repetitions), learners were able to exploit consonant probabilities but not vowel probabilities for segmentation. They conclude that the statistical learning of vowel dependencies is more difficult than the statistical learning of consonant dependencies. They argue that this is due to a functional distinction between consonants and vowels (namely that consonants are used for word identification and vowels carry information about syntax, see Nespor, Peña, & Mehler, 2003). In addition to demonstrating the relevance of consonant probabilities for word segmentation, Bonatti et al. showed that

learners learn the phonotactic structure of the words. When trained on a continuous artificial language containing 9 words that are structured in such a way that they exhibit statistical regularities on both the consonantal and the vocalic tier, learners pick up the consonant structure of these words and generalize these structures to test items containing a novel vowel structure. The preference of consonant phonotactics over vowel phonotactics is reflected in the significant preference for words respecting the consonant structure of the training language (with new vowel structures) over words respecting the vowel structure of the training language (with new consonant structures).

There are two ways in which the results of the study by Bonatti et al. (2005) can be interpreted. The first interpretation is that learners induce phonotactics directly from the speech stream. That is, it is possible that participants in the experiment did not use consonant probabilities for *word* segmentation, but instead learned only consonant dependencies. The consequence would be that word *roots* were learned, rather than actual words (Bonatti et al., 2005). This could explain the generalization to novel vowels, since simply no vowel information was extracted from the speech stream. The second interpretation is that learners induce phonotactics from a statistically learned lexicon of word forms. That is, it may be the case that participants used consonant probabilities to align their segmentation of the speech stream, and subsequently learned the artificial lexicon (containing the actual multisyllabic words) from the speech stream. This statistically learned lexicon could then have been used to derive the phonotactic structure of these words, allowing participants to generalize to novel words that have the same consonant structure. Indeed, it has been argued that statistical learning provides a basis for an initial lexicon, from which further linguistic generalizations can be derived (e.g., Swingley, 2005; Thiessen & Saffran, 2003). Note that these two interpretations sharply contrast the Speech-Based Learning hypothesis and the Lexicon-Based Learning hypothesis (as formulated in Chapter 1). The latter predicts that the phonotactic structure of words is derived from the (proto-)lexicon, while the former predicts that the structure of words is learned from continuous speech directly.

The current chapter attempts to disentangle these two possible explanations, and aims at showing that phonotactics can be learned from continuous speech without mediation of a lexicon of word forms. In order to provide evidence in favor of the Speech-Based Learning hypothesis, artificial languages are constructed in such a way that it is unlikely that learners construct a lexicon from sequences in the speech stream. This is done by either substantially increasing the number of words that would have to be learned, or by inserting randomly selected vowels into the speech stream, resulting in the absence of recurring words in the language. In addition, this chapter looks at whether

participants learn phonotactic constraints on specific consonants, or whether they induce feature-based generalizations from the continuous speech stream.

#### *The current experiments*

Since most previous studies have used artificial languages that consisted of only a few words (typically about 9 or 12) with many occurrences in the speech stream, it is possible that learners build up a lexicon of CVCVCV word forms while listening, and project any generalizations (e.g., CXCXCX word roots, Bonatti et al., 2005; XAXBXA structural generalizations based on vowels, Toro et al., 2008; or AXC structural generalizations based on syllables, Peña et al., 2002) from this lexicon. Indeed, it has been argued that cues facilitating word segmentation allow for the extraction of structural generalizations from the speech stream (Peña et al., 2002; Toro et al., 2008). Here, we ask whether CXCXCX word roots can be learned under conditions in which the segmentation of whole words is particularly difficult. This is done by constructing languages that have many different CVCVCV word forms with only a few occurrences (Experiment 1), or languages in which vowels occur at random, thereby eliminating the notion of recurring word forms altogether (Experiment 2). If learners can still pick up word roots in these conditions, distinguishing between low-frequency (Experiment 1) or novel (Experiment 2) word forms in the test phase, then this would constitute strong evidence for the SBL hypothesis. It would indicate that, in accordance with the SBL hypothesis, the phonotactic structure of words is learnable without reference to actual word forms. That is, it would show that phonotactic constraints are learnable directly from continuous speech input.

The learning of feature-based generalizations from continuous speech has not been addressed in earlier studies. Therefore, in addition to testing for lexicon-based versus speech-based learning, the chapter will look into whether learners induce feature-based phonotactic constraints from the speech stream (Experiment 3). Crucially, such constraints should generalize to test items containing novel segments taken from the same natural class as the segments used in the training language. This would constitute even stronger evidence for the SBL hypothesis, since it would show that not only word roots, but also generalizations in terms of abstract levels of representation (natural classes) would be learnable from continuous speech. This would be in direct contradiction with studies that argue that learning mechanisms with different forms of computation (statistical learning, generalization) operate on different sorts of input (Peña et al., 2002; Toro et al., 2008). Specifically, it would contradict the view spelled out in these studies that statistical learning operates on continuous input, and that generalization operates on segmented input.

The goal of the first experiment is to assess whether learners can exploit phonotactic probabilities for segmentation in an artificial language that consists of a multitude of word forms, each with only a few occurrences. The structure of the language has two important properties. First, as a result of a relatively large number of consonant frames, combined with a large number of intervening vowels, the language entails a vocabulary that is far larger than those used in earlier studies (e.g., Newport & Aslin, 2004; Bonatti et al., 2005). The large number of words makes it unlikely that learners acquire the actual vocabulary of word forms during familiarization, and use those word forms for the induction of phonotactics (as would be predicted by the Lexicon-Based Learning hypothesis). While this possibility cannot be ruled out, it seems more likely that phonotactic knowledge arises from abstraction over the speech stream, by tracking the probabilities of consonants across vowels. Second, in order to allow for the possibility of feature-based abstractions from specific training segments to natural classes, the consonant sequences in the language display substantial phonological similarities. As a consequence of this setup, the difference between ‘within-word’ and ‘between-word’ consonant probabilities is smaller than in any earlier study on consonant probabilities ( $TP_{within} = 0.5, TP_{between} = 0.33$ ).<sup>1</sup> Before testing the types of generalizations that may be extracted from such a language, Experiment 1 is concerned with the question of whether the statistical structure of the language is learnable under these relatively difficult learning conditions.

## 5.2 EXPERIMENT 1

The first experiment focuses on two natural classes of consonants that co-occur across intervening vowels in a continuous speech stream. In the speech stream, fricatives (natural class A) are always followed by plosives (natural class B). A third class of unrelated segments (arbitrary class X) was included to create a 3-syllable pattern. The AB sequence was presented to participants as either a within-word or a between-word sequence. This was done by manipulating the statistical structure of the continuous speech stream. The stream was a continuous concatenation of either ABX or BXA words, resulting in higher probabilities for sequences within words than for sequences between words. The languages will be referred to as the ‘ABX language’ (with AB as a within-word sequence), and the ‘BXA language’ (with AB as a between-word

<sup>1</sup> In contrast to earlier chapters in this dissertation, statistical dependencies in this chapter are expressed in terms of transitional probability (TP) rather than in terms of observed/expected (O/E) ratio. This is done because TP has been the standard measure used in studies on artificial language learning. The results, however, are compatible with either measure (see e.g., Aslin et al., 1998). In addition, computational studies have shown that the two formulas produce roughly equivalent results (e.g., Swingley, 1999; see also Chapter 3 of this dissertation).

sequence). If participants learn the statistical structure of the language they are trained on, then they should show a preference for words that conform to the words from their training language (either ABX or BXA) as compared to words from the other language.

A critical aspect of artificial language learning experiments is whether the preference for a particular set of items can be genuinely attributed to exposure to the artificial language (see, Reber & Perruchet, 2003; Redington & Chater, 1996, for a discussion). Two factors are potentially problematic for such a conclusion. First, preferences may be due to the participant's native language, rather than to knowledge that was induced from the artificial language. Second, preferences may be due to the structure of test items, rather than to the structure of the training language. In order to rule out these potential confounds, the current experiment employs two different training languages and one single set of test items. That is, participants are exposed to one of the two training languages, and are subsequently tested on the same set of items. If participants base their decisions solely on knowledge from their native language, or on the structure of test items, then they would display the same preferences in the test phase, regardless of their training condition. Conversely, if there is a significant difference between the preferences of participants in the two groups, then this has to be due to the structure of the different training languages (since test items are identical in both conditions). In the current experiment, the two training languages contain the same continuous sequences, but the statistical structures of the two languages were manipulated such that they would indicate different segmentations. A significant difference between preferences in the two training conditions thus would indicate that participants are sensitive to the statistical structure of the training languages.

It should be noted that the current experiment does not provide a test for generalization. Specifically, it is not a test for the construction of feature-based generalizations, since all segments had occurred in the familiarization stream. Also, it is not a test for the abstraction of word roots, since test items had occurred in the familiarization language (albeit with rather low occurrence frequencies). Rather, the experiment is a first exploration intended to see whether participants can learn the statistical structure of continuous artificial languages under conditions that are more difficult than those in earlier studies (e.g., Bonatti et al., 2005; Newport & Aslin, 2004). Specifically, the current experiment employs languages that are more complex in terms of the total number of words in the language, and the relatively small differences in transitional probability in the continuous speech stream.

Table 5.1: Artificial languages for Experiment 1

ABX language		BXA language	
Consonant frames (C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub> -)	Vowel fillers (-V <sub>1</sub> -V <sub>2</sub> -V <sub>3</sub> )	Consonant frames (C <sub>2</sub> -C <sub>3</sub> -C <sub>1</sub> -)	Vowel fillers (-V <sub>2</sub> -V <sub>3</sub> -V <sub>1</sub> )
f.b.x-	[-a][-e][-o]	b.x.f-	[-e][-o][-a]
f.d.l.	[-i][-u][-y]	d.l.f.	[-u][-y][-i]
s.p.n-	[-ɛi][-œy][-au]	p.n.s-	[-œy][-au][-ɛi]
s.b.l.		b.l.s-	
z.p.x-		p.x.z-	
z.d.n.		d.n.z-	

Note. A = fricatives, B = plosives, X = 'arbitrary' class.

### 5.2.1 Method

#### Participants

Forty native speakers of Dutch (33 female, 7 male) were recruited from the Utrecht Institute of Linguistics OTS subject pool (mean age: 21.6, range: 18-34). Participants received 5 euros for participation. Participants were assigned randomly to either the ABX or the BXA condition.

#### Materials

Words are defined by 6 consonant frames (C.C.C.), which are combined exhaustively with 9 different Dutch vowels<sup>2</sup> (three for each vowel position). As a result, the language contains a total of 162 different CVCVCV words (6 consonant frames times 3 × 3 × 3 vowel frames). Two different languages (ABX and BXA) were created, which contain different orderings of three classes of consonants. Class A consists of fricatives /f/, /s/, and /z/. Class B consists of plosives /p/, /b/, and /d/. The arbitrary class X consists of /x/, /n/, and /l/. The consonant frames and vowel fillers for each language are given in Table 5.1.

The two languages are different segmentations derived from the same underlying consonant sequencing:

$$(5.1) \dots ABXABXABXABXABXABX\dots$$

<sup>2</sup> Only tense vowels were used, as lax vowels are restricted to occur only in word-medial position in Dutch. Diphthongs were included to expand the vowel set.



The ABX language is characterized by having AB as a word-internal sequence. In this language, each fricative can be followed by 2 different plosives. Similarly, each plosive can be followed by 2 different consonants from the arbitrary class. This results in a ‘within-word’ consonant transitional probability (TP) of 0.5. Every final consonant (from the X class) can be followed by each of the 3 initial consonants (from the A class). The ‘between-word’ consonant TP is thus 0.33. Each of 3 different vowels can occupy a vowel slot, resulting in vowel transitional probabilities of 0.33 (both within words and between words). The statistical structure of the language, with lower probabilities for XA sequences than for AB and BX sequences, predicts the following segmentation of the continuous speech stream:

(5.2) ... ABX.ABX.ABX.ABX.ABX.ABX... (*ABX language*)

The BXA language has the same basic structure, but predicts a segmentation that is shifted one element to the right. In this case, each plosive can be followed by 2 different consonants from the arbitrary class, and each arbitrary consonant can be followed by 2 different fricatives. Crucially, since fricatives occur in the word-final syllable, they are followed by 3 different word-initial plosives. In this language, AB sequences are thus predicted to be interpreted as between-word sequences, due to lower transitional probabilities of between-word sequences:

(5.3) ... A.BXA.BXA.BXA.BXA.BXA.BX... (*BXA language*)

Note that the vowels are still linked to the same consonants (see Table 5.1). The only difference between the ABX and BXA languages is the location of the statistical word boundaries in the speech stream (due to lower probabilities between words than within words).

For each language, a continuous sequence of CV syllables was generated by concatenating 4 different pseudo-random orderings of the 162 different words in the language. Each of the 162 words thus occurred 4 times in the speech stream, and, since the stream was presented twice to participants, each word occurred a total of 8 times during familiarization. In order to control for syllable TPs, restrictions were defined on which syllables could follow each other. Specifically, since syllables could have 6 possible successors within words (since each syllable has 2 successor consonants combined with 3 different vowels), the allowed syllable sequences *between* words in the language were restricted to match this number. This ensured that participants could not rely on between-word versus within-word syllable TPs during segmentation. The exact statistical structures of the languages are as follows: Consonant TPs are 0.5 within words, and 0.33 between words (range: 0.28-0.4, ABX language; 0.3-0.36, BXA language). Vowel TPs are 0.33 within words, and also

between words (range: 0.25-0.39, ABX language; 0.30-0.38, BXA language). Finally, syllable TPs are 0.17 within words and between words (range: 0.08-0.24, ABX language; 0.11-0.24, BXA language). The resulting continuous language contains 1944 CV syllables. An audio stream was generated using the MBROLA text-to-speech system (Dutoit, Pagel, Pierret, Bataille, & Vrecken, 1996), using the Dutch 'nl2' voice. The stream was synthesized with flat intonation and had a total duration of 7.5 minutes (average syllable duration: 232 ms).

For the test phase, 36 items were selected from each language. The selection was such that each test item consisted of a consonant frame with a mixed set of vowels (i.e., one long vowel: /a/, /e/, /o/, one short vowel: /i/, /u/, /y/, and one diphthong: /ei/, /œy/, /au/), occurring at different positions. A list of test trials was created for use in a 2-alternative forced-choice task. Every trial consisted of one ABX word and one BXA word (e.g., /fibexau/ - /bœylosi/). The complete list of test items is given in Appendix D. The test items were synthesized with the same settings as the familiarization stream.

#### *Procedure*

Participants were tested in a sound-attenuated booth. Participants were given written instructions which were explained to them by the experimenter. Audio was presented over a pair of headphones. Participants' responses were given by selecting one out of two response options (indicated visually with '1' and '2') by clicking with a mouse on the screen. The instructions given to participants were that they would hear a novel ('Martian') language, and that their task was to discover the words of this language. Before starting the actual experiment, participants were given a short pre-test in which they had to indicate whether a particular syllable had occurred in first or in second position in a trial. This was done to familiarize participants with the setup of the experiment.

The 7.5-minute familiarization stream was presented twice, with a 2-minute silence between presentations of the stream. As a consequence, total familiarization time was 15 minutes. The stream started with a 5-second fade-in, and ended with a 5-second fade-out. There were thus no indications of word endings or word beginnings in the speech stream. After familiarization, participants were given a two-alternative forced-choice (2AFC) task in which they had to indicate for each trial which out of two words sounded more like the Martian language they had just heard. After the experiment had finished, participants were asked to fill in a debriefing questionnaire, which was intended to check whether there had been any specific items or potential L1 interferences (such as embedded words) systematically affecting participants' preferences.

The questionnaire asked participants to indicate the following:

- (i) which words from the Martian language they could remember (and were asked to make a guess if they could not remember the exact word),
- (ii) whether they had noticed any existing Dutch words while listening, and, if so, which ones,
- (iii) whether they had employed specific strategies in making their decisions in the test phase, or whether they had simply been guessing,
- (iv) what their general impression of the experiment was, and whether they had been clearly instructed,
- (v) whether they had participated in similar experiments before.

### 5.2.2 Results and discussion

The mean preference of each individual participant is plotted in Figure 5.1. A mixed logit model (see e.g., Jaeger, 2008) with subjects and items as random factors showed that there was no significant effect of training condition ( $p = 0.4484$ ).<sup>3</sup> These results thereby provide no evidence that participants learned the statistical structure of the continuous language to which they had been exposed. An additional analysis tested whether the conditions, when analyzed separately, are significantly different from chance-level performance. This analysis indicates that participants in both conditions have a significant preference for BXA words over ABX words (ABX condition:  $p < 0.01$ , BXA condition:  $p < 0.001$ ). There thus appears to be a bias in favor of BXA over ABX words, which was driven by factors independent of training. There is no evidence that participants learned the statistical structure of the language and used the structure to learn words from the speech stream.

The experiment, which failed to show an effect of statistical learning, illustrates the necessity for control groups in artificial language learning (see also, Reber & Perruchet, 2003; Redington & Chater, 1996). On the basis of the results for the BXA language alone, one could be inclined to conclude that participants indeed learned the statistical structure of the continuous speech stream, and segmented the continuous stream accordingly into word-like units. Presenting participants with a language with the opposite statistical structure (the ABX pattern), however, did not result in a different segmentation of the speech stream. Regardless of the statistical structure of the familiarization stream, participants had a general preference for BXA over ABX words.

Interestingly, many participants had quite consistent responses, as shown by their near-categorical preferences (see Figure 5.1). That is, only a few

<sup>3</sup> Many studies use  $t$ -tests (on participants' aggregated preferences) for the analysis of 2AFC data. See Jaeger (2008) for an extensive discussion of why  $t$ -tests should not be used for the analysis of categorical outcomes.

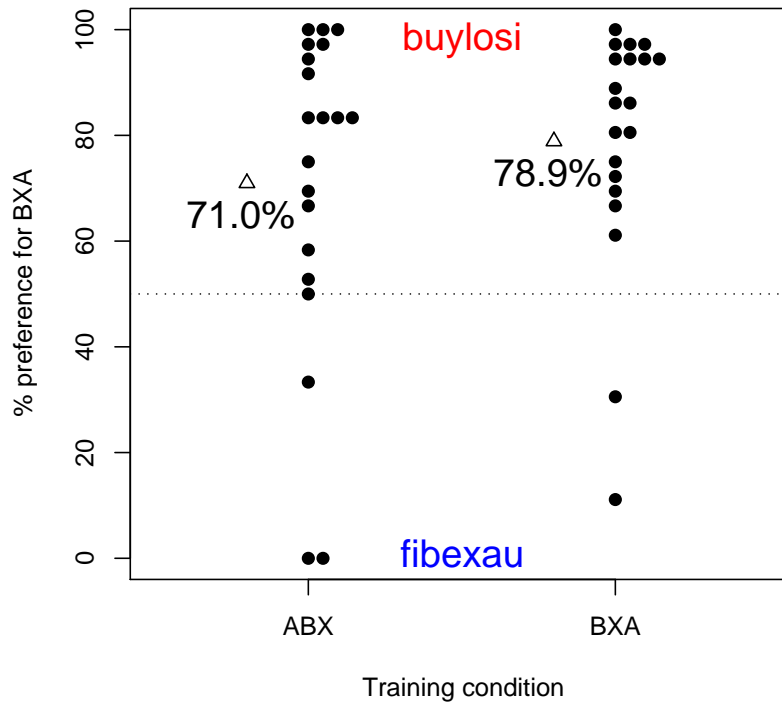


Figure 5.1: Results for the ABX and BXA training languages. The circles indicate mean preferences for individual participants. Triangles indicate the mean score for each condition.

participants had preferences that were close to random (i.e., approaching 50%). This could be interpreted to mean that participants picked up a repeating pattern, and consolidated this pattern through listening (rather than not being affected by the training language at all). This is also illustrated by two participants in the ABX condition who displayed a systematic preference for ABX words, choosing ABX over BXA in all trials. It is thus possible that the bias in favor of one of the two word types was active during segmentation in the training phase, rather than during the judgment of experimental items in the test phase.

The question then arises what caused the large majority of participants to segment the speech stream according to the BXA pattern. This was examined in two different ways. First, the debriefing questionnaires were checked for any recurring items, patterns, or strategies that participants had been aware of, and could have influenced their decisions. Second, the ABX and BXA items were analyzed in terms of various phonotactic probabilities in Dutch, the native language of the participants. Several studies have shown that L1 phonotactics can interfere with artificial language segmentation (Boll-Avetisyan & Kager, 2008; Finn & Hudson Kam, 2008; Onnis et al., 2005). Corpus analyses were conducted to see whether there were L1 segmentation cues that could have interfered with participants' segmentation of the artificial language. These post-hoc analyses of the results were not meant to provide conclusive evidence about what strategy participants had used in the experiment, but rather were meant to highlight potentially interfering factors, which could then be controlled in a follow-up experiment.

The questionnaires showed no indication of specific items or embedded words systematically affecting the results. Participants in both conditions mainly had remembered words that conformed to the BXA pattern. Out of a total 124 words that were written down as remembered items, about 74% conformed to the BXA pattern, and about 22% conformed to the ABX pattern. This roughly reflects the results shown in Figure 5.1. Several existing Dutch words had been heard in the familiarization phase. These were mainly monosyllabic words, such as *zij* ('they'), *doe* ('do'), *nu* ('now'). Interestingly, whenever participants had heard a sequence of multiple Dutch syllables, these were usually ABX sequences (*zij doen nu*, 'they do now', *zeg dan niet*, 'say then not'). Apparently, these sequences had not interfered with the segmentation of the speech stream. Few participants reported Dutch BXA sequences. When this did occur, these were low-frequent words, such as *delusie* ('delusion') or *doelloosheid* ('purposelessness'). Thus, the questionnaires showed no indication that participants' preference for BXA words was due to embedded Dutch words.

Several participants (22 out of 40) had written down a strategy that they had used during the test phase. Although participants' learning of the artificial words may have been implicit, and participants thus may not have been aware of the strategies they had actually employed, inspection of these strategies could reveal properties of the artificial languages that may have been particularly salient for the participants. The reported strategies varied from very general remarks, such as 'I paid attention to word endings' to specific reports on the patterns of the words, such as 'many words started with /p/, had medial /x/, and ended with /s/'. Some participants reported that they had based their decisions on the final vowel of the word. Specifically, they

reported that they had chosen words ending with /i/. Inspection of the words that participants had listed as remembered items confirms this: 44% of these words ended with /i/. Inspection of the BXA test items also indicates that BXA items ending with /i/ overall (i.e., averaged over different participants) received somewhat higher scores than BXA words ending with /a/ or /ɛi/ (ABX condition: /i/ - 73.8%, /a/ - 70.4%, /ɛi/ - 68.8%; BXA condition: /i/ - 84.2%, /a/ - 75.4%, /ɛi/ - 77.1%), suggesting that participants may have used final-/i/ as a segmentation cue.<sup>4</sup>

In order to assess whether word-final vowels indeed could have provided a segmentation cue, the vowels used in the experiment were compared on the basis of their frequency in word-final position in Dutch words. If there is a segmentation strategy based on word-final vowels, then this should be reflected in higher word-final frequencies for those vowels that occur in final positions in BXA sequences, than vowels at the ends of ABX or XAB sequences. The distribution of word-final vowels across Dutch words was calculated using the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). The counts indicate that final vowels in BXA words indeed have higher frequencies than final vowels in ABX or XAB words (BXA: /i/ - 3050, /a/ - 796, /ɛi/ - 616; ABX: /o/ - 598, /y/ - 45, /au/ - 16; XAB: /e/ - 331, /u/ - 63, /œy/ - 34). These counts make it seem plausible that participants may have been sensitive to the position of /i/ in the speech stream, since there are many more Dutch words ending with /i/ than with any of the other vowels. Note that the claim that word-final vowels affect segmentation has not been proven here. In order to so, an experiment should be designed specifically aimed at such a hypothesis. Rather, the data reported here indicate that it may be wise to be cautious when using /i/ in artificial segmentation experiments with Dutch participants. In general, artificial languages should be controlled for interferences stemming from the native language as much as possible. The current analyses indicate that possible effects of word-final vowels should be considered for the list of factors that need to be controlled for.

One way to dispose of any potential interferences from vowels is to make sure that vowels do not appear at fixed positions in the consonant pattern. That is, vowels can be disconnected from consonant frames by simply inserting them at random into vowel slots in the speech stream. If vowels are inserted at random, then they cannot be used by listeners to align their segmentations into a fixed, repeating pattern. Thus, random vowel insertions render vowels completely useless as a segmentation cue. The insertion of vowels at random positions has the additional advantage that it provides an even stronger test

<sup>4</sup> The 'final-/i/' strategy does not necessarily result in higher scores for /i/-final words. That is, if /i/ is used by participants to align their segmentation into the BXA pattern, then this does not imply that words ending with /i/ are better words than other BXA words, but rather means that /i/-alignment is used to prefer BXA segmentations over ABX segmentations.

for the Speech-Based Learning hypothesis. If vowels are randomly inserted into the speech stream, then there are simply no systematically recurring words (i.e., CVCVCV sequences) in the language. This makes it highly unlikely that learners induce the phonotactic structure of words from a statistically learned lexicon of CVCVCV word forms. Rather, it seems likely that learners would induce phonotactics directly from the speech stream.

An alternative explanation for the current findings is that acoustic properties of the speech stream, rather than L1 knowledge, may have pushed participants into systematic BXA segmentation. For example, the articulation of plosives requires closing the vocal tract in order to build up pressure before the release. This causes a short silence in the acoustic signal prior to the burst of air. Such fine-grained acoustic properties of the speech signal may potentially provide a cue for segmentation. However, no evidence for a systematic 'plosive-initial' strategy was found in the debriefing questionnaires. One way to rule out this possibility would be to create a control condition in which the familiarization stream consists of randomly ordered syllables. Such a control condition would take away any systematic phonetic bias in the speech stream, while still allowing for L1 transfer to items in the test phase.

Instead of running such a control condition, it was decided to construct new languages in Experiment 2, distributing plosives and fricatives equally across consonant positions in the speech stream. Phonetic differences between plosives and fricatives could therefore not lead to a systematic segmentation bias. It should nevertheless be noted that natural classes are defined by articulatory features and thus typically result in specific acoustic regularities (such as pre-burst silences) in the speech stream. The challenge is to construct artificial languages in such a way that they contain only a minimal amount of acoustic cues to boundaries. In Experiment 2, two new languages were constructed in which consonant frames were interleaved with random vowels.

### 5.3 EXPERIMENT 2

Experiment 2 is again concerned with the induction of CXCXCX word roots from continuous speech. In order to control for possible interference of vowel structures, vowel fillers are selected at random. Consonant frames consist of segments from three natural classes: voiceless obstruents (class A), voiced obstruents (class B), and dorsal obstruents (both voiceless and voiced, class C). These classes are used to create two languages which are defined by different statistical structures: an ABC language and a BCA language. Test items in this experiment are novel combinations of consonant frames and vowel fillers, which had not occurred in the speech stream. The experiment thus tests whether the knowledge that is acquired by participants generalizes

to novel words that have the same consonantal structure as the sequences in the familiarization language.

### 5.3.1 *Method*

#### *Participants*

Forty native speakers of Dutch (33 female, 7 male) were recruited from the Utrecht Institute of Linguistics OTS subject pool (mean age: 21.4, range: 18-39). Participants received 5 euros for participation. Participants were assigned randomly to either the ABC or the BCA condition.

#### *Materials*

Six new C\_C\_C\_ consonant frames are used to create the languages. Consonants were taken from natural classes of obstruents. Class A consists of voiceless obstruents /p/, /t/, /s/. Class B consists of the voiced counterparts of those obstruents: /b/, /d/, /z/. The third class, class C, has three dorsal obstruents /k/, /g/, and /x/. Note that two segments were intentionally withheld from the A and B classes. Specifically, voiceless obstruent /f/ was not presented as member of the A class, and voiced obstruent /v/ was not presented as member of the B class. These segments were kept for future testing of feature-based generalizations from the specific consonants in the class to novel segments in Experiment 3 (Section 5.4). The consonants are again sequenced in two different ways. One language is made up of ABC consonant frames, having higher probabilities for AB and BC ('within-frame') sequences than for CA ('between-frame') sequences. The second language consists of BCA consonant frames, resulting in higher probabilities in the speech stream for BC and CA ('within-frame') sequences than for AB ('between-frame') sequences. The materials are given in Table 5.2.

Continuous sequences of consonants were generated by concatenating 600 randomly selected frames for each language. The resulting speech stream had  $600 \times 3 = 1800$  consonants. The two languages differed with respect to their consonant co-occurrence probabilities (in the same way as in Experiment 1): Within-frame probabilities were 0.5 (consonants had 2 possible successors within frames), while between-frame probabilities were 0.33 (consonants had 3 possible successors between frames). The hypothesized segmentation of the two languages is thus as follows:

(5.4) ... ABC.ABC.ABC.ABC.ABC.ABC... (*ABC language*)

(5.5) ... A.BCA.BCA.BCA.BCA.BCA.BC... (*BCA language*)



Table 5.2: Artificial languages for Experiment 2

ABC language		BCA language	
Consonant frames (C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub> -)	Vowel fillers ( <i>random</i> )	Consonant frames (C <sub>2</sub> -C <sub>3</sub> -C <sub>1</sub> -)	Vowel fillers ( <i>random</i> )
p_d.g_	[_a,_e,_o,_i,_u,_y]	d_g.p_	[_a,_e,_o,_i,_u,_y]
p_z.k_		z_k.p_	
t_b.x_		b_x.t_	
t_z.g_		z_g.t_	
s_b.k_		b_k.s_	
s_d.x_		d_x.s_	

*Note.* A = voiceless obstruents, B = voiced obstruents, C = dorsal obstruents. Vowel fillers are inserted at random into the speech stream.

where ‘.’ indicates a boundary as predicted by a low transitional probability between consonants.

A set of six vowels (/a/,/e/,/o/,/i/,/u/,/y/) was used to fill the vowel slots in the continuous sequence of consonant frames. For each vowel slot, a filler was selected from the set of vowels at random. This procedure resulted in the following statistical structure for the languages: Consonant TPs were 0.5 within words (range: 0.48-0.52, ABX language; 0.48-0.52, BXA language), and 0.33 between words (range: 0.31-0.35, ABX language; 0.32-0.35, BXA language). Vowel TPs were 0.17 within words, and also between words (range: 0.14-0.2, ABX language; 0.14-0.2, BXA language). Note that there is a small difference between within-word and between-word syllable transitional probabilities in the setup used in this experiment. While any syllable can have 12 possible successors (2 consonants times 6 vowels) within words, there are 18 possible successors (3 consonants times 6 vowels) between words. Theoretically, syllable TPs are thus 0.08 within words, and 0.06 between words. However, this theoretical difference is not reflected in the actual speech stream. Due to the random insertion of vowels, there are many different syllable bigrams, each with very low occurrence frequencies. Each language contains about 675 different syllable bigrams. Since the total speech stream contains 1799 bigram tokens, each syllable bigram on average occurs 2.7 times. In contrast, the set of consonant bigrams is much smaller, and such bigrams have higher occurrence frequencies. There are 21 different consonant bigrams, and consonant bigrams on average occur 85.7 times each. The consequence is that consonant TPs

can be estimated much more reliably from the continuous speech stream than syllable TPs. The estimated values for consonant bigrams approach the theoretical values (as can be seen from the TP ranges listed above), whereas syllable TPs take on a wide range of values: 0.02-0.29 within words, and 0.02-0.21 between words (in both languages). Importantly, the fact that within-word and between-word TP ranges are largely overlapping, indicates that syllable TPs are an unreliable segmentation cue in this experiment. That is, within-word syllable probabilities are not systematically higher than between-word syllable probabilities.

Due to the design of the language, there are no systematically recurring word forms in the speech stream, and phonotactic learning is hypothesized to operate on the speech stream, rather than on a statistically learned lexicon. However, CVCVCV word forms might occasionally recur in the speech stream due to chance. In order to see how often this had occurred, the continuous ABC and BCA languages were analyzed for recurring CVCVCV sequences. Frequency counts indicate that there were 485 different ABC word forms in the speech stream (with each 'word' being a CVCVCV combination of one of the consonant frames and three randomly selected vowels). The average frequency of occurrence was 1.2. Similarly, the BCA language contained 482 different BCA forms, with an average frequency of 1.2. Most forms occurred only once, and a small number of forms occurred up to a maximum of 4 times in the speech stream. In contrast, the original work on word segmentation by Saffran, Newport, and Aslin (1996) used 6 different word forms which each occurred 300 times in the speech stream. It thus seems unlikely that participants in the current experiment would learn this large set of different word forms. Rather, it seems plausible to assume that any phonotactic knowledge was due to the learning of consonant structures directly from the speech stream.

Audio streams were generated for the two languages using MBROLA (Dutoit et al., 1996), using the Dutch 'nl2' voice. The streams were synthesized with flat intonation and had total durations of 7 minutes (with average syllable durations of 232 ms). For the test phase 36 novel three-syllabic sequences were created for each language. The test items were made by combining consonant frames with vowel structures in such a way that there were no vowel repetitions in the item (i.e., each word had 3 different vowels), and that the exact combination of consonants and vowels had not occurred in either of the two familiarization languages. Care was taken to ensure that no other factor than the consonant probabilities in the artificial language could affect decisions in the test trials. Importantly, any remaining possible effect of vowel structure in the test items was neutralized by employing pairs of test trials in which the consonant frame for each item was the same in both trials, but

the vowel frame was switched between items.<sup>5</sup> That is, for each trial (e.g., /tibaxo/ – /dugopa/) a counterpart was created in which vowel frames had been switched (e.g., /tuboxa/ – /digapo/). If participants would base their decisions on vowel frames of the test items, then chance-level performance should occur, since vowel frames were distributed evenly between ABC and BCA items. Any significant learning effect is thus due to the consonant structure of the words. Vowel frames for the test items were chosen in such a way that there was a minimal difference in cohort density (i.e., the number of Dutch words starting with the initial CV syllable) between two items in a trial. For example, /ti-/ and /du-/ have comparable cohort densities (as do /tu-/ and /di-/). Test items were synthesized with the same settings as those for the familiarization stream. The complete set of test trials is given in Appendix D.

#### *Procedure*

The procedure was the same as in Experiment 1. Each stream lasted 7 minutes, and was presented twice with a 2-min pause in between. Total familiarization time was 14 minutes.

#### 5.3.2 *Results and discussion*

The mean preference of each individual participant is plotted in Figure 5.2. A mixed logit model with subjects and items as random factors revealed a significant effect of training condition ( $p = 0.0168$ ). The significant difference between groups indicates that participants' preferences were affected by the statistical structure of the continuous language to which they had been exposed. Importantly, these results show that learners have the ability to induce phonotactics directly from the speech stream. Since participants in both conditions were tested on the same items, the different results cannot be reduced to effects from the native language, nor to learning during the test phase. An additional analysis tested whether the separate conditions were different from chance-level performance. Participants in the ABC condition showed no significant preference for either ABC or BCA items ( $p = 0.114$ ). In contrast, participants in the BCA condition showed a significant preference for BCA words over ABC words ( $p < 0.001$ ).

The results of the current experiment can be interpreted in different ways. It is possible that there was a general bias in favor of BCA words. If this is true, then participants in the ABC condition neutralized this bias as a result of exposure to the opposite statistical pattern (and leading to chance-level performance during the test phase). An alternative explanation is that there

---

<sup>5</sup> I thank Tom Lentz for this suggestion.

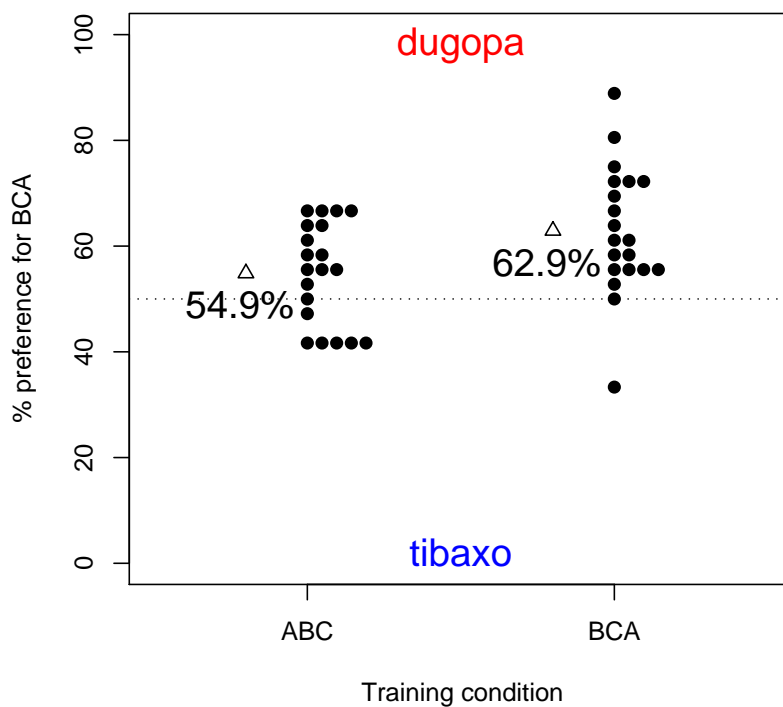


Figure 5.2: Results for the ABC and BCA training languages. The circles indicate mean preferences for individual participants. Triangles indicate the mean score for each condition.

was no bias for either type of structure. In that case, participants in the ABC condition simply did not pick up the pattern from their training language, whereas participants in the BCA condition did. Either way, the significant difference between training conditions shows that participants were sensitive to the statistical structure of the familiarization language. The results show that participants were able to induce CXCXCX consonant roots from the continuous artificial speech stream, generalizing to novel words conforming to those roots in the test phase.

The results indicate that learners' ability to learn consonant structures from continuous speech is stronger than previously shown. Compared to earlier studies (Bonatti et al., 2005; Newport & Aslin, 2004), the languages used in this chapter displayed smaller differences in consonant TPs (the current study:  $TP_{within} = 0.5$ ,  $TP_{between} = 0.33$ ; earlier studies:  $TP_{within} = 1.0$ ,  $TP_{between} = 0.5$ ). In addition, the languages in the current experiment contained randomly inserted vowels. The lack of systematically recurring CVCVCV word forms in the speech stream makes it unlikely that participants relied on a statistically learned lexicon to learn the word roots. Rather, it seems plausible that learners picked up phonotactics (consonant structures) directly from the continuous speech stream. This finding supports the Speech-Based Learning hypothesis, which states that phonotactics is learned from continuous speech, rather than from the lexicon.

Furthermore, the test words did not occur in the familiarization stream. Despite the difficult learning conditions, participants were able to learn the phonotactic structure of the training language from continuous speech, and generalize these structures to previously unheard test words. While the learning effect is not particularly large, it should be noted that the design allows for the conclusion that the learning effect is truly due to the statistical dependency of consonants in the artificial languages, and not to effects from the native language. The learning effect can thus be attributed to the induction of novel phonotactics from continuous speech.

The languages were made up of segments from three different natural classes, voiceless obstruents, voiced obstruents, and dorsal obstruents (both voiced and voiceless). However, not all members of these natural classes were presented to the learners. Specifically, the labial fricatives /f/ and /v/ did not occur in the familiarization languages. The model of phonotactic learning that has been proposed in this dissertation, STAGE (see Chapter 2), assumes that the learner constructs feature-based generalizations on the basis of statistically learned phonotactic constraints. In this view, constraints on specific consonants that form a subset of a natural class result in the construction of a generalization which affects the natural class as a whole. That is, constraints affecting classes A (voiceless obstruents) and B (voiced obstruents) should generalize to novel segments from the same natural class (/f/ and /v/, respectively) which were not heard during familiarization. Experiment 3 tests whether participants perform feature-based generalization to novel segments from the same natural class.

## 5.4 EXPERIMENT 3

Experiment 3 asks whether participants induce feature-based generalizations (phonotactic constraints targeting natural classes as a whole) from continuous speech. The continuous stream of the BCA language from Experiment 2 was used as familiarization language. A random control language serves to check that any effect found would not be due to L1 preferences. The control language consists of random combinations of consonants and vowels.<sup>6</sup> Instead of testing whether participants pick up specific CXCXCX word roots, participants are tested on whether they learn the natural class structure of word roots. This is done by presenting them with test items which contain novel word roots. Specifically, each test item contains one novel consonant that was not heard during familiarization.

The task of participants is to choose between novel items that conform to the natural class pattern of the training language, and novel items that violate the natural class pattern of the training language. If participants construct feature-based generalizations on the basis of occurrences of specific consonants in the training language, then participants should prefer test items in which a consonant has been replaced by a novel segment from the same natural class over test items in which a consonant has been replaced by a novel segment from a different natural class. In contrast, no such preference should be found for participants that are trained on the random control language. A significant difference between the training conditions would indicate that participants learn feature-based generalizations from the continuous speech stream. Importantly, since neither the legal nor the illegal novel segment occurred in the familiarization languages, a difference between conditions would indicate that participants relied on a phonotactic constraint at the level of natural classes, and not on a constraint affecting specific consonants in the familiarization language.

5.4.1 *Method**Participants*

Forty native speakers of Dutch (34 female, 6 male) were recruited from the Utrecht Institute of Linguistics OTS subject pool (mean age: 22.3, range: 18-28). Participants received 5 euros for participation. Participants were assigned randomly to either the BCA or the Random condition.

<sup>6</sup> The ABC language from Experiment 2 was not used in the current experiment, since the language produced no significant preference for either type of word structure. It was thus not expected that training participants on this language would lead to accurate feature-based generalizations.

### *Materials*

The familiarization stream was identical to the BCA condition in Experiment 2. The stream consists of a continuous sequence of consonant frames where voiced obstruents (class B) are followed by dorsal obstruents (class C) and voiceless obstruents (class A). Vowel slots were filled with randomly chosen vowels. (See Table 5.2 for the specific materials used in the language.) A control condition (Random) was created in which all consonants and all vowels occurred at random. The continuous stream of the Random condition was of the same length as the BCA stream (1800 CV syllables). The same consonants and vowels were used as in the BCA language, but consonant slots and vowel slots were filled randomly with segments from the inventory.

All test items consisted of BCA frames in which either the initial consonant (from class B) or the final consonant (from class A) was replaced with a novel consonant. The novel consonant was either from the same natural class as the replaced segment (i.e., legal substitution), or from a different natural class (i.e., illegal substitution). The novel consonants were /f/ (voiceless obstruent, belonging to class A) and /v/ (voiced obstruent, belonging to class B). In the substitutions of initial consonants, the legal item would be vCA (e.g., /vygate/, conforming to the BCA pattern), and the illegal item would be fCA (e.g., \*/fykape/, violating the BCA pattern). Conversely, in word-final position the legal item would be BCf (e.g., /dagefo/, conforming to the BCA pattern), and the illegal item would be BCv (e.g., \*/zakevo/, violating the BCA pattern). A preference for legal over illegal substitutions would indicate that participants had learned the BCA pattern.

Trials were constructed such that both the legal and the illegal item had the same vowel structure, but different consonant structures (e.g., /vygate/ – /fykape/, /dagefo/ – /zakevo/). Any effect found would thus be due to the consonant structure of the test items, and not to vowel structure. Trials were constructed in such a way that the items had comparable cohort densities. Participants received trials with substitutions in initial position, and trials with substitutions in final position. The order of test trials was randomized. The order of presentation of legal and illegal items within trials was balanced within and across participants. The complete set of test trials is given in Appendix D. All materials were synthesized using MBROLA (Dutoit et al., 1996), using the Dutch ‘nl2’ voice.

### *Procedure*

The procedure was the same as in Experiments 1 and 2. Total familiarization time was 14 minutes for both conditions (BCA and Random).

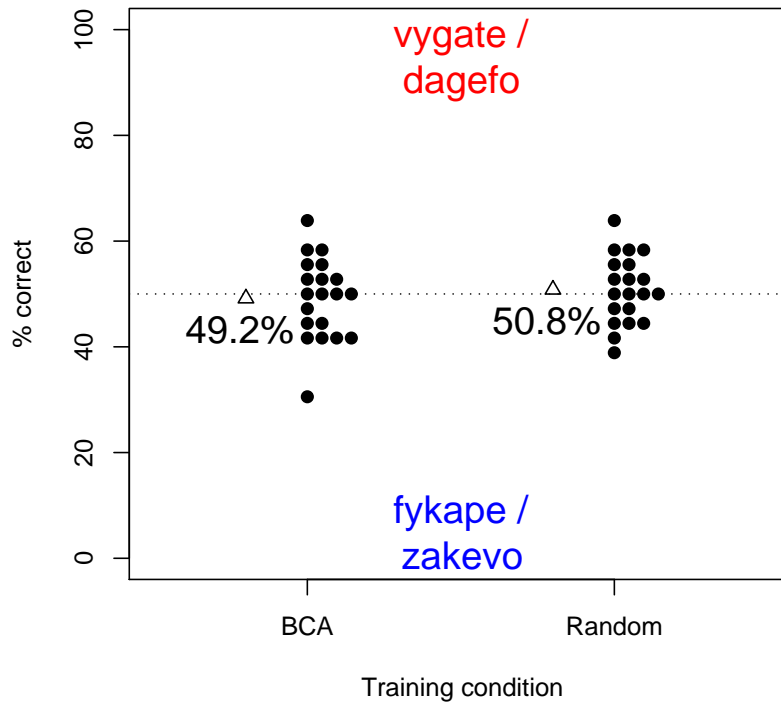


Figure 5.3: Results for the BCA and Random training languages. ‘% correct’ indicates how often participants chose legal items over illegal items. The circles indicate mean preferences for individual participants. Triangles indicate the mean score for each condition.

#### 5.4.2 Results and discussion

The percentages of correct responses (i.e., choosing legal over illegal items) for each participant are shown in Figure 5.3. A mixed logit model with subjects and items as random factors indicates that there was no significant effect of training condition, no significant effect of substitution position (initial or final), and no significant interaction between training condition and position. When analyzed separately, neither condition was significantly different from



chance. The experiment thus provides no evidence for the learning of feature-based generalization. That is, participants did not generalize constraints on consonants to which they had been exposed during familiarization to novel consonants from the same natural class. In addition, there appears not to have been any L1 preference for either items containing /f/ or items containing /v/, as indicated by chance-level performance in both the BCA and the control condition. It should be noted that this null result does not show that participants *cannot* induce feature-based generalizations, but rather that there is no evidence supporting the view that they can.

Chance-level performance in both conditions could be taken to indicate that participants had difficulty perceiving the difference between /f/ and /v/. If participants do not hear the difference between these two sounds, then they cannot distinguish between legal and illegal items in the test phase. There are two sources that could have caused perceptual confusion. First, the contrast between /f/ and /v/ does not surface reliably in spoken Dutch. Specifically, words with underlying /v/ are often pronounced as /f/. The strength of this effect varies between different regions in the Netherlands. Second, perceptual confusion could have been caused by the fact that participants heard synthesized speech, rather than natural speech. The use of synthesized speech inevitably comes with loss of phonetic detail and thus may contribute to the difficulty in distinguishing between voiced and unvoiced fricatives.

A second possible explanation of the null result is the complexity of the artificial languages used here, as compared to earlier studies (e.g., Bonatti et al., 2005; Newport & Aslin, 2004). As described in Experiment 2 (Section 5.3), the BCA language contains relatively small differences between within-word and between-word consonant transitional probabilities, and contained randomly inserted vowels. These factors may have made it difficult for participants to induce a robust feature-based generalization after 14 minutes of exposure.

Finally, one would have to consider the possibility that the hypothesis is wrong. It may be the case that phonotactic constraints on natural classes of consonants are not learnable from continuous speech input. It has been argued that different forms of computation operate on different sorts of input (Peña et al., 2002). Specifically, Peña et al. (2002) argue that statistical learning operates on continuous speech, while generalization mechanisms operate on word forms. In their study, it was found that generalizations were learnable only after the insertion of acoustic segmentation cues (brief silences) into the speech stream. It should be noted, however, that the study by Peña et al. (2002) focused on structural generalizations (e.g., AXC syllable patterns, where A predicts C across intervening syllables), and not on feature-based generalizations (where specific elements from class A would generalize to other, unheard elements from that same class). Further research is thus

needed to show whether or not feature-based generalizations are learnable from continuous speech.

Possible follow-up experiments could focus on reducing the difficulty of learning the specific consonant structures, either through the use of less complex languages, or through increasing the exposure time. In addition, it would be worth investigating whether a different set of consonant classes could be used for familiarization and testing. Specifically, the current design, which focuses on distinguishing between consonants that are either legal or illegal at a certain position, requires that the novel segments can be accurately perceived. The challenge is thus to find a pattern of natural classes which is not likely to be perceptually confused, and at the same time is not heavily biased in terms of L1 phonotactics.

In addition, in order to support the claim by Peña et al. (2002) that generalization operates on segmented word forms, rather than on continuous speech, one could consider conducting an experiment in which the BCA frames from the current experiment would be separated by brief silences in the speech stream. However, the conclusions that could be drawn from such an experiment are still limited. If BCA patterns are learnable from segmented speech, then this could also be taken to show that a reduction in the complexity of the learning task (i.e., facilitating segmentation through the insertion of acoustic cues into the speech stream) enables the learning of feature-based generalizations. A stronger case for learning generalizations from word forms could be made if learning from continuous speech still fails after substantially increased exposure time.

## 5.5 GENERAL DISCUSSION

Human learners, both adults and infants, have been shown to be capable of learning novel phonotactic regularities from a set of isolated word forms (e.g., Onishi et al., 2002; Saffran & Thiessen, 2003). The current chapter focused on whether adult learners can induce novel phonotactic constraints when presented with a stream of continuous speech from an artificial language. In line with the Speech-Based Learning hypothesis, it was predicted that phonotactic constraints can be learned directly from the continuous speech stream, without reference to a statistically learned lexicon of word forms. This prediction was tested by creating artificial languages in which the formation of a lexicon would be rather difficult. Compared to earlier studies (e.g., Bonatti et al., 2005; Newport & Aslin, 2004), these languages showed increased complexity both in terms of their consonantal structure and in terms of their vocalic structure. Importantly, the combination of a relatively large number of consonant frames and a large set of intervening vowels resulted in languages

which either had a large vocabulary of word forms (162 words, Experiment 1), or had no recurring word forms at all (due to the random insertion of vowels, Experiment 2). This setup makes it likely that participants would induce phonotactics directly from the speech stream, rather than from a statistically learned lexicon.

No evidence for phonotactic learning was found in Experiment 1. This was found to be due to a general bias stemming from the participants' native language. Experiment 2 showed that, when carefully controlling for such L1 factors, participants were able to learn phonotactics from continuous speech. The experimentally-induced phonotactic knowledge of consonant frames generalized to novel word forms in the test phase.

The findings of Experiment 2 shed light on the types of input that are used by human learners as a source for the induction of phonotactics. As outlined in Chapter 1, the Speech-Based Learning (SBL) hypothesis states that continuous speech is the input for phonotactic learning. Using the mechanisms of statistical learning and generalization, learners detect phonotactic regularities in the input. Phonotactics is subsequently used for the detection of word boundaries in the speech stream, and thus contributes to the development of the mental lexicon. In contrast, the Lexicon-Based Learning (LBL) hypothesis assumes that phonotactic constraints are the result of abstractions over statistical patterns in the lexicon. The LBL hypothesis thus states that phonotactics is not learned directly from the continuous speech stream.

In Experiment 2 it was found that adult learners can pick up phonotactic patterns directly from the speech stream. The results show that learners need not be presented with word forms to learn phonotactics. The experiment thus provides support for the SBL hypothesis. The findings are also in line with Bonatti et al. (2005) who suggest that learners learn the consonant structure of words from continuous speech, and do not necessarily extract a sequence of syllables to form a lexicon. It should be noted that the experiment does not contradict earlier studies that have demonstrated phonotactic learning from isolated word forms. Rather, it is a demonstration that word forms are not *required* for the learning of phonotactics. This is an important finding, since infant language learners are confronted with continuous speech input, and are faced with the task of developing a mental lexicon. The capacity to induce phonotactics from continuous speech could serve as a way of bootstrapping into lexical acquisition, since phonotactics can be used to hypothesize the locations of word boundaries in the speech stream. The finding thus provides support for the psychological plausibility of the computational learning approach that was proposed earlier in this dissertation. In Chapters 2 and 3 it was shown that phonotactic constraints can be learned from a corpus of transcribed continuous speech, and that such constraints provide a useful cue

for the segmentation of continuous speech. Here, it was shown that human learners can indeed induce phonotactics from continuous speech input.

In Experiment 3 the possible role of feature-based generalizations in phonotactic learning was investigated. By training participants on a subset of the consonants from a natural class, and testing them on novel consonants which were not presented during familiarization, the experiment examined whether the induced phonotactic knowledge generalized to novel segments from the same natural class. The experiment produced a null result, and thus provides no evidence that learners induce feature-based generalizations from continuous speech input. Possible explanations for the failure of the experiment include perceptual confusion of the novel consonants, and the structure of the artificial language, which was possibly too complex to result in a feature-based generalization during familiarization. Further research on this issue is needed to determine which types of phonotactic constraints can be learned from different sorts of input. With respect to the computational model presented in Chapter 2 it remains to be demonstrated that participants perform feature-based abstractions on the basis of statistically learned bigram constraints.

Several previous studies have argued that consonants and vowels are processed differently in phonotactic learning (Bonatti et al., 2005; Chambers et al., 2010). The results of Experiment 2 are compatible with this view. Consonant probabilities were learned across randomly occurring intervening vowels. That is, the occurrence of vowels between consonants did not hinder the learning of the dependencies between consonants. This provides support for an assumption that was made in Chapter 4. In Chapter 4, the computational model, STAGE, was applied to consonant bigrams that occurred with intervening vowels in the corpus. These intervening vowels were removed from CVC sequences before applying the model. This modification to the processing of input by the model was necessary to allow for the model to capture restrictions non-adjacent consonants. The findings of the current chapter provide psycholinguistic evidence that human learners can indeed process consonant sequences independently from intervening vowels. The modification that was made to the processing window of the model (allowing it to focus on non-adjacent consonants) thus seems to be justified by the current findings.

In Experiment 1, a bias was found for one type of segmentation over the other, independent of the training condition. Possible explanations of these biases were found in phonotactic properties of the native language, which may have affected the segmentation of the artificial language. Several previous studies have shown that properties of the native language can influence the segmentation of artificial languages (Boll-Avetisyan & Kager, 2008; Finn & Hudson Kam, 2008; Onnis et al., 2005). The fact that no artificial language is likely to be completely neutral with respect to the native language of the

participants calls for the use of proper control groups (Reber & Perruchet, 2003; Redington & Chater, 1996). That is, a learning effect cannot be reliably established by comparison of a single experimental condition to chance-level performance. Control groups should be employed before concluding that any significant result is due to the structure imposed by the artificial language. In the current chapter, effects due to interference from the native language were checked through the use of languages that have opposite statistical structures (such as in Experiments 1 and 2), or languages that have no structure at all (such as the random language in Experiment 3). As a consequence, it was possible to disentangle general biases from learning effects.

A final issue concerns possible formalizations of what exactly is learned by participants in artificial language learning studies such as those presented here. In Chapter 2 an explicit model for the induction of phonotactics from continuous speech was proposed. It should be noted that the data in this chapter support the learning of phonotactics from continuous speech, but the learning of feature-based generalizations from continuous speech remains to be demonstrated. While it is difficult to set up an experiment that provides conclusive evidence for a computational model, it is nevertheless interesting to speculate how learners could come up with general preferences for certain phonotactic patterns. What does STAGE predict with respect to the induction of phonotactics from the continuous artificial languages in the current chapter? Below, a sketch will be provided of how the model can be modified to induce phonotactic constraints from the artificial languages used in this chapter. The sketch should, however, not be taken as evidence for the model, nor as an account of the present findings. While some of the findings in this chapter support the ideas behind STAGE, further experimentation is needed to test detailed predictions that follow from the model.

STAGE categorizes bigrams into markedness constraints ( $*xy$ ) and contiguity constraints (CONTIG-IO( $xy$ )). This categorization relies on O/E thresholds that separate high-probability sequences from low-probability sequences. The markedness threshold ( $t_M$ ) and continuity threshold ( $t_C$ ) can be set to categorize the consonant sequences in the artificial languages into sets of specific markedness and contiguity constraints. This requires calculating the O/E values for the consonant sequences in the language. In the current set of artificial languages, the 'within-word' transitional probability of 0.5 corresponds to  $O/E = 4.5$ , and the 'between-word' transitional probability of 0.33 corresponds to  $O/E = 3.0$ . It should be noted that, due to the relatively small number of consonant bigrams in the artificial language, both the between-word and within-word O/E values are statistically overrepresented (that is,  $O/E > 1.0$ ). Capturing the distinction between 'high' and 'low' probabilities thus requires that we set both thresholds relatively high (e.g.,  $t_M = 3.5$ ,  $t_C = 4.0$ ).

The result of this threshold configuration is that the model induces  $*xy$  constraints for sequences between frames, and  $\text{CONTIG-IO}(xy)$  for sequences within frames. Importantly, with this configuration STAGE induces different constraints in different training conditions. For example, in the ABC language from Experiment 2, the model induces  $\text{CONTIG-IO}(AB)$  constraints which prevent the insertion of word boundaries into AB sequences (e.g.,  $\text{CONTIG-IO}(pVz)$ ,  $\text{CONTIG-IO}(tVb)$ ). In contrast, the statistical structure of the BCA speech stream is such that the model would induce  $*AB$  constraints (e.g.,  $*pVz$ ,  $*tVb$ ), which favor the insertion of boundaries into these sequences.

The phonological structure of the word frames used in the languages is such that STAGE's generalization mechanism, Single-Feature Abstraction, would construct a more general constraint based on these specific constraints. For example, in the BCA condition in Experiment 2 the specific constraints against AB sequences (e.g.,  $*pVz$ ,  $*tVd$ , etc.) would trigger the generalization  $*x \in \{p,t,f,s\}; y \in \{b,d,v,z\}$ . Conversely, in the ABC condition the model predicts the induction of  $\text{CONTIG-IO}(x \in \{p,t,f,s\}; y \in \{b,d,v,z\})$ . The model thus predicts that the novel segments /f/ and /v/, which did not occur in the ABC and BCA training languages, are affected by this generalization.

An interesting prediction that follows from the model is the strength of feature-based generalizations. STAGE typically assigns lower ranking values to generalizations than to specific constraints. Indeed, when applying the model to the artificial languages in the way described above, the generalizations affecting the A and B classes as a whole end up at the bottom of the constraint hierarchy. While the model predicts that novel consonants are affected by natural class constraints (as a result of feature-based generalizations over specific consonants), it may be that these constraints are too weak (as reflected by the low ranking value) to produce any observable effects in experiments such as the ones reported in this chapter.

To conclude, the current chapter provides support for the Speech-Based Learning hypothesis, since it shows that human learners are able to induce novel phonotactics from continuous speech input. With respect to the precise learning mechanisms that are involved, and the types of input these mechanisms operate on, many issues require further research. Specifically, it remains an open issue whether feature-based generalizations are learnable from continuous speech input.

# 6

## SUMMARY, DISCUSSION, AND CONCLUSIONS

---

When hearing speech, listeners are confronted with an acoustic signal which contains no clear indications of word boundaries. In order to understand spoken language, listeners thus face the challenge of breaking up the speech stream into words. The process of speech segmentation is aided by several types of linguistic cues which indicate word boundaries in the speech stream. Metrical cues, fine-grained acoustic cues, and phonotactic cues all contribute to the recognition of spoken words. These cues may be particularly important for language learners. Infants, who have not yet acquired the vocabulary of the native language, hear continuous speech in their environment, and need to develop strategies to extract word-like units from speech. Infants exploit metrical, acoustic, and phonotactic cues for the development of a lexicon from continuous speech input. The question arises how infants acquire these cues (which are language-specific) in the absence of a lexicon.

The focus of this dissertation has been on learning phonotactic cues for segmentation. Phonotactic constraints affect segmentation, since they define which sound sequences cannot occur within the words of a language (e.g., \*pf in Dutch). The occurrence of such sequences in the speech stream therefore indicates the presence of a word boundary within the sequence. It was predicted that phonotactic constraints can be learned from continuous speech, using the mechanisms of statistical learning and generalization. Moreover, it was predicted that these constraints would subsequently be used by the learner for the detection of word boundaries in continuous speech. This hypothesis, which was referred to as the Speech-Based Learning hypothesis, was investigated using a combination of (i) computational modeling, providing a formal account of how phonotactics can be learned, (ii) computer simulations, assessing the relevance of phonotactic constraints for speech segmentation, (iii) simulations of human data, assessing the ability of the computer model to account for segmentation behavior, and (iv) artificial language learning experiments, testing the capacity of human learners (adults) to induce novel phonotactics from a continuous speech stream. Through the use of computational modeling and psycholinguistic experiments, this dissertation has aimed at providing a formal account of the induction of phonotactics for speech segmentation, which is supported by experimental data from both computational and human learners.

A particular focus of this dissertation has been on bridging the gap between models of speech segmentation and models of phonotactic learning. While both types of models address the learning of co-occurrence patterns, they make radically different assumptions about the input of phonotactic learning, and the level of abstractness of phonotactic constraints. Specifically, segmentation models have assumed that phonotactics is learned *from continuous speech*, while phonotactic learning models have assumed that phonotactic constraints are learned *from the lexicon*. In addition, segmentation models have assumed that phonotactic constraints refer to *specific segments*, whereas phonotactic learning models have assumed that phonotactic constraints target patterns of *natural classes of segments*, as defined by abstract phonological features. Evidence from infant studies indicates that phonotactic constraints on natural classes are learned before the infant's lexicon has reached a substantial size. This led to the specific hypothesis that abstract phonotactic constraints are learned from continuous speech during early stages of phonological development. The approach taken in the dissertation thus combines the input assumption of segmentation models with the representational assumption from phonotactic learning models. Below, a summary is given of the dissertation (Section 6.1). The chapter then proceeds to critically assess the findings, and to provide suggestions for future work in this area (Section 6.2). Finally, some concluding remarks are made (Section 6.3).

## 6.1 SUMMARY OF THE DISSERTATION

### 6.1.1 *A combined perspective*

In light of the above, this dissertation has addressed the issue of how language learners induce phonotactic constraints that will help them to break up the continuous speech stream into words. Phonotactic constraints define which sound sequences are allowed, and which sound sequences are disallowed within the words of a language. Occurrences of sequences in the continuous speech stream that are not allowed within words have the potential to indicate the location of a word boundary to the learner. While the relevance of phonotactics has been demonstrated for speech segmentation by adults (e.g., McQueen, 1998) and infants (e.g., Mattys & Jusczyk, 2001b), little is known about the acquisition of these constraints. The acquisition of phonotactics by infants poses a particularly interesting problem, since infants are only beginning to develop a mental lexicon, and thus seem to be able to infer the sound structure of words, without knowing the words themselves.

Through an investigation of the induction of phonotactic constraints for speech segmentation, the dissertation has aimed at providing insight into



several related problems, which are typically studied in separate research areas. First and foremost, the dissertation takes a computational perspective. This entails that the research has been mainly concerned with specifying the exact computations that are performed by learners when confronted with input data. This algorithmic view results in an explicit proposal for the case of phonotactic learning, and for the application of phonotactic constraints to the speech segmentation problem.

Second, the dissertation takes a linguistic perspective. That is, it was attempted to provide a linguistic characterization of the knowledge that is learned by infants, and could be used by infants to break up the speech stream. The linguistic characterization critically involves phonological features, which are used by the learner to construct more abstract phonotactic constraints. The dissertation addresses the roles of both segment-based and feature-based constraints in segmentation. The output of the computational learning model is represented as a set of phonological constraints, which are interpreted using the linguistic framework of Optimality Theory (Prince & Smolensky, 1993). The linguistic interpretation of the output allows for an exact description of the knowledge that is constructed as a result of learning, and it allows for a specification of the types of phonotactic knowledge that are relevant for segmentation. Importantly, the approach is data-driven: It focuses on knowledge that is learned on the basis of input data, and, in contrast to the traditional view of Optimality Theory, it does not make the assumption of an innate constraint set.

Third, the dissertation takes a psychological perspective. The model that has been proposed is based on learning mechanisms that have been shown to be available to human language learners. Moreover, the model is presented with transcribed utterances of continuous speech, and thus does not rely on a previously acquired lexicon of word forms for the induction of phonotactics. The lexicon-free approach to phonotactic learning is compatible with psycholinguistic studies showing that infants possess knowledge of the phonotactics of their native language at the age when they start acquiring a lexicon. In addition to psycholinguistic motivations behind the learning model, the dissertation provides new data on human phonotactic learning mechanisms. In a series of artificial language learning experiments with human participants, it was examined which types of phonotactic knowledge can be learned from a stream of continuous speech input. In sum, this dissertation has aimed to contribute to our understanding of human phonotactic learning and speech segmentation. An explicit proposal was formulated which subsequently was evaluated in different types of empirical studies, involving both computer simulations and experiments with human participants.

The dissertation's theoretical focus is on the *input* that is used for phonotactic learning, and the *mechanisms* that are used in phonotactic learning. The input issue concerns the question of whether the learner is presented with a lexicon of word forms (either types or tokens) for phonotactic learning, or whether the learner is presented with utterances that contain a continuous stream of segments (with no apparent word boundaries). The issue of learning mechanisms deals with how the learner derives phonotactic knowledge from the input. Various sources of evidence point towards roles for two mechanisms: statistical learning and generalization. The former specifies how learners accumulate data about occurrences of specific structures in the input. The latter is concerned with learners' ability to derive more abstract linguistic knowledge, such as rules and constraints, from the data. This dissertation aims at specifying the nature of each mechanism individually, as well as explaining how the mechanisms interact. Specifically, the question is addressed how statistical learning can provide a basis upon which the learner can build phonotactic generalizations.

#### 6.1.2 *The learning path*

Phonotactic learning, as discussed in this dissertation, can follow two different paths. The traditional assumption in studies on phonotactic learning has been that phonotactic constraints are induced from the lexicon (e.g., Frisch et al., 2004; Hayes & Wilson, 2008; Pierrehumbert, 2003). I have called this the Lexicon-Based Learning (LBL) hypothesis. From an acquisition point of view, the LBL hypothesis may be problematic. Infants have notably small lexicons (consisting of less than 30 words by the age of 8 months, Fenson et al., 1994). Nevertheless, there is evidence that infants have fine-grained probabilistic knowledge of the phonotactics of their native language (e.g., Jusczyk, Friederici, et al., 1993; Jusczyk et al., 1994) by the age of 9 months. It seems unlikely that infants induce their initial knowledge of phonotactics from such a small lexicon. Also, there is the question *how* infants learn their first words from the continuous speech stream. Several studies have shown that infants use phonotactics for speech segmentation (Mattys et al., 1999; Mattys & Jusczyk, 2001b). Knowledge of phonotactics may thus be used by infants to bootstrap into word learning (e.g., Cairns et al., 1997). If this is the case, then infants use phonotactics for the development of the mental lexicon, rather than relying on the lexicon for the learning of phonotactics.

This dissertation has investigated the consequences of these considerations, and addressed what I have called the Speech-Based Learning (SBL) hypothesis. In this view, phonotactic constraints are learned from continuous speech input, and are subsequently used for the detection of word boundaries in the speech

stream. Phonotactics thus contributes to the development of the mental lexicon. The dissertation presents a computational model which implements the SBL hypothesis. Three different types of empirical studies in the dissertation are concerned with providing both computational and psycholinguistic support for the SBL hypothesis.

### 6.1.3 *The proposal and its evaluation*

Chapter 2 presents the computational model, STAGE (Statistical learning and Generalization), which is an implementation of the SBL hypothesis. Specifically, the chapter shows that a model based on learning mechanisms that have been demonstrated to be available to human language learners allows for the induction of phonotactic constraints from continuous speech. The model induces biphone constraints through the statistical analysis of segment co-occurrences in continuous speech (Frequency-Driven Constraint Induction), and generalizes over phonologically similar biphone constraints to create more general, natural-class-based constraints (Single-Feature Abstraction). The model learns phonotactic constraints of two different types: markedness constraints ( $*xy$ ) and contiguity constraints (CONTIG-IO( $xy$ )). The former type exerts pressure towards the insertion of boundaries into the speech stream, while the latter type militates against the insertion of boundaries. Conflicts between constraints are resolved using the OT segmentation model.

In Chapter 3, a series of computer simulations is presented which demonstrate the potential usefulness of the SBL hypothesis, as implemented by STAGE. The chapter focuses on the contributions of different learning mechanisms to segmentation by asking whether a crucial property of the model, feature-based generalization, improves the segmentation performance of the learner. Experiments 1–3 zoom in on the complementary roles of statistical learning and generalization in phonotactic learning and speech segmentation. Experiment 4 shows the development of the model’s segmentation performance as a function of input quantity. The main finding of the chapter is that STAGE outperforms purely statistical models, indicating a potential role for feature-based phonotactic generalizations in speech segmentation. The chapter thus provides support for the SBL hypothesis, since it demonstrates that phonotactic learning contributes to better segmentation, and thus facilitates the development of the mental lexicon.

Chapter 4 examines whether STAGE can provide a learnability account of a phonological constraint, OCP-PLACE (restricting the co-occurrence of consonants sharing place of articulation), and of its effect on segmentation. The chapter examines the output of the constraint induction procedure, applied to non-adjacent consonants in the speech stream, to see to what extent the

induced constraint set reflects OCP-PLACE. In addition, the model is used to predict word boundaries in an artificial language from a study by Boll-Avetisyan and Kager (2008). The segmentation output is evaluated on its goodness of fit to human item preferences in the artificial language learning study. The chapter provides a detailed examination of the contents of the constraint set that is induced by the model, and of the ability of this constraint set to account for human segmentation behavior. Experiment 1 shows that STAGE (which uses feature-based generalization) has a better fit to human data than models that rely on pure consonant distributions. In addition, the model outperforms a categorical interpretation of OCP-PLACE. The success of the approach was found to be due to the mix of specific and abstract constraints. Experiment 2 addresses the input issue. It was found that a continuous speech-based learner has a fit to the human data comparable to a type-based learner, and a better fit than a token-based learner. The chapter shows that STAGE successfully learns constraints that resemble OCP-PLACE through abstractions over statistical patterns found in continuous speech. The model shows tendencies towards similarity avoidance, without employing a metric to assess the similarity between consonants in a sequence, and without relying on a lexicon of word forms for constraint induction. In addition, the simulations show that both specific and abstract phonotactic constraints are needed to account for human segmentation behavior.

Chapter 5 investigates whether human learners can learn novel phonotactic structures from a continuous speech stream. In a series of artificial language learning experiments, the chapter examines the psychological plausibility of the phonotactic learning approach. In addition, the chapter provides a test for the assumption that learners can induce phonotactic constraints on non-adjacent consonants from continuous speech (while ignoring intervening vowels). Experiment 1 shows that a bias (probably due to interference from L1 phonotactics) influences the segmentation of the artificial language by adult participants, and hinders the learning of the statistical structure of the artificial language. Experiment 2 shows that, when carefully controlling for potential biases, human learners can induce novel phonotactics from continuous speech. The experiment demonstrates that learners generalize phonotactics, learned from continuous speech input during familiarization, to novel items in the test phase. The results were obtained with languages that had only subtle differences in 'within-word' and 'between-word' probabilities, and the experiment thus provides new evidence for the strong capacity of human learners to induce consonant dependencies using a statistical learning mechanism. However, in Experiment 3, no evidence was found for feature-based generalization to novel consonants (i.e., consonants not presented during

training). This null result was possibly due to perceptual confusion, or due to the complexity of the training language.

## 6.2 DISCUSSION AND FUTURE WORK

This dissertation has proposed a theoretical model, implemented as a computer program, and has presented a total of 9 empirical studies that assess the proposal. These studies use different methodologies: computer simulations, computer simulations linked to human data (taken from earlier studies), and phonotactic learning experiments with human participants (i.e., new human data). Four experiments assess the ability of the model to detect word boundaries in a corpus of unsegmented speech in a series of computer simulations (Chapter 3). Two experiments are concerned with linking the output of phonotactic learning simulations to results from artificial language segmentation by human learners (Chapter 4). Finally, three experiments present new data on the learning of artificial language phonotactics by human participants (Chapter 5). Before summing up the evidence for the SBL hypothesis, I will compare the proposed model to other models of constraint induction. While such models have provided learnability accounts of adult phonotactic knowledge, rather than infant phonotactic knowledge, and have been concerned with wellformedness judgments, rather than with segmentation, the models have some interesting similarities and differences. Below, some of these issues will be discussed. Suggestions will be provided for future work which might lead to a more direct comparison between these models and STAGE.

### 6.2.1 *Comparison to other constraint induction models*

Here, I will discuss STAGE in comparison to two earlier models of phonotactic constraint induction: the maximum-entropy learner of Hayes and Wilson (2008), and the feature-based model by Albright (2009). While STAGE was primarily designed as a model for the induction of phonotactic constraints for speech segmentation, and the models of Hayes and Wilson (2008) and Albright (2009) were meant to account for human wellformedness data, the models are similar in the sense that they *induce* abstract, feature-based phonotactic constraints. The models thus do not rely on a universal constraint set, but rather construct phonotactic constraints as a response to input data. It is therefore worthwhile to consider the differences and similarities between these models of constraint induction.

One basic assumption that is common to all three models is that the learner has access to the feature specifications of the segments in the language's inventory, and that these specifications are used to learn phonotactic generalizations.

In addition, the three models assume similar constraint formats. Hayes and Wilson (2008) define markedness constraints of the form  $*[a][b][c] \dots$  where  $[a]$ ,  $[b]$ , and  $[c]$  are feature bundles. The scope of the constraints (i.e., the number of bundles in a sequence) can be varied, but large values (say, larger than 3) should be avoided, since those would result in an unmanageable search space. Both the model of Albright (2009) and STAGE have a biphone scope (although they are not in principle limited to this scope). All three models assume constraints that are sequential in nature. No word edges are encoded in the format of the constraints. However, the models can be made to produce positional constraints, simply by training them on positional data (e.g., onset sequences).

There are differences in how phonotactic generalizations come into existence. Hayes and Wilson (2008) assume a space of possible phonotactic constraints, from which constraints are selected and incorporated into the language-specific phonotactic grammar. The constraint space contains all logically possible constraints, given the feature specifications and the defined constraint scope. Using maximum-entropy learning, constraints are selected from the space that are maximally accurate with respect to the training data, and maximally general. In contrast, Albright's model is based on feature-based generalization over observed sequences in the training data (rather than from a pre-defined constraint space). The model decomposes words into biphone units, and assigns a wellformedness score to the word based on probabilities over combinations of natural classes (rather than over combinations of specific segments). STAGE constructs feature-based generalizations on the basis of statistical patterns in the input. The model relies on a statistical learning component which induces specific markedness and contiguity constraints from the data. More general (feature-based) constraints are derived from the specific constraints by abstracting over single-feature differences between specific constraints. The mechanism used by STAGE is similar to the model of Albright (2009) in the sense that generalizations are made only *after* the learner is presented with input data. In contrast, the model by Hayes and Wilson (2008) creates an *a priori* set of possible constraints, and selects constraints from this set in response to input data.

Albright (2009) shows that, when tested against English onset cluster data (taken from Scholes, 1966), his model outperforms the model of Hayes and Wilson in accounting for gradient judgments of sequences that are well-attested in the language. The two models had comparable performance in accounting for unattested sequences. Importantly for the current dissertation, Albright (2009) reports independent effects of segment-based and feature-based probabilities. He suggests that these effects may be due to two different levels of evaluation. While the simulations in this dissertation address a different

problem, namely the segmentation of continuous speech, similar results were found. That is, feature-based generalizations were found to improve the segmentation performance of the learner as compared to segment-based biphone probabilities (Chapter 3). Moreover, when the statistical induction thresholds of STAGE were set such that all biphones participated in generalization, the added value of generalization was lost. The findings in this dissertation are thus compatible with the view of Albright (2009) that segments and features may have independent contributions in a model of phonotactics. This independence is a crucial aspect of the architecture of STAGE: STAGE uses segment-based statistical learning, and adds feature-based generalization to statistically induced biphone constraints. In contrast to Albright's model, the effects that follow from specific biphone probabilities and natural class-based generalizations arise from a single set of constraints with varying levels of generality. Decisions on whether or not to insert a word boundary into the speech stream are based on evaluation of this single constraint set, and thus no distinct levels of evaluation are assumed.

A point for future research concerns the evaluation of the induced phonotactic constraint set. Throughout this dissertation I have tested the induced phonotactics for its contribution to solving the speech segmentation problem. This approach is motivated by the fact that phonotactics is used in speech processing and segmentation. Traditionally, however, phonologists have tested theories of phonotactics against wellformedness data (e.g., Scholes, 1966; Bailey & Hahn, 2001; Albright, 2009; Hayes & Wilson, 2008). A comparison of STAGE against models like those of Albright (2009) and Hayes and Wilson (2008) would require generating predictions about the wellformedness of nonwords from the model. A first step towards modeling wellformedness was taken in Chapter 4, where STAGE was used to predict human judgments on experimental items in an artificial language learning task. I speculate that two properties of the model would be useful for generating wellformedness scores for isolated word forms. The first property concerns the three distinct phonotactic categories that are predicted by STAGE. STAGE categorizes phonotactic sequences into categories of low, high, and neutral probability. A gradient distinction between nonwords could perhaps be derived from the model using this categorization. For example, when trained on Dutch onset clusters in the Spoken Dutch Corpus, the model induces markedness constraints for sequences that are illformed (e.g., \*sr), contiguity constraints for sequences that are wellformed (e.g., CONTIG-IO(st)), and no constraints for sequences that have a questionable status (e.g., /sn/, which is legal, but relatively infrequent).

A second property of the model that could be further exploited concerns the ranking values of the constraints. Gradient predictions within the phono-

tactic categories could perhaps be derived by comparing respective rankings of constraints (see Coetzee, 2004). For example, when STAGE is trained on consonant clusters in Dutch continuous speech, the model predicts \*nr ≫ \*mr. Alternatively, the ranking values could be used as weights, rather than as strict rankings (see e.g., Hayes & Wilson, 2008; Legendre et al., 1990; Pater, 2009). The gradient prediction that would follow from these interpretations is that Dutch participants judge *mrok* to be a better nonword than *nrok*.<sup>1</sup> These proposals for deriving gradient predictions from STAGE are preliminary, and would need to be worked out (and empirically validated). Conversely, it would be interesting to see whether earlier models of phonotactic learning can be modified to provide a plausible account for human segmentation data. Since phonotactics affects both the rating of nonwords, and the segmentation of continuous speech, a complete model of phonotactic learning should be able to account for both effects.

#### 6.2.2 *What is the input for phonotactic learning?*

A primary concern of the dissertation has been the input issue: Is phonotactics learned from continuous speech or from the lexicon? All empirical studies in the dissertation bear on this issue. It was argued in Chapter 1 that infants would need to learn phonotactics from continuous speech in order to rely on phonotactics to detect word boundaries in continuous speech. In order to evaluate this proposal a theoretical model was outlined, from which testable predictions were derived. Importantly, in order for the learning approach to be successful, there should be a mechanism that allows learners to pick up information from the continuous speech stream. Moreover, the continuous speech stream should be rich enough for information to be derived from it using these mechanisms. I would like to stress the importance of showing an exact mechanism that can be used by learners to actually pick up the information that is hidden in the input. That is to say, linguistic information, present in the data, is rather useless if human learners do not possess the mechanisms that will allow them to acquire this information from the data.

The computational model proposed in Chapter 2 shows that an approach to phonotactic learning from continuous speech can be defined and implemented. The model is based on learning mechanisms that are available to infants, and allow them to pick up information from the speech stream. The example provided in Chapter 2 indicates that, using these mechanisms, the learner can learn phonotactic regularities, as well as specific exceptions to these

<sup>1</sup> A similar gradient prediction can be made with respect to segmentation. The model predicts that Dutch listeners would be faster at detecting the embedded Dutch word *rok* in *finrok* than in *finrok*. (See McQueen, 1998, for a study of categorical effects of these sequences on segmentation.)



regularities, from the continuous speech stream (see Figure 2.5). The structure of continuous speech input thus seems to be rich enough to allow for the induction of phonotactics using the mechanisms of statistical learning and generalization. The empirical studies in Chapters 3–5 further explore this issue.

The computer simulations in Chapter 3 show that the constraints that are learned from continuous speech are useful for segmentation. In Chapter 4, it was shown that the structure of consonant sequences across intervening vowels in the speech stream also appears to be rich enough to allow for the induction of OCP-PLACE, a constraint against the co-occurrence of consonants sharing place of articulation. That is, when applying the learning model to such consonant sequences (while ignoring intervening vowels), the resulting constraint set contains phonotactic generalizations that resemble OCP-PLACE. In addition, the continuous speech-based learner has comparable performance to a learner based on word forms (types) in accounting for human segmentation data. For the induction of OCP-PLACE (using the mechanisms of statistical learning and generalization), there thus appears to be no advantage of having access to a lexicon as an input source. Finally, Experiment 2 in Chapter 5 provides direct evidence that human learners can learn phonotactics from continuous speech input. While the results in Chapter 4 can be explained by both a speech-based and a lexicon-based learner, the results of the artificial language learning experiment in Chapter 5 can only be due to learning from continuous speech, since there were simply no recurring word forms in the speech stream.

The claim that phonotactics can be learned from continuous speech is thus supported by a learning model, as well as by three different types of empirical studies involving both computational and human learners. The implication of these findings is that studies on phonotactic learning should consider continuous speech as a possible source of phonotactic knowledge. Previous studies have accounted for phonotactic knowledge either by abstraction over patterns in the lexicon (Hayes & Wilson, 2008; Pierrehumbert, 2003) or by innate universal preferences (e.g., Berent, Steriade, Lennertz, & Vaknin, 2007). Specifically, some studies have concluded that the absence of a particular sequence in the lexicon implies that language learners never encounter such a sequence in their native language (e.g., Berent et al., 2007). However, sequences that are illegal within words are still likely to occur across word boundaries. As a consequence, virtually all sequences occur in the speech streams that learners hear. The findings in this dissertation show that occurrences of such sequences may provide a valuable source of data for the learning of phonotactic constraints.

With respect to the input issue, future studies could focus on pushing phonotactic learning from continuous speech to its limits. Specifically, studies could focus on the types of phonotactic generalizations that can and cannot be learned from continuous speech input. Importantly for the current proposal, it remains to be demonstrated that human learners can learn feature-based generalizations directly from the speech stream. A related issue is whether the phonological features that are used by models of phonotactic learning can be learned from continuous speech (e.g., Lin & Mielke, 2008).

An open issue is when and how the lexicon comes into play in phonotactic learning. Assuming that phonotactic acquisition starts with learning from the continuous speech stream, it is possible that the lexicon at some point during development takes over, and becomes the most important source of phonotactic knowledge. That is, while phonotactics contributes to the development of the mental lexicon, the lexicon might provide an important source of phonotactic knowledge as soon as it has developed to a substantial size. The question is which types of constraints would become available as a result of learning from the lexicon. For example, constraints that explicitly link sequences to word edges might only become available after the learner starts learning from lexical items. Much work is needed to explain the development of phonotactic knowledge from infancy to adulthood. Explaining phonotactic development requires experiments documenting the types of phonotactic constraints that learners acquire at different ages, and computer models which can explain how different constraints can be learned at different stages of language acquisition.

### 6.2.3 *Which mechanisms are involved in phonotactic learning and segmentation?*

An important issue in this dissertation concerns the nature of the mechanisms by which language learners acquire phonotactics. Psycholinguistic studies show that at least two mechanisms are available to human learners: statistical learning and generalization. An important question is what the contribution of each of these mechanisms is to language acquisition in general, and to phonotactic learning in particular. While it has been widely acknowledged that both mechanisms are active during acquisition (e.g., Gómez & Gerken, 1999; Marcus et al., 1999; Toro et al., 2008; White et al., 2008), very little is known about how these mechanisms jointly contribute to the acquisition process, and how these mechanisms interact. The computational model presented in Chapter 2 provides a proposal for the interaction of the two mechanisms. The main claim is that statistical learning provides a basis for further linguistic generalizations. The learner collects data about co-occurrences, and generalizes to more abstract representations. This view

was explored throughout Chapters 3–5 by asking what the added value is of phonotactic generalizations that are based on statistical co-occurrences of specific segments in the speech stream. That is, the main question is whether feature-based generalization contributes to better phonotactic learning, and, as a consequence, to better speech segmentation.

The computer simulations in Chapter 3 show that adding generalization to statistical learning improves segmentation performance. This indicates a previously unexplored potential role for the generalization mechanism in speech segmentation. The simulations in Chapter 4 show that adding generalization to statistical learning results in phonotactic knowledge which closely resembles abstract constraints that have been proposed in theoretical phonology (such as OCP-PLACE). The generalization mechanism captures phonological regularities that are missed (or incompletely captured) by pure statistical learners. In addition, adding generalization to statistical learning improves the model's fit to human segmentation data, thereby strengthening the claim that there is a role for the generalization mechanism in speech segmentation. The findings suggest that the speech segmentation problem is not solved using statistical mechanisms alone. In Chapter 5, evidence was provided for statistically learned phonotactics from continuous speech by human learners. Unfortunately, no evidence was found for the learning of feature-based generalizations from continuous speech by human learners. The null result of Experiment 3 in Chapter 5 leaves the demonstration of the induction of constraints on natural classes from continuous speech as a topic for further research.

This dissertation has aimed at connecting statistical approaches to language acquisition (e.g., Saffran, Newport, & Aslin, 1996) with more traditional linguistic theories of language acquisition (e.g., Chomsky, 1981; Prince & Smolensky, 1993). That is, a view has been proposed in which statistical learning provides a basis for the learning of abstract linguistic constraints. In contrast to earlier linguistic theories that have assumed abstract phonotactic constraints to be innate (e.g., Prince & Smolensky, 1993; Prince & Tesar, 2004), the model presented here derives abstract phonotactic constraints from input data. The gap between statistical patterns and abstract constraints is bridged by a generalization mechanism which constructs abstract constraints on the basis of statistical regularities in the input. The proposed interaction of statistical learning and generalization thus has the consequence that learners construct abstract linguistic knowledge through generalization over observed data. This view is in line with studies that aim to minimize the role of Universal Grammar in explaining language acquisition, while still acknowledging the existence of abstract representations and constraints (e.g., Hayes, 1999; Hayes & Wilson, 2008). In addition, the learning model presented in this dis-

sertation induces markedness and contiguity constraints with varying degrees of generality. The mixed constraint set, combined with the use of constraint interaction through strict domination, allows for the learning of linguistic generalizations, as well as exceptions to these generalizations. The model thus makes no principled distinction between ‘exceptions’ and ‘regularities’. These notions have been incorporated into a single constraint set with varying levels of abstractness (see Albright & Hayes, 2003, for a similar proposal).

The proposal is in line with a growing body of research that show that phonotactic constraints are encoded both at the level of segments and the level of features (e.g., Goldrick, 2004; Albright, 2009). STAGE makes this distinction as a consequence of combining segment-based statistical learning and feature-based generalizations based on those segments. The result is a single constraint set that contains constraints at both the segment and feature level. This allows the model to represent phonotactic regularities as well as exceptions to those regularities, in a single formal framework. A distinction between exceptions and regularities simply follows from the model through differences in the level of generality, and the respective ranking of constraints.

There are many issues in the current proposal that need further investigation. As mentioned above, the exact conditions under which feature-based generalizations are constructed by human learners remain unclear. In addition, several properties of the computational model remain to be tested with psycholinguistic experiments. A topic for further investigation could, for example, be the psychological plausibility of the statistical induction thresholds that are used by the model. It was shown in Chapters 2 and 3 that setting thresholds on statistical values allows for a categorization of bigrams into ‘low’, ‘high’, and ‘neutral’ probability categories. These categories have different effects on segmentation (pressure towards segmentation, pressure towards contiguity, and no pressure, respectively). While the exact values of these thresholds were found not to affect the claim that generalization over specific bigrams within these categories improves segmentation, the psychological status of induction thresholds is unknown. One way to address this issue would be to set up experiments in which the effect of sequences from different categories on segmentation is tested. For example, spotting a target word  $C_mVC_n$  in a sequence  $C_kVC_lC_mVC_n$  is predicted to be facilitated by a markedness constraint  $*C_lC_m$  (since it would result in a phonotactic boundary which is aligned with the target word, and would thus result in faster responses). Conversely, spotting the target word would be hindered by a contiguity constraint  $CONTIG-IO(C_lC_m)$  (which would prevent the insertion of a word boundary before the onset of the target word, and, hence, would result in slower responses). It would be interesting to see at which thresholds (i.e., O/E ratios) listeners start to show such effects (provided that such effects are found at all). Note that

earlier models have also relied on statistical thresholds for segmentation (e.g., Cairns et al., 1997; Swingley, 2005). It is currently an open issue whether such thresholds have any cognitive status, or whether they are simply parameters that are required for the implementation of computational models of these sorts.

From a modeling perspective, it would be interesting to investigate how the model's learning of generalizations is affected by using different types of transcriptions and/or different sets of phonological features. For example, segment and feature transcriptions could be used that are closer to the phonetic properties of the speech signal, thereby reducing the *a priori* assumptions in the current implementation that the learner has already acquired the full segment and feature inventory of the native language. Another possible area of future exploration would be to investigate whether the combination of statistical learning and generalization, as proposed in the constraint induction model, could also be applied to the induction of other linguistic segmentation cues. Swingley (2005) suggests that statistical clustering of syllable *n*-grams could serve as a basis to bootstrap into the Metrical Segmentation Strategy (Cutler & Norris, 1988). While it is unclear what the exact generalization mechanism would look like in this case, the general view that statistical learning serves as a basis for generalization is in accordance with the predictions of STAGE. A similar view has been proposed for a different type of phonological learning by infants: White et al. (2008) propose that learning phonological alternations requires two different forms of computation. Infants first learn about dependencies between specific sounds in the input, and then group similar sounds that occur in complementary distribution into a single phonemic category.

A final issue with respect to the induction mechanism concerns the model's use of memory. The model derives phonotactic constraints from accumulated statistical information about biphone occurrences. The model thus assumes a perfect memory, which is a simplification of the learning problem. A more psychologically motivated memory implementation could be obtained by adding memory decay to the model: Biphones that were encountered a long time ago should be "forgotten". A similar approach has been proposed by Perruchet and Vinter (1998), who argue for the implementation of laws of associative learning and memory, such as temporal proximity, in segmentation models.

Two lines of research could provide more insight into the segmentation procedure that is used by human learners. First, more work needs to be done on specifying the relative contributions of specific and abstract constraints to segmentation. While segmentation studies in psycholinguistics so far have focused on one or other type of constraint, the results from Chapters 3 and 4 indicate that both may play an important role in segmentation. This calls for

experiments in which both types of constraints are considered (e.g., Moreton, 2002). A particularly interesting case would be to set up conflicts between specific and general constraints, similar to the conflicts that predicted by the computational model (see e.g., Figure 2.5). This could also shed light on the conflict resolution mechanism that is used by the learner. While the current implementation of the computational model is based on strict constraint domination, in which a higher-ranked constraint neutralizes the effects of any lower-ranked constraints, one could also conceive of a version of the segmentation model that is based on weighted constraints (e.g., Legendre et al., 1990; Pater, 2009). In this case, multiple lower-ranked constraints can join forces in order to defeat a higher-ranked constraint.

A second line of future work on segmentation could focus on the role of context in speech processing. That is, to what extent do neighboring sequences (e.g., *wx* and *yz* in a sequence *wxyz*) affect the detection of a word boundary in a sequence (e.g., *xy* in *wxyz*)? Computational and psycholinguistic studies have indicated that both the probability of the sequence itself (*xy*) and that of its immediate neighbors (*wx* and *yz*) can affect speech segmentation. Interestingly, the OT segmentation model has the potential to capture both types of effects. That is, if the model is presented with a sequence *wxyz*, then the model will only insert a word boundary if (a) there is a constraint *\*xy*, and (b) *\*xy* is ranked higher than *\*wx* and *\*yz*. This way, both the wellformedness of *xy* and the wellformedness of its neighbors play a role in segmentation. In contrast, earlier statistical segmentation models implemented either strategies based on the probability of *xy* ('threshold-based segmentation', e.g., Cairns et al., 1997), or strategies based on the probability of *xy* relative to its neighbors ('trough-based segmentation', e.g., Brent, 1999a). The OT segmentation model is thus the first model to integrate these two segmentation strategies into a single evaluation mechanism. Future experimental studies could investigate whether human learners segment speech in a similar integrated fashion.

#### 6.2.4 *The formation of a proto-lexicon*

One property of STAGE that became apparent in the simulations in Chapters 3 and 4 is that the model has the tendency to undersegment. That is, while the model makes few errors, it also misses a substantial number of word boundaries. The consequence is that the stretches of speech that are found between consecutive boundaries in the speech stream tend to be units that are larger than words. (See Appendix C for some examples of the output produced by STAGE.) Indeed, developmental studies have argued that the initial contents of the child's lexicon do not correspond to adult word forms, but rather contain larger units that need to be broken down further into

word-sized units (e.g., Peters, 1983). Frequency effects of larger-than-word units have also been found in adult speech processing (Arnon & Snider, 2010), indicating that learners may indeed store multiword phrases.

The ‘proto-words’ that were the result of the model’s segmentation of the artificial language in Chapter 4 were used as a predictor for human judgments on experimental items. The words produced by STAGE were a better predictor of human item preferences than words produced by statistical learning models, and words produced by applying a categorical segmentation based on OCP-PLACE. This provides a first indication that the proto-words that are generated by the model could be a plausible starting point for lexical acquisition. More work needs to be done to determine whether the output produced by STAGE can indeed be used to bootstrap into lexical acquisition. Future work on STAGE’s output could be done along two lines. First, it would be interesting to investigate how closely the proto-words that are created by STAGE resemble the initial contents of the infant’s vocabulary. A close resemblance would provide strong support for the proposed view that phonotactics facilitates the development of the mental lexicon. Second, it would be interesting to see whether contents of the proto-lexicon can be decomposed into a more adult-like lexicon of word forms. This could be done through a combination of bottom-up segmentation based on phonotactics, and top-down segmentation based on recurring patterns in the proto-lexicon (Apoussidou, 2010). In this view, phonotactic learning provides the building blocks for lexical acquisition, which subsequently requires refining the lexical entries in the proto-lexicon.

Another line of research would be to implement multiple bottom-up segmentation cues (e.g., phonotactic cues and metrical cues) into a single model in order to create a more accurate proto-lexicon. That is, other segmentation cues potentially provide a means of further breaking up larger chunks into actual words. Indeed, the use of multiple segmentation cues in a single model has been found to improve segmentation performance (Christiansen et al., 1998), and studies with both adults and infants show that various cues affect segmentation (Mattys et al., 1999, 2005). An interesting open issue is whether such additional cues would be able to detect exactly those boundaries that are missed by STAGE, thereby increasing the total number of boundaries that are detected by the learner, or whether such cues would overlap with phonotactic cues, and would thus contribute simply by providing a more robust cue for the discovery of proto-word boundaries.

### 6.2.5 *Language-specific phonotactic learning*

All studies in this dissertation used Dutch language data, and the experiments were conducted with participants that had Dutch as a native language. It would be interesting to conduct simulations using speech corpora from different languages. (See Appendix A for software that could be used for this purpose.) Note that phonotactics may not provide a useful segmentation cue in all languages. In general, more work needs to be done on establishing the contribution of phonotactics to the speech segmentation problem for a variety of languages. Of specific interest would be to investigate what the different units are that are targeted by phonotactic constraints in different languages. For example, a continuous speech stream in Hawaiian does not contain consonant clusters, due to the absence of codas and complex onsets in the language (e.g., Pukui & Elbert, 1986; Adler, 2006). For the case of Hawaiian, phonotactic constraints for segmentation cannot be based on consonant clusters. Listeners of such a language potentially develop constraints for segmentation that target sequences of syllables, or sequences of non-adjacent consonants. It is an open issue to what extent STAGE can account for speech segmentation in languages that have a phonotactic structure that is fundamentally different from Dutch.

A related issue for future research would be to apply the model to the problem of second language (L2) acquisition. It has been shown that learners of a second language acquire the phonotactic properties of the target language (Weber & Cutler, 2006; Trapman & Kager, 2009; Lentz & Kager, 2009). It seems worthwhile to investigate whether STAGE can provide a learnability account of second language learners' knowledge of L2 phonotactics. Such an enterprise could start from the assumption that the learner's L1 phonological knowledge is transferred and used as the initial state for L2 acquisition (e.g., Escudero & Boersma, 2004). The simplest approach to simulating L2 acquisition with STAGE would be to present the model with L1 input to which different amounts of L2 input has been added (e.g., Flege, 1995). Note that this would require L1 and L2 corpora which are comparable in terms of level of transcription. The result of adding L2 input would be a statistical database with accumulated statistics over L1 and L2 input. This would lead to a phonotactic grammar that is different from the L1 grammar, but also different from the L2 grammar. The 'mixed' statistics lead to different biphone constraints, which trigger different generalizations. Segmentation predictions can be derived from this mixed L1+L2 model, and these predictions can be tested experimentally. For example, a comparison can be made between the model trained on L1-only, L2-only, or L1+L2 data. The latter should provide the best fit to experimental data obtained from L2 learners. An alternative modeling approach could be to train two independent models (for L1 and



L2), and to compare harmony scores for segmentation candidates in both languages. Such a view would be more compatible with different languages modes for L1 and L2 (e.g., Grosjean, 2001).

#### 6.2.6 *A methodological note on the computational modeling of language acquisition*

The central enterprise of this dissertation has been the computational modeling of language acquisition. While the dissertation has been mainly concerned with the problem of learning phonotactics, and the segmentation of continuous speech, the approach taken here could prove useful for other studies focused on unraveling the mechanisms of human language acquisition. In particular, I would like to stress the importance of using both computational learners and human learners for the study of language acquisition. I believe that a psycholinguistic perspective is essential for any model that aims to explain human learning mechanisms. Therefore, the general approach taken here was to create a computational model based on psycholinguistic evidence of infant learning mechanisms. The predictions of the model can be tested using computer simulations. While computer simulations by themselves do not provide a strong test for the psychological plausibility of the model, they are very useful to provide insight into the learnability of linguistic knowledge, and into the contributions of different mechanisms to solving the learning problem. For example, the computer simulations in this dissertation have shown that phonotactic constraints are learnable from continuous speech, and that feature-based generalization improves the accuracy of the segmentation model.

Importantly, the model should also be tested for its psychological plausibility. There are two ways to address this issue. The first and most common approach is to test the model against human data, such as a set of wellformedness judgments. In Chapter 4, a similar approach was taken by testing different models for their fit against human segmentation data. Note, however, that a potential weakness of this approach is that computational models typically have a large number of parameters that can be modified in order to improve the fit of the model to the data. That is, a potential criticism of computational modeling is that, with the right parameter settings, any phenomenon or data set can be modeled. In addition to reporting such ‘best-fit’ results, it would be insightful if modelers would report a more complete picture of the model’s performance, either through testing the complete parameter space (e.g., the ROC curves in Experiment 2, Chapter 3) or by including ‘worst-fit’ numbers.

I propose that a stronger test for the psychological plausibility of a computational model is to derive novel predictions from the model, and to set up new experiments that are specifically aimed at testing these predictions (Kager

& Adriaans, 2008). The artificial language learning paradigm provides an important methodological tool for the study of human learning mechanisms. By training participants on artificial languages with particular constraints and testing their acquired knowledge of the language, artificial language learning makes it possible to investigate the language learning abilities of infants and adults in a highly controlled environment. This controlled environment may, however, come at the expense of the naturalness of the input (e.g., Johnson & Tyler, 2010). This is why the combination of computational modeling and artificial language learning may prove to be especially fruitful: Computational models provide explicit descriptions of learning mechanisms, and can be trained on realistic data from natural languages (such as accurately transcribed speech corpora). Artificial language learning can provide a test for the psychological plausibility of these mechanisms by presenting human language learners with highly controlled, but simplified, data. In sum, the computational modeling of language acquisition would benefit from a cycle that includes both computer modeling and psychological experiments.

### 6.3 CONCLUSIONS

This dissertation has been concerned with the mechanisms that allow language learners to induce phonotactic constraints for speech segmentation. It was shown that phonotactics can be learned from continuous speech by combining mechanisms that have been shown to be available to infant and adult language learners. These mechanisms, statistical learning and generalization, allow for the induction of a set of phonotactic constraints with varying levels of abstraction, which can subsequently be used to predict the locations of word boundaries in the speech stream. The proposed combination of statistical learning and feature-based generalization provides a better account of speech segmentation than models that rely solely on statistical learning. This result was found both in computer simulations, and in simulations of human segmentation data. In addition, it was shown that human learners can learn novel phonotactic constraints from a continuous speech stream from an artificial language. The dissertation advocates an approach to the study of language acquisition that focuses on both computational and human learners. By integrating evidence from computational and human learners, models can be developed which are both formally explicit and psychologically plausible. Such converging evidence brings us closer to a thorough understanding of the mechanisms that are involved in language acquisition.



## THE STAGE SOFTWARE PACKAGE

---

This manual describes how to use the STAGE Software Package (SSP). The software can be used to run phonotactic learning and speech segmentation simulations, as described throughout this dissertation (see also, Adriaans & Kager, 2010). In addition, it allows the user to train models on new data sets, with the possibility to use other statistical measures, thresholds, and a different inventory of phonological segments and features. The package implements several learning models (STAGE, transitional probability, observed/expected ratio) and segmentation models (Optimality Theory, threshold-based segmentation, trough-based segmentation). Please see Chapters 2 and 3 for details about these models. This manual describes how to run simulations with SSP.

### GETTING STARTED

The software (SSP.zip) can be downloaded from:

<http://www.hum.uu.nl/medewerkers/f.w.adriaans/resources/>

The program uses configuration files, which specify how to run the software. In addition, the user will need to provide a training/test corpus and an inventory file specifying phonological segments and features.<sup>1</sup> The contents and required format of these files are described in the next sections.

### *Contents of the package*

- `ssp.pl` - The main script, written in Perl.
- `Input` - This is the folder where the corpus and inventory files should be placed.
- `Modules` - This contains several Perl modules which are used by the main script.
- `Output` - This is the folder where the program writes its output to. A new folder will be created for each configuration file that is used to run the program.
- `ssp_manual.pdf` - This manual.
- Several configuration files (`.txt`) have been included which can be used to replicate the simulations reported in the dissertation, and which can be modified to create new simulations.

---

<sup>1</sup> Examples of configuration files and an inventory file for Dutch are included in the package. The Spoken Dutch Corpus (*Corpus Gesproken Nederlands*) is distributed by the Dutch HLT Agency (TST-centrale; <http://www.inl.nl/tst-centrale>) and can be ordered there.

### *Running the program*

To start the program, use Command Prompt (Windows) or Terminal (Unix, Mac OS X) to go to the SSP folder and type:

```
perl ssp.pl configuration.txt
```

where the configuration file 'configuration.txt' can be replaced by any other file name. Please note that running Single-Feature Abstraction may take several minutes, while the segmentation of *wxyz* sequences ('segmentation using context') may take up to several hours, depending on the speed of your computer.

### INPUT FILES

All input files (corpus files, feature definitions) should be placed in the Input folder.

### *Corpora*

Input to the program consists of the following:

- **Training file** - A training file is a single text (.txt) file with phonemic transcriptions. Each segment should be represented by a single ASCII character (such as in the CELEX DISC alphabet). Each line contains a sequence of segments. This can either be a single word or multiple words glued together. The sequences are optionally preceded by a digit (utterance IDs) and optionally followed by a digit specifying a (token) frequency count. The entries should be tab-separated.

Some examples:

(a) *A corpus of transcribed utterances of continuous speech (with utterance IDs):*

```
5   dizKnv@rstElbardor@tsKnsxufpot@
6   d}s@hKstadnyOps@nlaxlaxst@stAnt
7   laG@rdAnsokAni
... ..
```

(b) *A lexicon of isolated words (with token frequencies):*

```
dQg  2061
k{t  1195
... ..
```

(c) *Sequences of segments:*

```
mademobetumopodemopotubipotumo
...
```

- **Test file** - The trained model can be used to predict the locations of word boundaries in a test set. Similar to the training file, a test file consists of sequences of segments, optionally preceded by a digit to identify the utterance.

- **Correct file** - A 'correct' file can be specified to evaluate the model's segmentation of the test set. In order to link the correct segmentation of an utterance to the predicted segmentation of the test set, the user can specify utterance ID digits in both files, so that the program knows which correct utterance corresponds to which predicted utterance in the test set.

For example:

```
6 d} s@ hKstad ny Op s@n lax lax st@stAnt (Test file)
6 d}s @ hK stad ny Op s@n lax laxst@ stAnt (Correct file)
```

If no utterance IDs have been specified, the program will assume identical orderings in both files (i.e., the first utterance in the test file corresponds to the first utterance in the correct file, etc.).

### *Segment/feature inventory*

A tab-separated text file is used to define the segment inventory with corresponding feature values. Each line contains one such specification:

```
Segment1 Feature1=Value1 Feature2=Value2 ...
Segment2 Feature1=Value1 Feature2=Value2 ...
...           ...           ...           ...
```

For example:

```
p ... Cons=1 Voice=0 Place=Lab ...
b ... Cons=1 Voice=1 Place=Lab ...
... ...           ...           ...           ...
```

Note that such an inventory only needs to be specified when using `STAGE`. The statistical learning models (TP, O/E) do not require inventory files. The user is free to use different names for features and feature values (as long as they are consistent within the file). There is no restriction with respect to possible feature values.<sup>2</sup> That is, any two different values for a single feature is counted as a feature difference (e.g.,  $\text{Voice}=0 \neq \text{Voice}=1$ ;  $\text{Voice}=+ \neq \text{Voice}=-$ ,  $\text{Voice}=\text{abc} \neq \text{Voice}=\text{cba}$ , etc.).

Some restrictions hold with respect to feature definitions. First, the order in which the features are specified for each segment should be fixed. Second, the file should not contain multiple segments with identical feature specifications. Segments with such zero-feature differences should be replaced by a single (segment) symbol representing that feature specification. Third, the model only processes constraint pairs that have feature vectors of the same length. Each segment should thus be fully specified. Note that the user may use different features for consonants and vowels. In this case, consonant and vowel vectors will have different lengths and are not compared to each other. The consequence is that the model only compares CV constraints to CV constraints, CC constraints to CC constraints, etc. An example of such a feature

<sup>2</sup> With the exception of the symbol '?', which is used for program-internal purposes

## THE STAGE SOFTWARE PACKAGE

definition is included in the software package (`inventory-dutch.txt`). Finally, make sure that all segments that occur in the corpus are defined in the inventory.

## CONFIGURATION FILES

A configuration file is a text file specifying how to run the software. The user may modify parameters in configuration files in order to apply models to new data sets, or may alter the way in which a model processes its input, when it induces constraints, etc.<sup>3</sup> (See also the example configuration files that are included in the package.) Table A.1 shows some basic parameters for the program, such as a specification of which learning model is to be used (`MODEL`); which training and/or test data the model should be applied to (`TRAININGSET`, `TESTSET`); where to find the the correct segmentations for the test set (`CORRECT`); and, if `STAGE` is used, which feature set should be used (`INVENTORY`).

Several input processing parameters are given in Table A.2. The user controls what type of sequences are extracted from the corpus (`TRAININGPATTERN`). The default sequence is a simple biphone pattern (`TRAININGPATTERN = xy`). The software also supports more sophisticated processing windows, such as non-adjacent dependencies. For example, the model can be trained on non-adjacent consonants by specifying a larger, feature-based pattern (`TRAININGPATTERN = [syll=0] [syll=1] [syll=0]`).<sup>4</sup> When such a larger window is used for training, the user needs to specify the positions for which the statistical dependency is to be computed. In the current example, a dependency between the first and third positions needs to be specified (i.e., ignoring the vowel in the second position): `DEPENDENCY = 1-3`. The user may specify whether token frequencies (which are optionally included in the training set) are to be used in training the model (`COUNT`).

It should be noted that the resulting model is always 'biphone'-based. That is, only the two dependent elements are encoded in the model (e.g., `*CC` rather than `*C(V)C`). The constraints are applied directly to the test set (i.e., the test set is not pre-processed in any way). Therefore, when non-adjacent constraints are used, the user may want to filter the test set accordingly (e.g., removing all vowels from a test set containing `CVCVCV...` sequences, resulting in a test set containing only `CCC...` sequences). During testing the model either uses context or not (`CONTEXT`). That is, the test set is processed using either an `xy` (i.e., `CONTEXT = no`) or a `wxyz` (i.e., `CONTEXT = yes`) window.

Several additional (optional) parameters control more technical aspects of the learning and segmentation models (see Tables A.3 and A.4). The user can change the statistical learning formula that is used to create the statistical distribution (`SLFORMULA`) using any of the measures described in Table A.5; change the statistical measure that

<sup>3</sup> The flexibility provided by the various parameters in SSP is provided to encourage and facilitate research on phonotactic learning and speech segmentation. It is the responsibility of the user, of course, to decide whether a particular new configuration makes any theoretical/empirical sense.

<sup>4</sup> The features that are used in the pattern are defined by the user in the segment/feature inventory. A pattern may also involve multiple (comma-separated) features (e.g., `[syll=0, son=0, cont=1]`)

is used to compute constraint ranking values (RANKING); change the constraint induction thresholds (THRESHOLD.M, THRESHOLD.C) for STAGE; and change the segmentation thresholds (THRESHOLD.OE, THRESHOLD.TP) for the threshold-based statistical learning models. Finally, a default boundary probability (BOUNDARYPROB) can be specified for missing biphones, random baselines, etc.

#### OUTPUT FILES

For each simulation, a new folder will be created within the Output folder. This folder carries the name of the configuration file that was used in the simulation and, depending on the particular configuration that was used, will contain one or more of the following output files:

- *Configuration-model.txt*
  - ... contains either the constraint set with ranking values (STAGE), or biphones with associated transitional probabilities or observed/expected ratios (TP, O/E).
- *Configuration-decisions.txt*
  - ... contains all sequences that were processed in the test set (i.e., either *xy* or *wxyz* sequences). For each sequence, a classification is given, which is the probability of a boundary appearing in the middle of the sequence. That is, a classification of '0' means that a boundary is never inserted, the classification '1' means that a boundary is always inserted, and values between 0 and 1 indicate the probability of the insertion of a boundary into sequence in the test set. The final column displays how often the sequence appears in the test set.
- *Configuration-segmentation.txt*
  - ... contains the segmentation of the complete test set, as predicted by the model.
- *Configuration-results.data*
  - ... contains evaluation metrics (hit rate, false alarm rate) reflecting the model's ability to predict word boundaries in the test set.

More precisely:

Predicted segmentation	Correct segmentation	Label
'x.y'	'x.y'	<i>TruePositive</i>
'x.y'	'xy'	<i>FalsePositive</i>
'xy'	'xy'	<i>TrueNegative</i>
'xy'	'x.y'	<i>FalseNegative</i>

#### THE STAGE SOFTWARE PACKAGE

Hit rate ( $H$ ):

$$H = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (A.1)$$

False alarm rate ( $F$ ):

$$F = \frac{FalsePositives}{FalsePositives + TrueNegatives} \quad (A.2)$$

- *Configuration-tableaus.txt*
  - ...contains OT tableaus for each sequence in the test set. The optimal candidate is indicated by an arrow '->' and decides the classification (boundary probability) of the sequence. (*only applies to STAGE*)

#### SSP CONFIGURATION FILE PARAMETERS

The tables on the following pages (Table A.1 - Table A.5) define parameters which can be specified by the user in the configuration files.



Table A.1: Basic model and file parameters.

Parameter	Possible values	Description
MODEL <sup>1</sup>	StaGe, 0/E, TP, random	The learning model
TRAININGSET <sup>1</sup>	File name	Training file
TESTSET	File name	Test file
CORRECT	File name	Correct file
INVENTORY <sup>2</sup>	File name	Segment and feature definitions

Note. <sup>1</sup> = required fields, <sup>2</sup> = required if StaGe is used.

Table A.2: Input processing parameters.

Parameter	Possible values	Description	Default
TRAININGPATTERN	xy, [F1=v1] [F2=v2] [...]	Phonological pattern used for training.	xy
DEPENDENCY	1-3, 1-2, 2-3, etc.	The positions of the elements for which the statistical dependency is to be computed.	-
COUNT	type, token	Specifies whether word frequency counts in the training file should be taken into account.	type
CONTEXT	yes, no	Specifies whether context is to be used during segmentation.	no

Table A.3: Learning parameters for StaGe.

Parameter	Possible values	Description	Default
SLFORMULA	(See Table A.5)	Formula that is used to implement statistical learning.	0/E
RANKING	(See Table A.5)	Statistical measure that is used to rank the constraints.	E
THRESHOLD_M	Any number.	Statistical threshold for the induction of a markedness constraint.	0.5
THRESHOLD_C	Any number.	Statistical threshold for the induction of a contiguity constraint.	2.0

Table A.4: Segmentation parameters for statistical learning models (TP, O/E) and random baselines (*random*).

Parameter	Possible values	Description	Default
THRESHOLD_OE	Any positive real number	Segmentation threshold for O/E threshold-based segmentation.	1.0
THRESHOLD_TP	Any probability between 0 and 1	Segmentation threshold for TP threshold-based segmentation.	$Threshold_{xy} = \frac{1}{ Y_x }$
DEFAULTPROB	Any probability between 0 and 1	Probability of inserting a word boundary in case the learning model cannot decide (e.g., unseen biphones, random baselines)	0.5

*Note.* If no fixed TP threshold is specified by the user, the program will calculate a threshold based on the default assumption that all successors of a segment are *a priori* equally probable. The threshold thus equals  $\frac{1}{|Y_x|}$ , where  $|Y_x|$  is the number of possible successors for  $x$ .

Table A.5: Statistical measures.

Symbol	Description
O	observed frequency (= biphone frequency of occurrence)
F <sub>x</sub>	observed frequency of "x." (= frequency of first-position uniphone)
F <sub>y</sub>	observed frequency of ".y" (= frequency of second-position uniphone)
E	expected frequency (= uniphone-based estimated biphone frequency)
O/E	observed / expected ratio
O-E	observed minus expected (= the difference between O and E)
logO/E	log observed/expected ratio, with base 10
MI	pointwise mutual information (= log O/E ratio, with base 2)
TP	forward transitional probability
logTP	log forward transitional probability, with base 10
TP <sub>b</sub>	backward transitional probability
logTP <sub>b</sub>	log backward transitional probability, with base 10

# B

## FEATURE SPECIFICATIONS

### CONSONANTS

DISC:	<b>p</b>	<b>b</b>	<b>t</b>	<b>d</b>	<b>k</b>	<b>g</b>	<b>f</b>	<b>v</b>	<b>s</b>	<b>z</b>	<b>ʃ</b>	<b>ʒ</b>
IPA:	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ
<i>syl</i>	–	–	–	–	–	–	–	–	–	–	–	–
<i>cons</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>appr</i>	–	–	–	–	–	–	–	–	–	–	–	–
<i>son</i>	–	–	–	–	–	–	–	–	–	–	–	–
<i>cont</i>	–	–	–	–	–	–	+	+	+	+	+	+
<i>nas</i>	–	–	–	–	–	–	–	–	–	–	–	–
<i>voice</i>	–	+	–	+	–	+	–	+	–	+	–	+
<i>place</i>	<i>lab</i>	<i>lab</i>	<i>cor</i>	<i>cor</i>	<i>dors</i>	<i>dors</i>	<i>lab</i>	<i>lab</i>	<i>cor</i>	<i>cor</i>	<i>cor</i>	<i>cor</i>
<i>ant</i>	+	+	+	+	–	–	+	+	+	+	–	–
<i>lat</i>	–	–	–	–	–	–	–	–	–	–	–	–

DISC:	<b>x</b>	<b>ɣ</b>	<b>ç</b>	<b>m</b>	<b>n</b>	<b>ŋ</b>	<b>r</b>	<b>l</b>	<b>w</b>	<b>j</b>	<b>h</b>
IPA:	x	ɣ	ç	m	n	ŋ	r	l	w	j	h
<i>syl</i>	–	–	–	–	–	–	–	–	–	–	–
<i>cons</i>	+	+	+	+	+	+	+	+	–	–	+
<i>appr</i>	–	–	–	–	–	–	+	+	+	+	–
<i>son</i>	–	–	–	+	+	+	+	+	+	+	–
<i>cont</i>	+	+	–	–	–	–	+	+	+	+	+
<i>nas</i>	–	–	–	+	+	+	–	–	–	–	–
<i>voice</i>	–	+	+	+	+	+	+	+	+	+	+
<i>place</i>	<i>dors</i>	<i>dors</i>	<i>cor</i>	<i>lab</i>	<i>cor</i>	<i>dors</i>	<i>cor</i>	<i>cor</i>	<i>labdors</i>	<i>cor</i>	<i>glot</i>
<i>ant</i>	–	–	–	+	+	–	+	+	+	–	–
<i>lat</i>	–	–	–	–	–	–	–	+	–	–	–

FEATURE SPECIFICATIONS

VOWELS

DISC:	i	!	I	y	(	u	e	E	)		}	*	o	O
IPA:	i	i:	ɪ	y	y:	u	e:	ɛ	ɛ:	ø:	ʉ	œ:	o:	ɔ
<i>high</i>	+	+	+	+	+	+	-	-	-	-	-	-	-	-
<i>low</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>back</i>	-	-	-	-	-	+	-	-	-	-	-	-	+	+
<i>round</i>	-	-	-	+	+	+	-	-	-	+	+	+	+	+
<i>long</i>	-	+	-	-	+	-	+	-	+	+	-	+	+	-
<i>tense</i>	+	+	-	+	+	+	+	-	-	+	-	-	+	-
<i>nasd</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-

DISC:	<	@	a	A	K	L	M	q	o	~	^	3	#
IPA:	ɔ:	ə	a:	ɑ	ei	œy	au	ã:	ẽ:	õ:	ũ:	ɜ:	ɑ:
<i>high</i>	-	-	-	-	-+	-+	-+	-	-	-	-	-	-
<i>low</i>	-	-	+	+	-	-	-	+	-	-	-	-	+
<i>back</i>	+	+	+	+	-	-	+	+	-	+	-	+	+
<i>round</i>	+	-	-	-	-	+	+	-	-	+	+	-	-
<i>long</i>	+	-	+	-	+	+	+	+	+	+	+	+	+
<i>tense</i>	-	-	+	-	+	+	+	-	-	-	-	-	-
<i>nasd</i>	-	-	-	-	-	-	-	+	+	+	+	-	-

# C

## EXAMPLES OF SEGMENTATION OUTPUT

---

Orthography: *Ik zou natuurlijk idioot zijn als ik ja zou zeggen he?*  
 Translation: 'I would of course be an idiot if I would say yes.'  
 Transcription: ik zau natylək idijot sein als ik ja zau zεγə hε

---

STAGE: ik zaunatylək idijot sein alsik jazauzeγə hε  
 O/E: ik zau natyl ək idijo tsein al sik ja zau zε γə hε  
 TP: ik zaunat ylək idij ots ein als ik ja zauze γə hε

---

Orthography: *Toch een heel bekend standaardwerk.*  
 Translation: 'Just a well-known standard work.'  
 Transcription: tɔx ən hel bəkent standardwɛrək

---

STAGE: tɔxən helbəkent standard wɛrək  
 O/E: tɔxə nhel bək ən tst and ard wɛr ək  
 TP: tɔxən hel bək ɛnt st and ard wɛ rək

---

Orthography: *Liefde is de enige manier om je tegen de dood te verzetten.*  
 Translation: 'Love is the only way to resist death.'  
 Transcription: livdə is ət enəγə manir ɔm jə teγə də dot tə vɛrzetə

---

STAGE: liv də isətənəγəmanirɔm jətɛγədədot təvɛr zɛtə  
 O/E: liv də isət ən əγə ma ni rɔm jət ɛγə dədot təvɛr zɛ tə  
 TP: liv də isət ənə γə manir ɔm jət ɛγə dəd ot tə vɛr zɛ tə

---

Orthography: *Vond ik heel erg boeiend.*  
 Translation: 'I found (it) very interesting indeed.'  
 Transcription: fɔnd ik hel ɛrəx bujənt

---

STAGE: fɔndɪk helɛrəx bujənt  
 O/E: fɔ nd ik hel ɛr əx bujənt  
 TP: fɔnd ik hel ɛrəx bujənt

EXAMPLES OF SEGMENTATION OUTPUT

Orthography:	<i>En ik heb ook in ieder geval uh ja een paar collega's waarmee ik heel goed daarmee zou kunnen samenwerken.</i>
Translation:	'And I also have in any case uh yes a couple of colleagues with whom I might very well collaborate.'
Transcription:	ɛn ik hɛp ok ɛn idə xəfəl ə ja ɛm pɑr kɔlexas wame ik hel xut dame zɑu kʉnə saməwɛrəkə
STAGE:	ɛnik hɛpɔkənɪdɛxəfəlɔjə ɛm pɑrkɔlexas wame ik hel xut damezɑukʉnəs aməwɛrəkə
O/E:	ɛn ik hɛp ok ɛni də xə fələ ja ɛmp ar kɔ le xa swa me ik hel xut dame zɑu kʉn əs am əwɛr əkə
TP:	ɛn ik hɛp okɛn idə xə fə lə ja ɛm pɑr kɔ le xas wa me ik hel xut damez auk ʉnəs amə wɛ rə kə
Orthography:	<i>Ik weet nog niet precies hoe ik zal gaan.</i>
Translation:	'I don't know yet precisely how I will go.'
Transcription:	k wɛt nɔy nit prɛsis hu wɪk sɑl xɑn
STAGE:	kwɛt nɔy nit prɛsis hu wɪk sɑlxɑn
O/E:	kwɛt nɔy nit pr ɛsis hu wɪks ɑl xɑn
TP:	kwɛt nɔy nit prɛs is hu wɪks ɑl xɑn
Orthography:	<i>Maar in ieder geval in die film heeft ie wat langer haar.</i>
Translation:	'But in any case in this film his hair is somewhat longer.'
Transcription:	mɑ m i fɑl m di fɪlm heft i wɑt lɑŋə hɑr
STAGE:	mɑ mɪfɑlmdɪfɪlm hef ti wɑt lɑŋə hɑr
O/E:	mɑ mi fəl m dɪfɪl mheft tiwɑt lɑŋ əh ar
TP:	mɑ mi fəl mɪdi fɪlm he ft iwɑt lɑ ŋə hɑr
Orthography:	<i>Een paar jaar geleden heeft ze haar restaurant verkocht.</i>
Translation:	'A few years ago she sold her restaurant.'
Transcription:	ɛm pɑ ja xələdɛn heft sə hɑ rɛsturɑː fɛrkɔxt
STAGE:	ɛm pɑjɑxələdɛn heft sə hɑrɛsturɑːfɛr kɔxt
O/E:	ɛmp əjɑ xə le də nheft ts əh ərə sturɑːf ər kɔxt
TP:	ɛm pɑjɑ xə ledɛn he ftsə hɑr ɛst ʉr ɑːfɛr kɔxt

EXAMPLES OF SEGMENTATION OUTPUT

Orthography:	<i>Binnen in de vuurtoren zit een groot dier in een schommelstoel.</i>
Translation:	'Inside the lighthouse a large animal is sitting in a rocking chair.'
Transcription:	bɪnən ɪ də vʏrtorən zɪt əɲ xrod dir ɪn ən sɔ̃mɔ̃stul
STAGE:	bɪnən ɪndəvʏrtorən zɪ təɲ xrod dɪrɪn ən sɔ̃mɔ̃ stul
O/E:	bɪn ən ɪndəv ʏrt or ənzɪ tə ɲx rod di rɪn ən sɔ̃ ɔ̃mɔ̃ st ul
TP:	bɪn ən ɪn dəv ʏrt orən zɪ tə ɲxrod dir ɪn ən sɔ̃ ɔ̃m ɔ̃st ul
Orthography:	<i>Die horizon kan toch ook naar ons komen.</i>
Translation:	'That horizon may also come to us.'
Transcription:	di horizɔ̃n kən tɔ̃x ok nar ɔ̃ns komə
STAGE:	dihorizɔ̃n kən tɔ̃xok narɔ̃ns komə
O/E:	di hor iz ɔ̃n kən tɔ̃x ok nar ɔ̃ns komə
TP:	di hor izɔ̃n kant ɔ̃x ok nar ɔ̃ns komə





# D

## EXPERIMENTAL ITEMS

---

Table D.1: Test trials for Experiment 1

Trial	ABX word	BXA word	Trial	ABX word	BXA word
1	fabuxau	dœylyfa	19	sabœyly	punosœi
2	sapunau	dœylofi	20	zapœyxy	punausa
3	sapœyny	penyseï	21	zapuxau	dulofœi
4	fabœxy	penausi	22	zadunau	dulaufa
5	feibexy	dunozeï	23	zadœyny	pœynysa
6	feidely	dunauza	24	fadœyly	pœynosi
7	feidulo	bœxyyfa	25	fidelau	belyseï
8	seibulo	bœxyofi	26	zidenau	belausi
9	seibely	denauzi	27	zidœyno	pœyxozi
10	zeideny	denyzeï	28	sibœylo	pœyxyza
11	zeiduno	buxofœi	29	sibelau	bulausa
12	feibuxo	buxaufa	30	sipenau	buloseï
13	fibexau	bœylosi	31	sipœyno	puxauza
14	zipexau	bœylysa	32	fidœylo	puxozœi
15	zipœyxo	pexauzi	33	seipeny	dœynyza
16	fibœyxo	pexyzeï	34	zeipexy	dœynozi
17	fadulau	delaufi	35	zeipuxo	bexyfeï
18	sabulau	delyfeï	36	seipuno	bexaufi

*Note.* The order of test trials was randomized. The order of presentation of ABX and BXA items within trials was balanced within and across participants.

EXPERIMENTAL ITEMS

Table D.2: Test trials for Experiment 2

Trial	ABC word	BCA word	Trial	ABC word	BCA word
1	padego	zekypa	19	tozagu	zikepy
2	pedyga	zakepo	20	tizegy	zokapu
3	pazeky	zegotu	21	tibaxo	dugopa
4	pezoku	zagety	22	tuboxa	digapo
5	pedugo	bixeta	23	tizoge	duxeso
6	pidega	bexuto	24	tuzego	dixose
7	peziku	bikasy	25	sabiku	dexyso
8	pizaky	bekisu	26	sebyko	daxisu
9	pydige	duxosy	27	sadyxe	degopu
10	pudogy	dyxise	28	sedoxu	dagype
11	pyzoke	dugypa	29	sybeka	zokape
12	puzyka	dygope	30	sobake	zykepa
13	tabixe	bokesa	31	sydaxu	zogity
14	tobexa	bakise	32	sodixy	zygatu
15	tazugo	boxeta	33	sabeky	bixute
16	tozega	baxuto	34	sibuke	baxety
17	tobaxe	ziguto	35	sadexu	bikosa
18	tibuxo	zogate	36	sidoxa	bakesu

*Note.* The order of test trials was randomized. The order of presentation of ABC and BCA items within trials was balanced within and across participants.

## EXPERIMENTAL ITEMS

Table D.3: Test trials for Experiment 3

Trial	Initial		Trial	Final	
	Legal (vCA)	Illegal (fCA)		Legal (BCf)	Illegal (BCv)
1	vygate	fykape	19	dagefo	zakevo
2	viguta	fikusa	20	bukyfa	dugyva
3	vikopy	figoty	21	zekofa	degova
4	vekaso	fegato	22	dygafu	bykavu
5	vigaty	fixasy	23	dugefy	buxevy
6	vagepo	fakeso	24	zikofy	bixovy
7	vyxasu	fygatu	25	byxufe	dyguve
8	vukesy	fugepy	26	byxifo	zykivo
9	vexuta	fekusa	27	zekafo	dexavo
10	vygupe	fyxute	28	zygafe	bykave
11	vukose	fuxote	29	daxofu	zakovu
12	vexota	fegopa	30	bikyfa	zigyva
13	vyxetu	fykepu	31	byxefu	zygevu
14	vugypa	fuxysa	32	zagife	daxive
15	vykipo	fyxiso	33	zigufa	bixuva
16	vaxise	fagipe	34	dexufa	zeguva
17	vixysa	fikypa	35	bikafy	dixavy
18	vakopu	faxotu	36	duxofe	bukove

*Note.* The order of test trials was randomized. The order of presentation of legal and illegal items within trials was balanced within and across participants.



## REFERENCES

---

- Adler, A. N. (2006). Faithfulness and perception in loanword adaptation: A case study from Hawaiian. *Lingua*, 116, 1024-1045.
- 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 generalisation as a source of gradient acceptability. *Phonology*, 26, 9-41.
- Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with Minimal Generalization. In M. Maxwell (Ed.), *Proceedings of the sixth meeting of the ACL special interest group in computational phonology* (p. 58-69). Philadelphia: Association for Computational Linguistics.
- Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90(2), 119-161.
- Altmann, G. (2002). Learning and development in neural networks - the importance of prior experience. *Cognition*, 85, B43-B50.
- Apoussidou, D. (2007). *The learnability of metrical phonology*. Doctoral dissertation, University of Amsterdam. (LOT dissertation series 148).
- Apoussidou, D. (2010). UP on StaGe: A lexical segmentation strategy based on phonotactics. *Paper presented at The Eighteenth Manchester Phonology Meeting*.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67-82.
- 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.
- Baayen, H. R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44, 568-591.
- Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, 83, 167-206.
- Berent, I., Marcus, G. F., Shimron, J., & Gafos, A. I. (2002). The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition*, 83, 113-139.
- 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., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104, 591-630.

## References

- Berkley, D. (2000). *Gradient Obligatory Contour Principle effects*. Unpublished doctoral dissertation, Northwestern University.
- Boersma, P. (1998). *Functional phonology. Formalizing the interactions between articulatory and perceptual drives*. Doctoral dissertation, University of Amsterdam. (LOT dissertation series 11).
- Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th international congress of phonetic sciences* (p. 1013-1016). Barcelona.
- Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32(1), 45-86.
- Boll-Avetisyan, N., & Kager, R. (2008). Identity avoidance between non-adjacent consonants in artificial language segmentation. In *Poster presented at Laboratory Phonology 11*. Victoria University of Wellington, New Zealand, June 30 - July 2, 2008.
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science*, 16, 451-459.
- Booij, G. (1995). *The phonology of Dutch*. Oxford, UK: Oxford University Press.
- Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16, 298-304.
- Brent, M. R. (1999a). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 34, 71-105.
- Brent, M. R. (1999b). 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. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87, B69-B77.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2010). A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36, 821-828.
- Chomsky, N. (1981). *Lectures on government and binding*. Berlin: Mouton de Gruyter.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper and Row.

- 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.
- Christiansen, M. H., Onnis, L., & Hockema, S. A. (2009). The secret is in the sound: from unsegmented speech to lexical categories. *Developmental Science*, *12*, 388-395.
- Christie, W. M. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, *55*, 819-821.
- Christophe, A., Dupoux, E., Bertoni, J., & Mehler, J. (1994). Do infants perceive word boundaries? an empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*, 1570-1580.
- Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: The MIT Press.
- Coetzee, A. W. (2004). *What it means to be a loser: Non-optimal candidates in Optimality Theory*. Doctoral dissertation, University of Massachusetts.
- Coetzee, A. W. (2005). The Obligatory Contour Principle in the perception of English. In S. Frota, M. Vigfó, & M. J. Freitas (Eds.), *Prosodies* (p. 223-245). New York, NY: Mouton de Gruyter.
- Coetzee, A. W., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, *26*, 289-337.
- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (p. 133-163). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coleman, J. S., & Pierrehumbert, J. B. (1997). Stochastic phonological grammars and acceptability. In *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology* (p. 49-56). Somerset, NJ: Association for Computational Linguistics.
- Cristià, A., & Seidl, A. (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development*, *4*, 203-227.
- Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. Altmann (Ed.), *Cognitive models of speech processing* (p. 105-121). Cambridge, MA: The MIT Press.
- Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (p. 87-99). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218-236.

## References

- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385-400.
- Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.
- Daland, R. (2009). *Word segmentation, word recognition, and word learning: a computational model of first language acquisition*. Doctoral dissertation, Northwestern University.
- 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*, 25, 1568-1578.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Vrecken, O. van der. (1996). The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Proceedings of the icslp'96* (p. 1393-1396). Philadelphia, PA.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, 62, 2187-2209.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551-585.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874.
- Fenson, L., Dale, P. S., Bates, E., Reznick, J. S., Thal, D., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59, 1-173.
- Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, 61, 423-437.
- 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.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 15822-15826.



- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.
- Fowler, C., Best, C., & McRoberts, G. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, *48*, 559-570.
- Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287-295.
- 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. A. (2001). The psychological reality of ocp-place in Arabic. *Language*, *77*(1), 91-106.
- Goddijn, S., & Binnenpoorte, D. (2003). Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus. In *Proceedings of the 15th international congress of phonetic sciences* (p. 1361-1364). Barcelona.
- Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*, 586-603.
- Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*, 1155-1164.
- Goldsmith, J. A. (1976). *Autosegmental phonology*. Cambridge, MA: Doctoral dissertation, MIT.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431-436.
- Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109-135.
- Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*, 567-580.
- Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*, 183-206.
- Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*, 229-252.
- Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access ii. infant data. *Journal of Memory and Language*, *51*, 548-567.
- Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 344-359.
- Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (p. 1-22). Oxford, UK: Blackwell.

## References

- Hamann, S., Apoussidou, D., & Boersma, P. (to appear). Modelling the formation of phonotactic restrictions across the mental lexicon. In *Proceedings of the 45th meeting of the Chicago Linguistic Society*.
- Hamann, S., & Ernestus, M. (submitted). Adult learning of phonotactic and inventory restrictions. *Manuscript submitted for publication*.
- Hanulíková, A., McQueen, J. M., & Mitterer, H. (2010). Possible words and fixed stress in the segmentation of Slovak speech. *The Quarterly Journal of Experimental Psychology*, 63, 555-579.
- Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, 3, 367-382.
- Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in Laboratory Phonology vi* (p. 58-74). Cambridge, UK: Cambridge University Press.
- Hayes, B. (1999). Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. J. Newmeyer, & K. M. Wheatley (Eds.), *Functionalism and formalism in linguistics* (p. 243-285). Amsterdam: John Benjamins.
- Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 379-440.
- Hockema, S. A. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development*, 2, 119-146.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.
- 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., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology*, 46, 65-97.
- 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: The MIT Press.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1-23.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.

## References

- Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & 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., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, *40*, 62-82.
- Jusczyk, P. W., Hohne, E., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*, 1465-1476.
- 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*, 630-645.
- Jusczyk, P. W., Smolensky, P., & Alocco, T. (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition*, *10*, 31-73.
- Kager, R., & Adriaans, F. (2008). The acquisition of artificial languages: Models and human data. *Contributed symposium at the 16th Annual Meeting of the European Society for Philosophy and Psychology*.
- Kager, R., Boll-Avetisyan, N., & Chen, A. (2009). The language specificity of similarity avoidance constraints: Evidence from segmentation. *Paper presented at NELS 40*.
- Kager, R., & Shatzman, K. (2007). Phonological constraints in speech processing. In B. Los & M. van Koppen (Eds.), *Linguistics in the netherlands 2007* (p. 100-111). John Benjamins.
- Kager, R., & Shatzman, K. (submitted). Similarity avoidance in speech processing. *Manuscript submitted for publication*.
- Kawahara, S., Ono, H., & Sudo, K. (2006). Consonant cooccurrence restrictions in yamato japanese. *Japanese/Korean Linguistics*, *14*, 27-38.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*, B35-B42.
- Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279-312.
- Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy*, *14*, 591-612.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months

## References

- of age. *Science*, 255, 606-608.
- Leben, W. R. (1973). *Suprasegmental phonology*. Cambridge, MA: Doctoral dissertation, MIT.
- Legendre, G., Miyata, Y., & Smolensky, P. (1990). Harmonic Grammar: A formal multi-level connectionist theory of linguistic wellformedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the cognitive science society* (p. 388-395). Cambridge, MA: Lawrence Erlbaum.
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, 5, 1-54.
- Lentz, T., & Kager, R. (2009). L1 perceptive biases do not stop acquisition of L2 phonotactics. *Poster presented at the 34th Boston University Conference on Language Development (BUCLD 34)*.
- Lentz, T., & Kager, R. (in preparation). Phonotactic cues initiate lexical look-up. *Manuscript in preparation*.
- Lin, Y., & Mielke, J. (2008). Discovering place and manner features: What can be learned from acoustic and articulatory data? In *Proceedings of the 31st penn linguistics colloquium* (p. 241-254). Philadelphia, PA: University of Pennsylvania.
- MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (second ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marcus, G., Vijayan, S., Rao, S. B., & Vishton, P. (1999). Rule learning by seven-month old infants. *Science*, 283, 77-80.
- 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. (2001a). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27, 644-655.
- Mattys, S. L., & Jusczyk, P. W. (2001b). 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.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, 11, 122-134.

- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.
- McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, *17*, 207-263.
- McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica*, *43*, 84-108.
- McCarthy, J. J., & Prince, A. S. (1995). Faithfulness and reduplicative identity. In *Papers in Optimality Theory: University of Massachusetts occasional papers* (Vol. 18, p. 249-384). Amherst, MA: Graduate Linguistics Student Association.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21-46.
- McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, *45*, 103-132.
- Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive models of speech processing* (p. 236-262). Cambridge, MA: The MIT Press.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoni, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.
- Mielke, J. (2008). *The emergence of distinctive features*. Oxford, UK: Oxford University Press.
- Mitchell, T. M. (1997). *Machine learning*. New York, NY: McGraw-Hill.
- Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*, 55-71.
- Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*, 911-936.
- Myers, J., Jusczyk, P. W., Kemler Nelson, D. G., Charles-Luce, J., Woodward, A. L., & Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language*, *23*, 1-30.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, *62*, 714-719.
- Nazzi, T., Bertoni, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756-766.
- Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingua*

## References

- Linguaggio*, 2, 203-230.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.
- Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34, 191-243.
- Oller, D. K. (1973). Effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54, 1235-1247.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13-B23.
- Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53(2), 225-237.
- Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science*, 33, 999-1035.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113, 244-247.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80, 674-685.
- Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.
- Peperkamp, S., Le Calvez, R., Nadal, J.-P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, 101, B31-B41.
- Perruchet, P., & Desautly, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, 36, 1299-1305.
- Perruchet, P., & Peereeman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, 17, 97-119.
- Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39, 246-263.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge, UK: Cambridge University Press.
- Pierrehumbert, J. B. (1993). Dissimilarity in the Arabic verbal roots. In A. Schafer (Ed.), *Proceedings of the North East Linguistics Society* (Vol. 23, p. 367-381). Amherst, MA: GLSA.
- Pierrehumbert, J. B. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, 16, 691-698.

- Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics*. Cambridge, MA: The MIT Press.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
- Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Tech. Rep.). New Brunswick, NJ: Rutgers University Center for Cognitive Science, Rutgers University.
- Prince, A., & Tesar, B. (2004). Learning phonotactic distributions. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (p. 245-291). Cambridge, UK: Cambridge University Press.
- Pukui, M. K., & Elbert, S. H. (1986). *Hawaiian dictionary*. Honolulu, HI: University of Hawaii Press.
- Quené, H. (1992). Durational cues for word segmentation in dutch. *Journal of Phonetics*, 20, 331-350.
- Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, 56A, 97-115.
- Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, 125, 123-138.
- Richtsmeier, P. T., Gerken, L., & Ohala, D. (2009). Induction of phonotactics from word-types and word-tokens. In J. Chandlee, M. Franchini, S. Lord, & G.-M. Rheiner (Eds.), *Proceedings of the 33rd annual Boston University Conference on Language Development* (p. 432-443). Somerville, MA: Cascadilla Press.
- Rytting, C. A. (2004). Segment predictability as a cue in word segmentation: Application to modern Greek. In *Current themes in computational phonology and morphology: Seventh meeting of the acl special interest group on computational phonology (sigphon)* (p. 78-85). Barcelona: Association for Computational Linguistics.
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172-196.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.

## References

- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484-494.
- Scholes, R. J. (1966). *Phonotactic grammaticality*. The Hague: Mouton.
- Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development, 1*, 289-316.
- Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations*. Dordrecht, the Netherlands: Foris.
- Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*, 1-16.
- Soderstrom, M., Conwell, E., Feldman, N., & Morgan, J. (2009). The learner as statistician: three principles of computational success in language acquisition. *Developmental Science, 12*, 409-411.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111*, 1872-1891.
- Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.
- Swingle, D. (1999). Conditional probability and word discovery: A corpus analysis of speech to infants. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (p. 724-729). Mahwah, NJ: Lawrence Erlbaum Associates.
- Swingle, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86-132.
- Tesar, B., & Smolensky, P. (2000). *Learnability in Optimality Theory*. Cambridge, MA: The MIT Press.
- 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.
- Tjong Kim Sang, E. F. (1998). *Machine learning of phonotactics*. Doctoral dissertation, Rijksuniversiteit Groningen. (Groningen Dissertations in Linguistics 26).
- Toro, J. M., Nespors, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19*, 137-144.
- Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition, 16*, 178-221.
- Trubetzkoy, N. (1936). Die phonologischen Grenzsingnale. In D. Jones & D. B. Fry (Eds.), *Proceedings of the second international congress of phonetic*



- sciences. Cambridge, UK: Cambridge University Press.
- Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, 61, 846-858.
- van de Weijer, J. (1998). *Language input for word discovery*. Doctoral dissertation, Katholieke Universiteit Nijmegen. (MPI Series in Psycholinguistics 9).
- van der Lugt, A. H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*, 63(5), 811-823.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325-329.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood 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.
- Vroomen, J., Tuomainen, J., & Gelder, B. de. (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.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, 107, 238-265.
- Woodward, J. Z., & Aslin, R. N. (1990). Segmentation cues in maternal speech to infants. *Paper presented at the 7th biennial meeting of the International Conference on Infant Studies*.
- Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, 8, 451-456.



## SAMENVATTING IN HET NEDERLANDS

---

Bij het luisteren naar spraak worden luisteraars geconfronteerd met een akoestisch signaal dat geen duidelijke aanwijzingen bevat voor woordgrenzen. Om gesproken taal te kunnen verstaan, moeten luisteraars het doorlopende spraaksignaal opdelen in individuele woorden. Dit proces, spraaksegmentatie, wordt ondersteund door verschillende typen taalkundige kennis. Klemtoonpatronen, gedetailleerde akoestische informatie en fonotactische restricties dragen bij aan het herkennen van woorden in het spraaksignaal. Dit soort kennis is van cruciaal belang voor taalleerders. Eén van de belangrijkste uitdagingen voor baby's is om de woorden van de moedertaal te leren. Baby's horen echter ook een continu spraaksignaal en moeten dus strategieën ontwikkelen om woorden uit spraak te segmenteren. Ook voor baby's is het aangetoond dat ze voor het leren van woorden gebruik maken van klemtoonpatronen, akoestische informatie en fonotactische kennis. De vraag is hoe baby's dit soort kennis (die voor elke taal anders is) kunnen verwerven, nog voordat ze een woordenschat hebben opgebouwd.

Dit proefschrift gaat over het leren van fonotactische kennis die gebruikt kan worden voor spraaksegmentatie. Fonotactische restricties definiëren welke klankreeksen zijn toegestaan binnen de woorden van een taal. Nederlandse woorden bevatten bijvoorbeeld niet de reeks /pf/ (in tegenstelling tot Duitse woorden, zoals *pfeffer*). Kennis van klankreeksen is nuttig voor spraaksegmentatie, omdat het aangeeft waar zich mogelijke woordgrenzen bevinden in het spraaksignaal. Bij het horen van de continue zin /dɛlɑmpfil/ (*de lamp viel*) geeft de restrictie \*pf aan dat er een woordgrens zit tussen /lamp/ en /fil/. (Een Duitse luisteraar daarentegen zou wellicht geneigd zijn om /pfil/ als één woord waar te nemen.)

De stelling die in dit proefschrift verdedigd wordt, is dat fonotactische restricties geleerd worden uit het continue spraaksignaal. Dit gebeurt met behulp van twee leermechanismen: statistisch leren en generalisatie. De voorstelling is dat de fonotactische kennis die op deze manier geleerd wordt, gebruikt kan worden voor het ontdekken van woordgrenzen in het spraaksignaal. Deze veronderstelde methode, Spraak-Gebaseerd Leren (SGL), wordt onderzocht middels een combinatie van (i) computermodellering, om formeel te beschrijven hoe fonotactische kennis geleerd wordt, (ii) computersimulaties, om te testen of de geïnduceerde fonotactische kennis nuttig is voor spraaksegmentatie, (iii) simulaties van menselijke data, om te testen of het computermodel menselijk segmentatiegedrag kan verklaren en (iv) experimenten met kunstmatige talen, om te toetsen of menselijke leerders (volwassenen) in staat zijn om nieuwe fonotactische restricties te leren uit een continu spraaksignaal. Door gebruik te maken van zowel computermodellering, als psycholinguïstische experimenten, tracht dit proefschrift een formele

verklaring te geven voor het induceren van fonotactiek voor spraaksegmentatie, die wordt ondersteund door bevindingen met zowel computationele als menselijke leerders.

Dit proefschrift richt zich specifiek op het verbinden van eerder onderzoek op het gebied van het modelleren van spraaksegmentatie en het modelleren van fonotactisch leren. Beide typen modellen richten zich op het leren van klankreeksen, maar maken sterk verschillende aannames over de input van het fonotactische leerproces en over het niveau van abstractie waarop fonotactische restricties gerepresenteerd worden. Segmentatiemodellen gaan er van uit dat klankreeksen geleerd worden *uit continue spraak*. Fonotactische leermodellen daarentegen gaan er van uit dat kennis van klankreeksen verworven wordt *uit het lexicon*. Segmentatiemodellen gaan er bovendien van uit dat fonotactische restricties verwijzen naar *specifieke fonologische segmenten*, terwijl fonotactische leermodellen er doorgaans van uit gaan dat deze restricties verwijzen naar *natuurlijke klassen van segmenten* (die gedefiniëerd worden door fonologische *features*). Psycholinguïstische studies met baby's maken het aannemelijk dat fonotactische restricties op het niveau van natuurlijke klassen inderdaad geleerd worden, al voordat het lexicon zich volledig ontwikkeld heeft. Deze bevindingen leiden tot de specifieke hypothese dat abstracte restricties (dat wil zeggen, restricties met betrekking tot natuurlijke klassen) geleerd worden uit continue spraak in een vroeg stadium van fonologische ontwikkeling. Dit proefschrift benadert het probleem van fonotactisch leren door de input-aanname van segmentatiemodellen (namelijk, continue spraak) te combineren met de representatie-aanname van fonotactische leermodellen (namelijk, abstracte restricties).

In Hoofdstuk 2 wordt een nieuw computationeel model gepresenteerd: STAGE. Dit model implementeert de SGL-hypothese. Het hoofdstuk laat zien dat fonotactische restricties geleerd kunnen worden uit continue spraak door gebruik te maken van leermechanismen waarvan is aangetoond dat baby's ze tot hun beschikking hebben: statistisch leren en generalisatie. Het model leert twee typen fonotactische restricties: (i) 'markedness constraints' ( $*xy$ ), die aangeven welke reeksen niet voor mogen komen binnen woorden (en dus waarschijnlijk woordgrenzen bevatten), (ii) 'contiguity constraints' (CONTIG-IO( $xy$ )), die aangeven welke reeksen juist wel binnen woorden mogen voorkomen (en dus waarschijnlijk geen woordgrenzen bevatten). Eventuele conflicten tussen deze restricties worden opgelost met behulp van het OT segmentatiemodel (gebaseerd op principes uit Optimality Theory).

Hoofdstuk 3 laat de potentiële bruikbaarheid zien van de SGL-hypothese door middel van computersimulaties. Het hoofdstuk richt zich specifiek op de toegevoegde waarde van generalisatie. De voornaamste bevinding is dat STAGE beter in staat is om woordgrenzen te detecteren in continue spraak

dan puur statistische modellen. Dit geeft aan dat generalisatie, zoals gebruikt door STAGE, mogelijk ook gebruikt wordt door menselijke leerders. Het hoofdstuk geeft computationele evidentie voor de SGL-hypothese, doordat het laat zien dat fonotactisch leren uit continue spraak bijdraagt aan betere woordherkenning.

In Hoofdstuk 4 wordt gekeken of STAGE een verklaring kan geven voor de leerbaarheid van een fonotactische restrictie uit de theoretische fonologie. OCP-PLACE zegt dat consonanten met dezelfde plaats van articulatie niet naast elkaar voor mogen komen. Dit hoofdstuk bekijkt of de set van restricties die geïnduceerd wordt door STAGE sporen van OCP-PLACE bevat. Daarnaast wordt in dit hoofdstuk bekeken of het model het effect kan verklaren dat OCP-PLACE heeft op menselijk segmentatiegedrag (zoals aangetoond door Boll-Avetisyan & Kager, 2008). De bevinding is dat STAGE inderdaad restricties leert die lijken op OCP-PLACE. Bovendien is het model beter in staat om menselijke segmentatie te verklaren dan puur statistische modellen en modellen die gebaseerd zijn op een categorische interpretatie van OCP-PLACE. Het hoofdstuk laat zien dat de combinatie van leermechanismen, zoals gebruikt door STAGE, in staat is om data over menselijke segmentatie te verklaren.

Hoofdstuk 5 onderzoekt of menselijke leerders in staat zijn om nieuwe fonotactische restricties te leren uit continue spraak. Dit wordt gedaan door volwassen proefpersonen te laten luisteren naar continue spraak uit een kunstmatige taal. Vervolgens wordt gekeken welke woorden de proefpersonen menen te hebben gehoord. Deze experimenten testen of de SGL benadering van STAGE psychologisch plausibel is. De bevinding dit hoofdstuk is dat menselijke leerders in ieder geval restricties op het niveau van specifieke segmenten kunnen leren uit continue spraak. Er werd geen evidentie gevonden voor het leren van fonotactische generalisaties. Het leren van generalisaties uit continue spraak zal in de toekomst nog nader onderzocht moeten worden.

Het proefschrift laat zien dat fonotactische kennis voor spraaksegmentatie geleerd kan worden met een combinatie van mechanismen waarvan is aangetoond dat menselijke taalleerders ze tot hun beschikking hebben. Het voorgestelde model leert fonotactische restricties met verschillende abstractieniveaus. Deze restricties kunnen vervolgens gebruikt worden om woordgrenzen te herkennen in het spraaksignaal. De combinatie van statistisch leren en generalisatie in het model geeft een betere verklaring van spraaksegmentatie dan modellen die alleen gebruik maken van statistisch leren. Dit resultaat werd gevonden voor zowel computationele, als menselijke leerders. Daarnaast werd aangetoond dat menselijke leerders nieuwe fonotactische restricties kunnen leren uit een continu spraaksignaal. Het vergelijken van bevindingen met computationele en menselijke leerders stelt ons in staat om leermodellen te ontwikkelen die formeel expliciet zijn en die bovendien psychologisch plausibel zijn. Het proefschrift draagt hierdoor bij aan een beter begrip van de processen die een rol spelen bij taalverwerving.



## CURRICULUM VITAE

---

Frans Adriaans was born on March 14, 1981 in Ooststellingwerf, the Netherlands. He spent most of his childhood in Amersfoort, where he graduated from the Stedelijk Gymnasium Johan van Oldenbarnevelt in 1999. In that same year, he enrolled in the Artificial Intelligence program at the University of Amsterdam. After obtaining his *propedeuse* and *kandidaats* diplomas, he pursued the Language & Speech track within the Artificial Intelligence master's program. He conducted research for his master's thesis with the Information and Language Processing Systems group. The thesis focused on evaluating computational techniques for bridging the spelling gap between modern and 17th century Dutch, with the purpose of improving the retrieval of historic documents. He obtained his M.Sc. degree in 2005.

Frans Adriaans was a research assistant at Utrecht University in 2005 and 2006. During that period, he worked on developing computational techniques for the analysis of transcribed speech corpora. He subsequently started his PhD research at the Utrecht Institute of Linguistics OTS (UiL OTS). This dissertation is the result of research conducted at UiL OTS between 2006 and 2010.