Master Thesis

The Effect of Lexical Distribution and Similarity Avoidance in Speech Segmentation

Thesis for the Master Language and Development (Taal en Ontwikkeling),
at Utrecht University

Author:
Bart Penning de Vries
0374598

Thesis Supervisors:
Prof. dr. René Kager
Natalie Boll-Avetisyan, MA

Acknowledgements

Both during the writing of this thesis, and waiting for participants in the lab, I created a special bond with the coffee machine in the cafeteria. I do not know who installed it, but I am grateful.

On a less material level, and more related to the subject of linguistics, I would like to thank my supervisors René Kager and Natalie Boll-Avetisyan. In the first place for making it possible to write this thesis and conduct my own experiments, but also for their suggestions, support and expert guidance; and their time, of course. I would also like to thank Phonotactools, a programme by Frans Adriaans, which has proven to be very useful indeed, and has run for many hours to come up with explanations.

For technical support in the lab, thanks to Theo Veenker and Warmolt Tonckens; and for helping to recruit participants in less than 12 hours, thanks to Iris Mulders.

And many thanks to the home front, obviously.

In return, I hope this thesis reflects a high level of enthusiasm and curiosity for phonology research, and, clearly, sheer intelligence.

Table of Contents

Introduction

To understand spoken language, a listener must be able to segment the speech stream into meaningful units. Compared to (most) written text, where words are separated by blank spaces, word boundaries in speech are elusive. Spoken language does not seem to have clear, uniform cues to signal word boundaries: when analysing the acoustic signal of spoken sentences, there is no pause, or silence, reliably available to indicate a word boundary (Cole & Jakimik 1980). Listeners must therefore be making use of other information contained in speech to identify words.

Word boundaries can be found by attending to phonological cues, for instance, they may be indicated by intonation, or the location of word stress, duration of syllables, or the quality of vowels. However, these cues are not reliably present in speech. As a result, there must logically be a more general mechanism for segmentation, one which will allow a word to be recognised when there happen to be no acoustic cues signalling that word's boundaries.

Beside phonological cues, sub-lexical cues are proven to be used and useful for word segmentation. Sub-lexical cues are cues that arise from patterns in the lexicon. To illustrate, consider this orthographic example: "people don't talk like this, theytalklikethis" (Bryson 1990: 83). In the continuous string that imitates continuous speech there is no indication where one word ends, and the other begins. Still, English readers will be able to identify words in the string: they do not rely solely on the blank spaces between words. Apparently people can recognise the patterns of letters that belong together and group them into a word. In speech, a similar mechanism may be used, though instead of recognising patterns of letters, phoneme patterns are detected.

There is evidence that sub-lexical cues can be detected and used for speech segmentation (e.g. Saffran, Newport & Aslin 1996). They found that humans can detect patterns of co-occurring syllables and use them to find words in continuous speech stream. In a simplified example: in an unknown artificial language, if *ba* occurred together more frequently with *pi* than with *tu*, then their test subjects considered *ba-pi* to belong together as opposed to *ba-tu*. Further studies show that humans quickly acquire phonemic patterns from artificial speech, and can apply them to segment speech (both infants, Saffran & Thiessen 2003, and adults Onishi et al. 2002). The sensitivity that humans exhibit throughout their lives for the patterns which are present in spoken language suggest that they are important for language learning and language use.

A study by McQueen (1998) shows that native speakers are aware of language specific phoneme patterns, and that these patterns facilitate speech segmentation. For instance, for Dutch speakers, the subconscious knowledge that [mr] cannot be at the beginning of a syllable, will cause them to expect a word boundary between those phonemes. As a result, a Dutch speaker will spot the word *rem* 'brake' quicker in [mrem] than in [prem], because [pr] is possible as a word beginning. Moreover, native speakers have been shown to be sensitive to language specific statistics (Vitevitch & Luce 2005), since they are quicker to recognise frequent patterns from their language than infrequent patterns. This statistical knowledge also influences word segmentation (Mattys & Jusczyk 2001). In Chapter 1, this will be discussed in greater detail. There is evidence, therefore, that the frequency in which patterns occur in the lexicon influence speech segmentation.

Interestingly, not all statistical patterns in language are picked up with similar ease (e.g. Peña et al. 2002, Newport & Aslin 2004, Onnis 2005). This is not surprising, since any string of speech contains an enormous amount of possible statistics to keep track of; an example of irrelevant information would be how many times the third phoneme of a word is a /p/. Clearly, not all statistical information available in speech is relevant for speech segmentation. It is necessary, therefore, that learners are able to filter out only the statistical information that is relevant for language. Various studies show that learners track only statistics over specific linguistic elements (e.g. Newport & Aslin 2004, Onnis 2005, Toro et al. 2008). However, if learners are sensitive to specific information in language input, then the learning process itself may shape language. Saffran (2002) argues that if there are constraints on statistical learning, certain statistics in the input speech may be easier to process. If this reasoning is applied to statistical learning, the constraints found in statistical learning contribute to the forming of patterns in language, and these patterns may be detected in the lexicon.

Since the frequency of lexical patterns in language is found to influence segmentation, it becomes difficult to tell apart the difference of a constraint on statistical learning, and the constraint based on lexical frequency. This thesis intends to do exactly that: to compare the influence of low-lexical frequency with the influence of a constraint on statistical learning. In this thesis, segmentation of an infrequent pattern is compared to segmentation of an infrequent pattern that is also subject to a phonotactic constraint.

In several languages, there is evidence of a phonological constraint on similarity. McCarthy (1986) finds that in Arabic, sequences of consonants with a similar place of articulation are under-represented: they are less frequent in the lexicon than would be expected. This gives rise to assume a constraint restricting the occurrence of similar consonants in sequence. In subsequent studies, the effect of this constraint has been noted in various other languages (e.g. Arabic (McCarthy 1986), Hebrew (Berent & Shimron 2003), English (Coetzee 2005), Japanese (Kawahara et al. 2006), suggesting that the constraint may be a universal constraint banning similarity, but at least is not language-specific, and thus is not based on lexical frequency.

The similarity constraint is also found in Dutch (Shatzman & Kager 2007, Boll-Avetisyan & Kager 2008, Kager & Shatzman (in progress)). These studies show that the Dutch lexicon has an under-representation for similar consonants, and that native speakers of Dutch avoid sequences of similar consonants. Arguably, the effects of the similarity constraint that these studies find are attributable to the subjects' subconscious knowledge of Dutch lexical distribution. This argument can be backed by the evidence of human capacity for statistical and phonotactic learning and their ability to apply that to artificial language learning.

The question that this thesis addresses, therefore, is whether there is a difference between the influence of lexical distribution, and the influence of the constraint on similarity for speech segmentation. Three experiments presented in this thesis attempt to tease apart the effect of the similarity constraint and the effect of under-representation in the lexicon. Learners are assumed to focus on specific elements in the speech input which enables them to detect regularities, and to extract rules more efficiently. As a result, the phonological features of the phonemes in the language may influence statistical learning. A final experiment examines how the features of the consonants interact and play a role in statistical learning.

This thesis is set up in five chapters. Chapter 1 is an overview of relevant literature on statistical learning, artificial language learning, and the constraint on similarity, with a focus on the similarity constraint in Dutch. The prediction and hypothesis for the experiments is stated at the end of Chapter 1. Chapter 2 describes the first experiment that intends to determine whether lexical distribution of consonant pairs can predict segmentation. Chapter 3 deals with the second experiment, which examines the effect of a constraint on consonant similarity for segmentation behaviour; crucially this artificial language has the same lexical distribution characteristics as the language in Experiment 1, to enable comparison of segmentation behaviour in both languages. Experiment 2 is minimally different from the first experiment but includes a homorganic consonant pair (labials), this enables a comparison of a phonotactic constraint on similarity with lexical under-representation. Chapter 4 tests a hypothesis that arises from the previous two experiments concerning the conditions for similarity constraints. The third experiment further examines the phonotactic constraint, but uses a different set of labial consonants than Experiment 2. Chapter 5 combines the results of the three experiments, and presents a general discussion with suggestions for further research.

1. Literature Overview

This chapter presents an overview of the literature on speech segmentation, focusing on statistical learning for language (1.1), linguistic constraints on statistical learning (1.2), phonotactics (1.3), and the similarity constraint (1.4). Since the experiments in this thesis are artificial language learning experiments, (1.5) reviews relevant artificial language learning studies. The hypothesis and predictions are stated in (1.6).

1.1. Statistical learning

Statistical distributions in the lexicon can provide a reliable cue for speech segmentation. This is illustrated very nicely in an example by Johnson & Jusczyk (2001: 550):

> The two-word string *prettybaby* consists of four syllables: *pre*, *ty*, *bay*, and *by*. The first two syllables (*pre* and *ty*) consistently appear together because they form a word. Likewise, the latter two syllables (*bay* and *by*) also tend to occur together. However, the second and third syllables (*ty* and *bay*) occur together relatively rarely. Across a corpus of English, the syllable *ty* follows the syllable *pre* more frequently than the syllable *bay* follows the syllable *ty*, because many different words can follow the word *pretty* (e.g., *pretty flower*), but only a few syllables can follow *pre*. This greater predictability of word internal syllables than syllables spanning word boundaries may be helpful in discovering word boundaries.

In this fashion, the likelihood of a syllable sequence can predict a word boundary, or that the syllables together form a word. This knowledge may be useful for speech segmentation.

The likelihood of a syllable sequence can be expressed by transitional probability (TP).Transitional probability is computed as :

(1) TP = frequency of pair XY / frequency of X
(Saffran, Newport & Aslin 1996: 610).

From (1) follows that a high TP indicates that X strongly predicts that Y will occur. A low TP indicates that, given X, XY is an infrequent combination, and that there may be a word boundary between the syllables X and Y.

Saffran, Newport & Aslin (1996) show that adult speakers of a language can use distributional information as given by TPs to segment speech. They exposed adult learners to an artificial language that contained no cues for speech segmentation other than distributional properties, computed as transitional probability (TP). In (2), the TPs predict that the consonant (C)-vowel(V) syllables $CV_2CV_3CV_4$ belong together and make up a 'word' in the artificial language. The low TPs between CV_1-CV_2, and CV_4-CV_5 predict a word boundary.

(2) Syllable: $CV_1 \rightarrow CV_2 \rightarrow CV_3 \rightarrow CV_4 \rightarrow CV_5$
TP: 0.1-0.2 0.3-1.0 0.3-1.0 0.1-0.2

After listening to this language for a total of twenty-one minutes, subjects could identify words; that is, distinguish part-words (e.g $CV_1CV_2CV_4$) and non-words (e.g. ($CV_xCV_yCV_z$) from words. Subjects placed word boundaries where the TP was lowest, and, in this way, segmented continuous speech using only transitional probabilities between syllables.

However, this ability to keep track of statistical information in the input signal is not particular to humans. When tamarin monkeys were exposed to the same artificial language as that of Saffran et al. (1996), the monkeys were also seen to distinguish words from part-words (Hauser, Newport & Aslin (2000).

Additionally, Saffran et al. (1999) show that the ability to track statistics is not limited to the linguistic domain. They replicated the setup of Saffran, Newport & Aslin (1996), but replaced the syllables by tones. In this fashion, tone sequences made up the equivalent of a word. Again, after listening to a continuous sound stimulus that contained only distributional information, infant- (8 months) and adult listeners were found to be able to use the distributional information to discriminate the 'words' from non-'words' and part-'words'. Other studies show that adult subjects can also use distributional information (adjacent dependencies) for visual stimuli (Asaad (1998), and Kirkham (2002) quoted in: Onnis et al. 2005: 234).

As a result, the ability to track statistics appears to be domain-general and not specifically linguistic; not even specifically human. However, this general capacity applies to the input signal in a specifically linguistic manner: in this way filtering out only the statistical information that is relevant to language, and language learning. The next section looks at language learning of linguistically relevant statistics.

1.2. Linguistic constraints on statistical learning

Though humans are clearly capable statistical learners, logically, there must also be some filter that stipulates what information is useful, and what is not. For instance, keeping track of the average number of vowels in a sentence does not directly seem to be a relevant statistic. In language input, there are an infinite number of possible statistical computations; clearly, this is a problem for a learner with a finite processing capacity. To avoid this problem, statistical learning must be constrained in such a way that learners only perform a subset of the logically possible computations (Saffran 2002).

Learners arrive at a correct representation of language from language input, and do not learn structures that are not present in the natural language. Saffran (2002) argues that the learning mechanisms that apply to language, such as statistical learning, may be constrained to preferentially learn certain types of patterns. As evidence, she shows that adults, and infants, learn words better in an artificial language if it contains dependencies that are also found in natural languages, as opposed to systems in which there was no natural dependency[1]. As a result, she reasons that "[i]f the structures that are most learnable are also those that recur cross-linguistically, then the similarity of human languages may have roots in the learning process itself: constraints on language learning may shape the structure of natural languages" (Saffran 2002: 173). If this reasoning is applied to statistical learning, the constraints found in statistical learning contribute to the forming of patterns in language, and these patterns may be detected in the lexicon.

An example of a linguistic constraint on statistical learning is provided by Bonatti et al. (2005) who find that statistics in a linguistic input are calculated only over specific linguistic elements. Bonatti et al. (2005) found that French subjects could learn statistic dependencies between consonants to segment continuous speech, but were unable to learn dependencies between vowels[2]. Moreover, Toro et al. (2008) show that there are different

---

[1] e.g. in a phrase *AP*: a natural dependency: if there is an A, then there is the possibility of a D; opposed to phrase *AP* containing either an A or a D. In summary: AP= [A + (D)] versus AP = [(A) + (D)]

[2] Remarkably, a reverse distinction was found for non-human primates: word segmentation by tamarin monkeys depended on regularities between vowels, not consonants (Newport Hauser et al;. 2004).

roles for consonants and vowels in speech segmentation. Vowels are more useful to learn rule generalizations, while consonants are used for word segmentation.

In the studies by Saffran, Newport & Aslin (1996), Bonatti (2005), and Toro et. al. (2008), word segmentation relied on adjacent dependency of the syllables. Since natural language contains many non-adjacent dependencies, Aslin & Newport (2004) tested whether statistics of non-adjacent dependencies between syllables could be used for segmentation. They created a continuous artificial language, containing three-syllabic words as seen in (3). The only information available for word segmentation is that CV_1 is always followed by CV_3 (TP = 1.0).

(3) …CV_1 [CV_{2a}] CV_3… $CV_1 \rightarrow CV_3$: TP: 1.0
[CV_{2b}]
[CV_{2c}]
[CV_{2d}]

The test subjects could not identify words on the basis of non-adjacent dependencies between syllables[3]. However, they could detect non-adjacent regularities among segments, both with consonant and vowel regularities, as seen in (4):

(4) a. consonant regularity: $C_1[V_1]$ C_2 $[V_3]$ $C_3[V_5]$ C_x->C_y: TP= 1.0
[V_2] [V_4] [V_6]

b. vowel regularity: $[C_1]V_1$ $[C_3]V_2$ $[C_5]V_3$ V_x->V_y: TP= 1.0
[C_2] [C_4] [C_6]

Both the regularities of vowels and of consonants could be used for segmentation, though subjects performed better on the consonants. Aslin & Newport (2004: 138) suggest a correlation with natural languages: “nonadjacent regularities among syllables is not common in languages around the world, whereas nonadjacent regularities among segments is common in languages (e.g. it is found in Hebrew and Arabic)”. Taken together with the studies of

---

[3] In a similar experiment, Peña et al. (2002) found that adults could use regularities between non-adjacent syllables, but this depended crucially on subliminal pauses between words. Additionally, Onnis et al. (2005) show that their study contained a phonological confound –the perceived ability of the subjects to detect regularities was seen to correlate with a phonological preference (preference for words with plosives in initial and final position).

Bonatti et al. (2005) and Toro (2008), this suggests that word segmentation relies on segments, where distributional information of non-adjacent phonemes is used, rather than on syllables.

Onnis et al. (2005) examined the learning of non-adjacent dependencies more closely. They replicated the study by Peña et al. (2002) which claims that non-adjacent syllable dependencies can be learned, and showed that when the dependency is removed (such that all TPs are 0.33), there is still evidence of word learning among the participants. Apparently, the non-adjacent dependency between syllables was not driving word segmentation; segmentation was based on phonological properties of the consonants in the language.

A follow up experiment determined how the phonological properties interact with non-adjacent statistics. They constructed an artificial language similar to the first language: three $A_i\_B_i$ pairs (TP=1.0) with three intervening syllables (X) to make up 27 words of an A_iXB_i syllable frame. In this way, syllable A_i is always followed by syllable B_i, with three possible X syllables in the middle. In the language, the A-syllables begin with a continuant, and the X- and B-syllables with a plosive, such that ‘words’ begin with a continuant, and part-‘words’ with a plosive. The results showed a preference for part-words: the non-adjacent regularity was not detected, or was overruled by a phonological preference for a plosive onset. However, the results showed a stronger preference for part-words of the structure plosive-continuant-plosive, than for part-words, structured as plosive-plosive-continuant. Onnis et al. suggested that, in the plosive-continuant-plosive words, the plosives are distinct because of the intervening continuant syllables and therefore may be perceptually more salient.

This was tested in a similar A_iXB_i structure where continuants are in the A and B positions, and plosives in the X positions, see (5). The continuants are fricatives and the $A_i\_B_i$ pairs have different place of articulation.

5).

A	**X**	**B**
[cont.]	[plos.]	[cont.]
ze	pu	vo
thi	ta	shu
fo	gi	sa

The words in (5), with a non-adjacent $A_i\_B_i$-structure, were learned: the continuant-plosive-continuant structure was preferred over part words. Consonants that have a dissimilar manner of articulation seem to be more distinctive, causing phonologically similar consonants to be more salient. Learning of a non-adjacent dependency is possible, therefore, but only under certain circumstances where the first and the third syllable are phonologically similar.

However, the effect of phonological similarity cannot drive segmentation on its own. No learning occurred when the language in (5) did not contain a non-adjacent dependency. (There was learning of a plosive-continuant-plosive structure without a non-adjacent dependency, but this was shown to correlate with language specific preferences of phonological structure). As a result, Onnis et al. (2005) concluded that phonological similarity can influence segmentation performance, and this phonological similarity is seen to interact with statistical information and phonological preference in an additive way. Additionally, the plosives and continuants are grouped together into distinctive classes, as shown by the phonological similarity effect; this indicates a level of abstract representation of the consonants in natural distinctive classes.

As shown by Aslin & Newport (2003), adults are able to detect statistical information of non-adjacent consonants across a vowel (C_xVC_y), but not across non-adjacent syllables (CV_xCVCV_y). Onnis et al. (2005), however, found that learning of non-adjacent syllables is possible, but that this is highly dependent on phonological properties of the consonants. Additionally, the studies by Onnis et al. (2005) Bonatti et al. (2005), and Toro et al. (2008) suggest that non-adjacent consonants play an important role for word segmentation. The phonological similarity effect found in Onnis et al (2005) suggest that learners group consonants on basis of similar phonological features. In this thesis, another possible learning bias is assumed: statistical learning may be constrained by similarity, which may explain why dependencies between similar elements are more salient to the learner than those between non-similar ones (this is explained in more detail in the hypothesis in 1.6) The next section examines in more detail how phoneme patterns interact for word segmentation.

1.3. Phonotactics

Languages differ in the phoneme patterns that can be found in the lexicon. For instance, *ng* /ŋ/ is not found in the onset of a word in English, but it is found in onsets in Vietnamese (example from Onishi 2002). Similarly, /kn/ is a Dutch a word onset, but is not an onset in English. The combinatorial rules that phoneme patterns are subject to are called phonotactics.

Jusczyk et al. (1993) found that infants become aware of the phonotactics of their language very early in life. English and Dutch infants of 9 months-old could distinguish words concurring with their native phonotactics from words with non-native phonotactics; infants 6 months old could not distinguish the native from non-native words (they were found to be sensitive to native prosody): apparently, phonotactic patterns are acquired somewhere from 6 months onwards. Infants are also able to acquire new phonotactic regularities very quickly. Saffran & Thiessen (2003) showed that 9-month-old infants can (after 2 min of exposure) detect phonological regularities; Onishi et al. (2002) found that this capability is also still present in adults. Further evidence of learned phonotactics comes from Seidl & Buckley (2005): infants 9 months old could learn arbitrary phonological patterns; phonotactic learning appeared to be unconstrained. Furthermore, Cristià & Seidl show that cross-linguistic frequency does not predict ease of acquisition, but does provide evidence for ease of acquisition being correlated with phonological complexity (2007: 11). This notion ties in with Saffran's (2002) hypothesis that constraints on learning mechanisms may shape language (acquisition). Apparently, humans are sensitive to phonotactic regularities throughout their lives, and apply their phonotactic knowledge to segment speech. Since a central problem of language acquisition is how infants first learn to find words in spoken language, the early sensitivity to native phonotactics may prove to be very useful for basic word boundary detection.

Phonotactic regularities can facilitate speech segmentation by predicting where a syllable or word boundary is likely: if a phoneme sequence is phonotactically illegal in a language, a native speaker will assume a boundary in between the phonemes. For example, in Dutch [mr] is not permitted within the same syllable, a Dutch speaker hearing [mr], therefore, will expect a syllable boundary between the phonemes. McQueen (1998) provides evidence that this knowledge is used for word segmentation. In a word-spotting experiment, subjects responded quicker if a word was offset by phonotactic cues. Dutch subjects were presented aurally with a target Dutch word (e.g. *pil* 'pill') embedded in a non-word context, such as [pil.vrem] (the dot indicates a syllable boundary). In [pil.vrem] the word boundary is aligned with a phonotactic boundary, because Dutch does not allow [lv] syllable internally, but does allow [vr] as syllable onset (e.g. in *vrij* 'free'). In [pilm.rem] the word boundary is misaligned with a phonotactic boundary, because Dutch does not allow [mr] as syllable onset, but does allow [lm] as possible coda (e.g. in *palm* 'palm'). Dutch subjects were quicker to spot words when they were aligned with phonotactic boundaries. As a result, subjects appear to recognise the legality of phonotactics, and use phonotactics to indicate word or syllable boundaries.

1.3.1. Syllable-based versus sequential phonotactics

In the experiment by McQueen, phonotactics signalled the location of a word boundary. However, the location of a word boundary coincides with the location of a syllable boundary. This observation raises the question whether phonotactics signal syllable boundaries, or that syllable boundaries influence phonotactics: in other words, whether phonotactics are syllable-based. To test for the influence of syllable structure, McQueen presented Dutch subjects with words only where the syllable boundary was not phonotactically determined, but pronounced by a speaker (e.g. *pilkrem* allows *pil.krem* and *pilk.rem*: both [lk] and [kr] are phonotactically legal sequences). The results do not show an influence of syllable information. The target word (e.g. *pil* 'pill') was detected equally quickly, regardless of syllable boundary. Syllable boundary information has been shown to be used for segmentation (e.g. in Dutch: Zwitserlood et al. 1993), but, as McQueen suggests, this is only possible if the boundaries are made unambiguous by phonotactics: in his experiment the syllable boundaries were not clearly provided by phonotactics. It has been observed that not all words are syllabified unambiguously, for example, Dutch (and English) native speakers show a varying syllabification of the word *letter* (McQueen 1998: 38). This suggests that perhaps phonotactics signal syllable boundaries.

In the literature on phonotactics two views are found. Phonotactics are considered to apply sequentially (e.g. a ban on consonant clusters: *CC), or syllable-based (e.g. no consonant clusters in onset). If phonotactics are assumed to signal syllable boundaries, then phonotactics cannot operate on syllable level, and phonotactics must be assumed to apply sequentially. On the other hand, phonotactics are syllable based, then syllables must be detected before phonotactics can apply.

A syllable-based explanation allows for a simple and unifying approach to phonotactics, and is clearly present in rhythmic processes such as metrical stress, and intuitions on string division (Coté 2002:15). However, as McQueen argues, syllabification may not occur reliably if the syllable boundaries are not clearly provided by phonotactics. Coté (2002), and Steriade (1999) argue that the syllabic approach in phonotactics is unnecessary, since the patterns that are compatible with a syllabic analysis, can equally be rendered in a sequential account that makes no use of syllable well-formedness. Additionally, Coté argues that the syllabic approach cannot account for all occurrences of epenthesis and deletion. Processes of assimilation to certain neighbouring phonemes, such as /mp/ in *te*n p*ounds* also favour a sequential approach. An example of a phonotactic constraint that clearly does not operate on syllable level is seen in Vroomen,Tuomainen, de Gelder (1998). In a word-spotting experiment, native speakers of Finnish were shown to be sensitive to vowel harmony in words. Vowel harmony requires that within a word, all vowels must be either back or front. In this way, the constraint that applies cannot be syllable-based, since it applies to the whole word. Another example is the sequential obligatory contour principle (OCP), a constraint that restricts the occurrence of similar consonants. McCarthy (1986) assumes that consonants and vowels are treated separately in segmentation: consonants and vowels are represented on different tiers[4], and in this way, the constraint restricts similar consonants on the consonant tier (this is discussed in 1.4). This approach abstracts away from the syllable-based, and applies at sequential level on the consonant tier. However, the effect of this constraint often crosses syllable boundaries (e.g. as a result of OCP, the final /s/ causes the suffix [-s] to be altered in *pass-passes* /pæsɪz/).

Clearly, in phonotactics, some processes are limited to the syllable, but not all. How phonotactics interact with syllable boundaries is still subject to debate, but will not be dealt with here. However, it is clear that there are phonotactic processes that are not syllable based, and operate on a sequential level. The constraint under investigation in this thesis is a sequential constraint, where the constraint is not syllable-based, but is based on distance: the level of adjacency on the consonant tier.

1.3.2. Probabilistic versus categorical phonotactics

Phonotactics can be divided neatly into three categories, as suggested by Chomsky & Halle (1968):

(6)	existent	brick
	non-existent but well-formed	blick; and
	non-existent but ill-formed	bnick

(example from Hayes & Wilson 2008: 381)

With the help of a language-specific phonotactic grammar, learners should be able to separate the non-existent well-formed from the non-existent ill-formed. Although this is confirmed by

---

[4] To explain their finding of the different roles in speech segmentation for vowels and consonants, Bonatti et al. (2005) also suggest that they are represented on different tiers.

perception studies (in e.g. McQueen 1998, and for newly learned phonotactic patterns, e.g. Onishi 2002), in word judgement experiments, native speakers show a finer grained distinction: between the non-existent words, there are degrees of acceptability. As an example by Albright (2005) shows, non-words can be rated in levels of acceptability:

(7) “How good would . . . be as a word of English?”

Best	*stin* [stIn] , *mip* [mIp]
	blick [blIk], *skell* [skEl]
Intermediate	*blafe* [bleIf], *smy* [smaI]
	bwip [bwIp], *smum* [sm2m]
	dlap [dlæp], *mrock* [mrak]
Worst	*bzarshk* [bzarSk], *shöb* [Soeb]

These degrees of the acceptability of non-words correlate with their distributional properties (non-word judgement: Coleman & Pierrehumbert 1997, Frisch & Zawaydeh 2001, Luce & Large 2001, Hayes & Wilson 2008; spoken repetition: Vitevitch & Luce 2005). Mattys & Jusczyk (2001) showed that infants segment speech better when target words are offset by phonotactic cues with a low frequency in the lexicon than when the target words are offset by phonotactic cues with a high frequency[5].

The frequency with which phonemes occur next to each other in natural speech sequences is known as probabilistic phonotactics. It seems that the more often a learner has encountered types of a phonotactic sequence, the more acceptable that sequence will be rated. On the other end, if a learner has never heard a particular sequence, it will be judged as less acceptable. In this fashion, the probability of phonotactic sequences determines their acceptability. A categorical distinction between legal and illegal seems to be insufficient to explain the degrees of acceptability found in non-word judgements. As a result, a different constraint must be considered to explain the gradient rating of well-formedness of non-words.

---

[5] this study is similar to that of McQueen (1998). However, the crucial difference with the study by McQueen (1998) is that McQueen uses legal and illegal phonotactic cues, where Mattys & Jusczyk use good (low frequency) and bad (high frequency sequences) cues.

1.3.3. Gradient versus categorical phonotactic constraints

Traditional generative grammar aims at classifying any particular form as grammatical or ungrammatical. However, it cannot explain degrees of acceptability. If phonotactics are grammatical or ungrammatical on the basis of their lexical distribution, then the grammar should exclude patterns that are very infrequent in the lexicon. Patterns that are statistically under-represented are assumed to be phonologically illegal, even though a few exceptional forms may be found. Patterns of phonotactic probability are seen to correlate with native speaker judgements of degrees of well-formedness, but, it does not categorically separate legal from illegal patterns. In the literature, there are two relevant views to be distinguished that can account for the gradient nature of phonotactic probability:

- (i) the gradient distribution in the lexicon is the result of a single constraint that applies in a gradient manner. In a model, Frisch et al. (2004) show that the constraint that best approximates Arabic lexical data is a single gradient constraint that is quantitatively sensitive to violations of various degrees, such that forms that violate the constraint to a lesser degree are more frequent than forms that violate the constraint to a greater degree.
- (ii) the gradient distribution is the result of categorical constraints whose interaction is stochastic (e.g. Boersma & Hayes 2001; Coetzee 2008). This approach is based on Optimality Theory, where constraints are ranked according to the language input. Crucially, constraints can be violated in order to satisfy a higher constraint. Gradient outputs can thus be explained by violation of one or more constraints.

The difficulty for this view is explaining which constraints should apply, and how their ranking is achieved. In standard OT a universal set of constraints is assumed, which is re-ranked according to language input. In Boersma & Hayes' (2001) Gradual Learning Algorithm, categorical rankings emerge based on evidence from the language input: but these categories can overlap. When constraints are close together and overlap, alternating forms may surface as 'winner': this may account for degrees of acceptability. Another option is found in Coetzee & Pater (2008), where constraints are not ranked, but weighted. In this fashion a constraint may be more important in determining the 'winner' form (they show that this resembles more closely the data found in Frisch et al. 2004). Hayes & Wilson (2008) move away from using fixed constraints and instead have their model devise its own constraints, which are weighted.

Regardless of which view is adopted, a difficulty that arises is that in both cases the constraint is based on frequency in the input. In this way, a phonological constraint on similarity will cause a pattern of low frequency in the lexicon, but at the same time a low frequency in the lexicon causes the gradient constraint on frequency. In this way, frequency just by itself may be insufficient as a model of phonotactic legality. A central problem that this thesis addresses is differentiating the effect of lexical frequency from a phonological constraint on similarity (i.e. similarity constrains statistical learning).

1.3.4. Segment-specific versus abstract phonotactics

In the discussion of McQueen's (1998) study, phonotactic legality was seen to influence word-spotting. The sequence [mr], for instance, was separated by a syllable boundary by Dutch native speakers. However, this constraint *[mr] at syllable beginning is arguably a sub-case of a more general constraint banning all nasal+liquid combinations (mr, ml, nr, nl) from syllable beginning. In this way, the constraint is based on natural classes, and applies more generally. Evidence for this view is found in Shatzman & Kager (2007), who find that Dutch subjects can more quickly identify non-words when they were subject to an abstract phonotactic constraint on the distribution of labials (this study is dealt with in more detail in 1.4.1). Additionally, in Onnis et al.(2005) subjects were found to group consonants together with similar manner of articulation. This suggests that learners are attentive to more than just the segments in language, and instead are sensitive to a more abstract level, namely the features (and were able to generalise over them) to find regularities in the speech signal. The model in Hayes & Wilson (2008), which devises constraints and constraint weightings that apply to English syllable onsets, shows that the model looks for rules that apply as general as possible. In the constraints that were created by the model, features were seen to be highly relevant for phonotactic learning. This is evidence that a simple segment-based learner is insufficient. The current experiment depends on learners' ability to abstract away from segment level, and to perceive consonants with similar in place of articulation as belonging to the same group (as in Frisch et al. 2004).

Phonotactics seem to be a useful tool to segment speech. It is not surprising, therefore, that humans can quickly acquire them. They appear to be predisposed to attend to the input signal and deduce this information. However, phonotactic knowledge displays degrees of acceptability, (Luce & Large 2001, Vitevitch & Luce 2005, and the degrees of well-formedness of the phonotactic cues is also detected in segmentation Mattys& Jusczyk 2001). Although many phonotactic constraints appear to apply at a syllable-based level, there is also evidence of phonotactic constraints that apply sequentially, with the possibility of applying across syllable boundaries. The constraint on consonant similarity, as dealt with in this thesis, is a sequential constraint applying to non-adjacent consonants. How phonotactic constraints apply exactly to explain the gradient pattern of well-formedness judgements is subject of ongoing research. However, the model by Wilson & Hayes (2008) suggests that features are important to defining constraints. Additionally, Frisch et al. (2004) show that a gradient constraint on phonotactics based on consonant features can explain lexical data. Taken together with the results of Onnis et al. (2005), features grouped into natural classes may explain phonotactic patterns found in the lexicon.

1.3.5. Calculating phonotactic patterns in the lexicon

To map the frequency of specific phoneme patterns occurring in the lexicon, the observed-expected ratio (O/E) can be used. The ratio is more useful than only frequency counts, since it gives information on whether a pattern occurs less, or more than expected, in this way highlighting preferences or restrictions in the lexicon. The O/E ratio expresses the ratio of the number of observed forms in the lexicon (O) to the number of occurrences which would be expected if phonemes can combine at random (E).

The O/E ratio is particularly useful to indicate whether a consonant pair can combine freely, or that it is constrained in some way. (8b) gives the O/E measure as used for CVC consonant patterns relevant for this experiment. The expected count is calculated on the basis of probabilities of consonants appearing in a specific context, such as (C_iVC) as opposed to counting individual occurrences of the consonant such as (C_i). Positional information of the segment is needed to compare with the Observed value, since, if the probability of C_i (written as $P(C_i)$) is calculated, this will not include any information about the context in which the consonant occurs: thus any constraints from preventing the consonant in a particular position will not be registered (e.g. if consonant X never occurs word initially, but occurs often in codas, $P(C_xVC$ will show a much lower count than $P(C_x)$. Thus, also any applying co-occurrence constraints will not be detected when the Expected value is calculated without context (Kager Shatzman, in progress). By including a measure of contextual information, the Expected value (as given in 8b) will perhaps provide a more conservative number, but a more lexically based value.

8.) a. $$O/E = \frac{\text{the observed counts of consonantal combinations}}{\text{the consonants can freely combine at random}}$$

(Pierrehumbert 1993: 3)

8.) b. $E = \text{Expected} = P(C_1VC) * P(CVC_2) * N_{CVC}$

(Kager & Shatzman 2007)

If the O/E-ratio of a certain lexical item is below 1, it means that that lexical item does not appear as often as would be expected: that lexical item is *under-represented* in the lexicon. On the other hand, an O/E-ratio above 1 indicates that a specific lexical item is present more often than expected: that lexical item is *over-represented*. Patterns that are under-represented in the lexicon may be caused by a phonotactic constraint, such as the OCP constraint.

To compute the O/E ratio, the same method is used as described in Shatzman & Kager (2007: 1409). The O/E ratio of consonant sequences C_1VC_2 is computed by taking the number of the CVC sequences found in the Dutch lexicon (the observed value) and dividing that by the expected value. The expected value is the probability that C1 occurs in initial position in a CVC sequence, multiplied by the probability that C2 occurs in final position in a CVC, multiplied by the total number of CVC sequences in the lexicon (see (8)). The CELEX database (Baayen, Piepenbrock, Gulikers (1995) is used to search the Dutch lexicon.

1.4. The similarity constraint

In several languages, Javanese, English, French, Latin, Russian, Japanese, Rotuman (in: Coetzee & Pater 2006), Hebrew, Arabic, and Dutch (Berent & Shimron 2003, McCarthy 1986, Shatzman & Kager 2007) examination of the lexicon shows evidence of a constraint on similar consonants occurring together. McCarthy (1986) finds categorical avoidance of similar consonants sequences in various Semitic languages, and postulates a constraint: the obligatory contour principle (OCP constraint). This phonotactic constraint disfavours combinations of homorganic consonants in proximity to each other (McCarthy (1986)).

Pierrehumbert (1993) and Frisch, Pierrehumbert, Broe (2004) study the OCP in the Arabic lexicon, and on the basis of lexical analysis (the O/E ratio) find a gradient distribution of consonants that are subject to the OCP-constraint. Both the distance and level of similarity of the consonants are found to influence the strength of the constraint. Distance is defined as distance on the consonant tier. Following McCarthy (1986), Frisch et al. (2004) separate consonants and vowels in Arabic word forms onto different autosegmental tiers. In this way, a word form is represented as (9):

(9) vowel tier: u i
| |
skeletal tier: C V C V C
| | |
consonant tier: k t b
(example from Frisch et al. 2004:188)

On the consonant tier, consonants can be adjacent or non-adjacent. In, for instance /d s w/, the pairs /d,s/ and /s,w/ are an adjacent pair, and /d,w/ a non-adjacent pair. Frisch et al. show that the strength[6] of the similarity constraint is dependent on the level of adjacency; its strength decreases when the consonants are non-adjacent. Moreover, the identity of the intervening consonant is irrelevant. As a result, the similarity constraint clearly applies non-locally.

The level of similarity also influences the strength of the constraint. The level of similarity is calculated as given in (10):

(10)
$$\text{similarity} = \frac{\text{number of shared natural classes}}{\text{number of shared natural classes + natural classes not shared}}$$

Similarity is computed over the natural classes of a segment inventory, rather than over features, this is because “features that are not contrastive do not define unique natural classes, and so will not contribute to similarity” (Frisch et al. 2004: 197). They argue that the OCP should be considered a gradient constraint that can be related to a measure of similarity between homorganic consonant pairs. Though the gradient measure (in both Pierrehumbert 1993, Frisch et al. 2004) provides a better explanation of lexical distribution of Arabic consonants (as given by the O/E ratio), than a categorical constraint banning similarity (as proposed by McCarthy 1986), Coetzee & Pater (2008) find a closer approximation with use of

---

[6] The word ‘strength’ implies the notion of a gradient constraint. However, the data that is found by Frisch et al. (2004) can also be explained by categorical constraints that interact in a stochastic manner, as shown by Pater & Coetzee (2008). For clarity however, ‘strength’ of the constraint is used to indicate restriction in the lexicon.

a model that uses weighted constraints. It is undecided, therefore, whether the constraint on consonant similarity is a single gradient constraint, or whether several interacting constraints (or weighted constraints) on co-occurring features give rise to the gradient pattern attested as a result of OCP. However, the exact nature of the constraint is not essential for this thesis.

In a study of the phonological constraint on similar consonants in Hebrew, Berent & Shimron (2003) find that identical consonants behave differently from similar homorganic consonants. In an experiment where native speakers of Hebrew had to rate the acceptability of non-words in their language, identical consonants were rated to be more acceptable than homorganic consonants. This clearly contrasts with the prediction of similarity avoidance (as given by Frisch et al. 2004), which predicts that identical consonants are maximally similar, and are thus avoided to a maximal degree. As a result, it is necessary to include the identity hypothesis which allows for the constraints on identical consonants to differ from those governing similar consonants. In this way there are two separate constraints: identity avoidance and similarity avoidance.

In this thesis, the constraint on consonants with similar place of articulation across a vowel (e.g. $*C_{labial}VC_{labial}$) will be referred to as the similarity constraint. Although the similarity constraint abstracts away from the lexical representation and applies on the consonant tier, due to processes of consonant assimilation in Dutch it is necessary to consider CVC sequences as a lexical measure (see 1.4.1.). The effect of the similarity constraint is found to be gradient (Frisch et al. 2004), however, the nature of the constraint itself, either gradient or categorical, is left unspecified. For interpretation of the results the nature of the constraint is not crucial. What is crucial is that the constraint restricts the occurrence of consonants that share place of articulation, in such a way that consonant sequences that do not share place of articulation will be preferred over consonant sequences that share place of articulation.

As Berent & Shimron find in Hebrew, identical consonants are subject to a different (or additional) constraint. In other languages, it is also found that strictly identical consonants escape avoidance effects: in Javanese (Yip 1989), Muna (Coetzee & Pater 2006), Japanese (Kawahara et al. 2006). There is also evidence that identical consonants in Dutch escape avoidance effects. Shatzman & Kager (in progress) report that in Dutch, lexical distribution shows the following pattern:

features different	0	1	2	3	4
O/E value	0.7309	0.3592	0.4902	0.4597	0.4597

Table 1. Average O/E values for labial consonant pairs which differ in a given number of features (voice, continuant, nasal, sonorant), calculated in CELEX stems. (table from Shatzman & Kager (in progress)

When there are null different features, the forms are identical. But these identical forms are the least under-represented, though they are maximally similar. It seems, therefore, that similar consonants behave differently than identical consonants. Moreover, in a word segmentation experiment by Boll-Avetisyan & Kager (2008), Dutch subjects were found to prefer non-words that contained identical consonant sequences (e.g. [p_p_t]) over words that had the identical consonants further apart (e.g. [p_t_p]). Instead of avoidance of identical consonant sequences, therefore, identical consonant sequences were preferred. These findings clearly contrast with what the similarity constraint as proposed by Frisch et al (2004) predicts. As a result, identical consonants appear to be subject to a different constraint, or perhaps because of their marked form they have an escape-hatch to the similarity constraint.

The similarity constraint is found cross-linguistically. Besides the effect of the similarity constraint found in various Semitic languages by McCarthy (1998), in Hebrew (Berent & Shimron 2003), in Arabic (Frisch, Pierrehumbert, Broe 2004); similarity effects are also attested in English (Coetzee 2005); in Muna (Pater & Coetzee 2008); (Kawahara, Ono & Sudo 2006) and in Dutch (Shatzman & Kager (2007). The experiments in this paper will focus on Dutch, and the next sections deal with the OCP-constraint in Dutch in more detail, however, it is first necessary to look at the method of computing lexical representation.

1.4.1. Similarity avoidance in Dutch

The Dutch lexicon also shows evidence for the OCP-constraint. It is important to note that the OCP and similarity constraint as found in the examples above, in Hebrew and Arabic, applies to the consonant tier (autosegmental model, McCarthy 1986). As shown in (9), the consonant tier abstracts away from vowels, such that consonants interact without interference of the vowel (evidence for this view is found in Bonatti et al. 2005 who finds that consonants and vowels perform different roles in speech segmentation). In Arabic phonotactics, however, consonant clusters are infrequent, especially when compared Dutch. Dutch phonotactics allow complex clusters, as the CVCCCC word *herfst* 'autumn' illustrates. In particular, Dutch CC clusters behave differently and show, for instance place assimilation (e.g. /mp/ in *impressie* 'impression', and in English also: /mp/ in *impression*). Consequently, if only the C-tier were examined, then /mb/, /mp/ would show an over-representation due to place assimilation. As a result, the similarity constraint in Dutch is investigated by observing C_1VC_2 patterns as they occur in the lexicon. In this fashion, the similarity constraint is researched on non-adjacent consonants across a vowel.

The Dutch lexicon shows an under-representation of consonants sharing place of articulation, in particular the labial consonants (Kager & Shatzman 2007, Boll-Avetisyan 2008). Evidence for the OCP-constraint in Dutch functioning during speech processing in Dutch is found in Kager & Shatzman (2007). In their experiment, Dutch participants and were asked to decide as quickly as possible whether the non-word they heard was an existing Dutch word. The reasoning was that non-words which are subject to a phonotactic constraint, such as the similarity constraint, would be rejected quicker. Moreover, they carefully control for lexical influences, since non-words which are subject to a constraint will logically also have fewer lexical neighbours in the language, and thus influence processing. In this fashion, they find evidence that the phonotactic constraint on the distribution of labials influences word processing, which crucially does not rely on their lexical properties. Following this study, the experiments in this paper also control for lexical factors.

Additional evidence for the OCP constraint in Dutch comes from an experiment of speech segmentation of an artificial language by native speakers of Dutch (Boll-Avetisyan & Kager (2008). This experiment tested whether the OCP-constraint on non-adjacent labials ($*C_{labial}VC_{labial}$) could predict word segmentation. Dutch subjects listened to an artificial language containing only CV syllables. The syllables' initial consonant was either a labial consonant (given as 'P'), or a coronal consonant (given as 'T')[7]. These CV syllables were concatenated to make up trisyllabic words (CVCVCV), yielding a structure as seen in (11):

(11).a. …. $C_PVC_PVC_TVC_PVC_PVC_TVC_PVC_PC_TV$ ….

If these syllables are heard in a continuous speech stream, with no other cues to segment speech, the AL language given in (11a) can be parsed in 3 different ways as given in (11b) (the vowels are not shown in the figure, since the similarity measure abstracts away from vowels to the consonant tier):

---

[7] The consonants used: position 1 (labial): {p,b,m}, position 2 (labial): {p,b,m}, position 3 (coronal): {t,d,n}

(11).b. … PPTPPT**PPT**PPTP**PTP**PTPP**TPP**TPPTPPT…

The artificial language contained nine syllables: three per set. This results in 81 possible parsings ((3*3*3)*3 = 27 PPT words + 27 PTP words + 27 TPP words).

The language was carefully controlled for statistical confounds (e.g. the transitional probability between all syllables was 0.33), lexical confounds (e.g. probability of syllable occurrence in word-positions; (resemblance to words or morphemes), and for phonological confounding factors (e.g. segmental durations; number of identical vowels in sequence). As a result, the assumption is that the only cue for segmentation present in the language was the OCP-labial constraint.

Forty-two native speakers of Dutch were exposed to the artificial language for a duration of 10 minutes, after which they completed a forced choice test between two words that were both possible parsings in the AL.

The results showed that the subjects could use OCP-labial to segment the AL. Subjects were found to prefer a PTP segmentation over a PPT segmentation (58%), and a PTP segmentation over a TPP segmentation (55%). The subjects appeared to avoid a PP-sequence, as predicted by the OCP-labial constraint; however, the preference results also show that the cue is very subtle[8].

1.5. Artificial Language Learning

Similar to the experiment by Boll-Avetisyan & Kager (2008), this thesis will use an artificial language to examine speech segmentation. An artificial language is a miniature language in which the language rules and regularities can be controlled. Since the speech is synthesised by the computer, factors such as duration, prosody, pitch, can all be manipulated and kept constant. In this fashion the segmentation cues present in the AL can be carefully controlled (however, of course, not all speech segmentation cues may be known at present). Subjects listening to an AL language are expected to perceive it as linguistic speech[9], and they are also

---

[8] Noteworthy is the fact that there are identical consonant pairs in the language, which are shown by Berent & Shimron (2003) to behave differently in Hebrew, and also Boll-Avetisyan & Kager (2008) find in a separate experiment they find that identical consonants behave differently , and are not subject to OCP-labial. This may have had an effect on the OCP results.

[9] In the discussion of Saffran et al. (1999), they consider whether artificial language input could be perceived as non-linguistic input, as the language is devoid of prosody, rhythm, duration, or pause, which is unnatural.

told prior to the test that they are going to listen to ‘a nonsense language’ in which they are to identify ‘words’. To segment the AL subjects are expected to use patterns of their native language. As shown by Vroomen, Tuomainen & de Gelder 1998, adults listeners will impose a segmentation strategy appropriate for their native language on an artificial language. When they compared word segmentation of Finns, Dutch and French on one AL, they found specific influence of each of the separate native languages. Additionally, Finn & Kam (2008) find influence of the native language when segmenting an AL. They exposed learners to an AL in which the transitional probabilities were pitted against English word-formation patterns: words defined by TPs began with consonant clusters that violated English phonotactic rules (480). Even with 72 minutes of exposure time subjects did not learn the TP pattern over an illegal English phonotactic pattern. As a result, phonotactic patterns of the native language may be assumed to influence segmentation when listening to an AL.

Humans are skilled statistical learners (e.g. Saffran et al 1996) and are sensitive patterns to phonotactic patterns at early age (e.g. Jusczyk et al. 1993), acquire phonotactic regularities with ease (e.g. Onishi (2002), Saffran & Thiessen (2003)) and can use this knowledge to segment speech (e.g. McQueen (1998), Mattys Jusczyk (2001). Additionally, increased phonotactic probability facilitates word processing (e.g. Vitevitch & Luce 2005). Since the native language influences segmentation of an artificial language (e.g. Vroomen, et al. (1998), Finn & Kam (2008), lexical statistics of the Dutch language may be assumed to influence segmentation behaviour of Dutch subjects in an AL.

1.6. Hypothesis

The coming about of phonotactics may be a process on a feedback loop: if frequent patterns in the lexicon are considered to be more acceptable (Luce & Large (2001), Vitevitch & Luce (2005), the coinage of new words and loanwords are influenced by this statistic, in this way reinforcing the pattern further. For example, *Christmas* as a loanword in Japanese takes on the Japanese phonotactic pattern (which prohibits consonant clusters), and becomes [kurisúmasu]. In this way, the presence of a phonotactic preference for CV syllables causes CV syllables in

---

However, as a counter argument they say that some test-subjects had asked them who had recorded the speech. This indicates, Saffran et al. argue, that the subjects thought they heard a real speaker, and thus may be assumed to have perceived the artificial language as linguistic input. Subjects in the present experiment also often enquired with wonder who the incredibly monotone speaker of the language was; they may therefore also be assumed to have perceived the AL as linguistic input.

new words, and hereby strengthens the CV preference[10]. Saffran (2002) argues for constraints on statistical learning, and that certain statistics in the input speech may be easier to process; in this way learning influences the preferred forms of language. She reasons that "human languages may have roots in the learning process itself: constraints on language learning may shape the structure of natural languages" (2002:173). If this reasoning is applied to statistical learning, the constraints found in statistical learning contribute to the forming of patterns in language, and these patterns may be detected in the lexicon. As a result, phonotactic patterns in the lexicon may be said to determine the phonotactic constraints (Frisch et al. 2004: 180).

Lexical distribution patterns arise as a diachronic result. Frisch et al. (2004) suggest that lexical items that are not subject to the OCP constraint are easier to process and so will be favoured in language acquisition, and language use. In this way, a phonological constraint may become a phonotactic constraint because of a selection process favouring certain phonological forms over time. Ultimately, as a result of diachronic changes in a language, and the lexical patterns reinforcing themselves as in a feedback loop, "the phonological grammar and the lexicon conform to one another, and the grammar contains constraints that are phonologically grounded" (Frisch et al. 2004: 221). This suggests that the OCP constraint, and the resulting lexical under-representation of patterns subject to this constraint, are difficult to separate.

In the Dutch lexicon, the OCP constraint causes CVC sequences with consonants that share place of articulation to be under-represented (Shatzman & Kager 2007). Though the presence of the OCP constraint is shown in studies by Shatzman & Kager (2007), and Boll-Avetisyan & Kager (2008), it may also be argued that the subjects are using the knowledge of lexical distribution from their native language for word segmentation. This argument can be backed by the evidence of human capacity for statistical and phonotactic learning and their ability to apply that to artificial language learning. Moreover, as Berent & Shimron (2003: 39) note: "[d]istributional facts may be shaped by non-linguistic or diachronic factors that are not currently active in a language; hence they may not necessarily reflect the linguistic competence of modern speakers"; in other words, the OCP constraint may not even contribute to the segmentation behaviour, but the effect found may be due only to lexical distribution. As a result, a distinction should be made between an under-represented pattern that is caused by a constraint, a systematic gap, and an under-represented pattern that is not caused by a

---

[10] The phonotactic form of loanwords is not a conscious process: perception is influenced by the native phonotactic pattern: non-word *ebzo* is perceived by Japanese speakers as *ebuzo*, in contrast to French speakers, who hear *ebzo* as French phonotactics does allow VCCV (example from Kuo 2008).

constraint, an accidental gap. This distinction is also used by Frisch & Zawaydeh (2001) to compare lexical judgments of non-words belonging to a systematic gap as the result of the OCP constraint, with non-words belonging to an accidental gap[11].

To compare the influence of the OCP-constraint to lexical distribution effects, the Dutch lexicon is scanned for systematic gaps and accidental gaps. By using the O/E ratio, which compares observed counts (O) to expected counts (E), under-represented consonant pairs can be identified in the Dutch lexicon. In Dutch, the consonant sequences of a labial (P) followed by a labial (P) are under-represented, caused by the OCP-constraint. The consonant sequences of a labial (P) followed by a velar (K) are equally under-represented, but are not subject to any known constraint. This allows for comparison of word segmentation behaviour in an AL by Dutch native speakers.

The question is thus whether the similarity constraint directly causes test subjects to avoid sequences of consonants with similar place of articulation, or whether that avoidance may be caused by an effect of lexical distribution, and that subjects avoid similar consonants because they are statistically under-represented in the lexicon.

The experiments presented in this thesis attempt to tease apart the effect of the similarity constraint and the effect of under-representation in the lexicon statistics. In this way, this thesis addresses two hypotheses. The first hypothesis states that for Dutch native speakers, word segmentation in an artificial language will be influenced by the constraint on consonants sharing place of articulation, and consequently the participants will avoid words in which non-adjacent consonant share place of articulation.

The second hypothesis predicts that lexical distribution of consonants in Dutch influences word segmentation, such that consonant sequences that are under-represented in the lexicon will be avoided. By testing both hypotheses with closely matched artificial languages, the effect of the lexical and the phonotactic constraint can be disentangled.

This gives rise to the following predictions: if lexical distribution as given by the O/E ratio influences segmentation, then subjects will segment the artificial language at places where the O/E ratio is lowest. Since both the consonant sequences PK (labial-velar) and PP (labial-labial) are under-represented in the Dutch lexicon, Dutch native speakers are predicted to place a word boundary between those consonants. If the similarity constraint influences segmentation, then subjects will avoid words in the artificial language that have sequences of similar consonants (e.g. avoid sequences of labial-labial (PP)), however, the similarity

---

[11] Frisch and Zawaydeh (2001) find that non-words that belonged to a systematic gap were considered less word like: showing that the OCP constraint is independent of lexical statistics in Arabic

constraint will have no influence on the PK-sequence. If the similarity constraint applies separate from the influence of lexical distribution, then a difference should be found between segmentation of the PK sequence and the PP sequence.

In this thesis, three artificial language learning experiments are conducted to compare and examine the effects of lexical under-representation and the OCP-constraint on word segmentation. The first experiment (PKT) tests the effect of lexical under-representation on speech segmentation. The second experiment (PPT) is minimally different from the first experiment but includes a homorganic consonant pair (labials), this enables a comparison of the effect of the OCP constraint with lexical under-representation. The third experiment (PPT2) further examines the OCP constraint, but uses a different set of labial consonants than the PPT experiment. This was done to investigate under what conditions this constraint applies. In this fashion these experiments should shed light on the differences between the phonotactic OCP constraint, and the influence of lexical distribution.

2. Experiment 1 (PKT): The Influence of Lexical Distribution on Speech Segmentation.

In this experiment, the influence of lexical distribution on word segmentation is tested. By comparing sequences on basis of their observed expected values (O/E), the prediction is made that segmentation will occur where the O/E value is lowest.

2.1. Method

Participants

Forty-two Dutch native speakers were recruited for this experiment. They were paid 5 euros each for participation. The participants were divided into three groups of fourteen subjects (Table 2).

Group 1		Group 2		Group 3	
male:	1	male:	0	male:	0
female:	13	female:	14	female:	14
age (years; months):	23; 8	age (years; months):	22; 1	age (years; months):	22; 5

Table 2. Participants Experiment 1 (PKT).

Materials

In Dutch, labial consonant sequences are more strongly under-represented than other (coronal or velar) consonant sequences (Shatzman & Kager 2007). This makes the labial consonants best suited for researching the OCP constraint in Dutch. To compare the OCP constraint with influences of lexical distribution, an under-represented consonant sequence containing a labial is sought in the lexicon.

Under-representation is defined as an O/E ratio value lower than 1 (as discussed in Chapter 1.3.5.). The CELEX database (Baayen, Piepenbrock, Gulikers (1995) is used to search the Dutch lexicon. The labial consonant sequences are on average under-represented (see Table 3), which is attributed to the similarity constraint in the lexicon. In both Shatzman & Kager (2007) and in Boll-Avetisyan & Kager (2008), the test-subjects were found to be influenced by the similarity constraint and to avoid the Labial-Labial sequences.

Analysis of the O/E ratios for C_1VC_2 sequences in the lexicon reveals that under-representation is also found among consonant sequences that do not share place of articulation. By choosing consonant sequences that are matched as closely as possible on O/E ratio value, and other lexical factors (such as occurrence in word initial position), and phonological factors, two artificial languages can be created in which the crucial difference is shared of place of articulation (i.e. under-represented sequences are either subject to OCP constraint, or not).

Table 3 shows the O/E values of labial-labial sequences (PP), and of labial-velar (PK) sequences. As a group, both sequences are under-represented: the PP sequences (O/E = 0.47) and for the PK (O/E = 0.71) sequences. The fact that the PP sequences are more strongly under-represented is not problematic, since from the consonant sequences in Table 2 a smaller, and more precise, selection is made of consonant pairs which are matched for O/E values. Importantly, both PP and PK show a comparable level of under-representation as a group, and although PP under-representation, there is no reason to assume a constraint causing the PK under-representation. As a result, PP and PK seem to be suitable candidates for comparison of segmentation behaviour.

PP sequences								**PK sequences**						
	p	b	f	v	m	w	avg		k	g	x	G	N	avg
p	1.13	0.71	0.22	0.17	0.21	0.38	0.47	p	0.77	0.63	0.25	0.62	0.42	0.54
b	0.25	0.79	0.18	1.53	0.57	3.53	1.14	b	1.01	1.89	0.18	0.92	0.34	0.87
f	0.11	1.55	0.13	0.13	0.62	0.12	0.44	f	0.81	0.46	0.17	0.55	0.96	0.59
v	0.03	0.19	0.15	0.10	0.07	0.17	0.12	v	0.23	0.23	0.42	0.37	0.67	0.38
m	0.19	0.49	0.27	0.13	0.31	0.25	0.27	m	0.76	0.59	0.65	1.08	0.75	0.76
w	0.50	0.29	0.57	0.28	0.46	0.12	0.37	w	0.86	1.04	1.32	1.64	0.77	1.13
avg							0.47	avg						0.71

Table 3. Average O/E values of C_1VC_2 sequences, CELEX type counts. On the left are labial-labial sequences (PVP) on the right labial-velar sequences (PVK).

From the under-represented sequences four consonants (two continuants /f, x/ and two non-continuants /m,k/) are selected. These consonants are then combined with vowels to make up CV-syllables, which can be concatenated into three syllabic CVCVCV words (Table 4):

P	K	T
me	ke	so
mu	ki	sy
fa	xo	da
fu	xi	dy

Table 4. The PKT artificial language.

By assigning consonants to specific 'slots' in the language, words can only be made up in fixed consonantal frames. In the language, labial consonants (P) were assigned to slot 1, velar consonants (K) to slot 2, and coronal consonants (T) to slot 3. This results in the language as given in Table 4, which can be segmented in three ways, as illustrated in Figure 1.

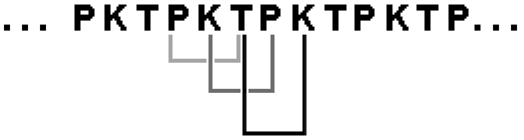

Figure 1. Illustration of the three word-segmentation possibilities in continuous speech.

Since there are four syllables in each position, and there is no restriction on syllable sequences, the transitional probability is identical: both for adjacent, and non-adjacent syllables (the same holds for consonants, though for consonants all TPs are 0.5; for the vowels the TP pattern is identical between positions -each position has one vowel that occurs twice).

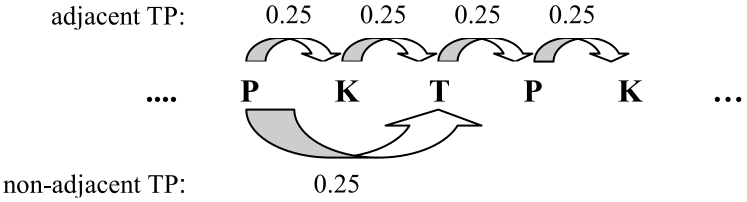

Figure 2. Transitional probabilities (TP) of syllables in AL PKT (the P,K, or T denotes a syllable starting with a labial, velar, or coronal consonant, respectively).

In this fashion, the speech stream contains no cues for segmentation. The subjects must rely on phonotactics to segment either a PKT, KTP, or a TPK word. The prediction is that the low O/E value between PK will determine this segmentation and that KTP words are preferred.

Word boundaries can be placed at three positions: between the labial (P) and velar (K) consonants, between the velar (K) and coronal consonants (T), and between the coronal (T) and labial (P) consonants. As seen in Table 4, the PK sequences have the lowest mean value. Word boundaries are expected where the O/E is lowest, thus at PK-sequences.

$P_1(V)K_2$		$K_2(V)T_3$		$T_3(V)P_1$		$P_1(V)K_2$	
labial- velar	**O/E**	**velar-coronal**	**O/E**	**coronal-labial**	**O/E**	**labial- velar**	**O/E**
fx	0.17	kd	0.37	df	0.53	fx	0.17
mx	0.65	ks	0.66	dm	0.91	mx	0.65
mk	0.76	xs	1.16	sf	1.55	mk	0.76
fk	0.81	xd	2.02	sm	1.81	fk	0.81
Average	0.60		1.05		1.20		0.60

Table 5. Consonant frame of the PKT artificial language.

From the under-represented PK-sequences (Figure 2 above), the velar plosive /k/, and the velar fricative /x/, are selected. Coronal consonants are selected to complete the three syllabic consonant frame. In this fashion there cannot be any OCP-influence between the labial, velar and coronal consonants. The coronals /d,s/ have fairly neutral O/E distribution that should not lead to lexical confounds (average O/E of TP-sequences: 1.20, of TK sequences 1.05)

For the labials (P) in the PK sequence, only few consonants were suitable for testing. Labials are shown to be preferred word-initially: Kager & Shatzman (in progress) report the average O/E of initial P = 2.71, and are thus a possible confound. (Still, a preference for initial labials will result in a preference for a PKT segmentation: which is contrary the hypothesis.) The /v/ and the /b/ are strongly over-represented at word initial position and may cause a confound, and are excluded. The /w/ is excluded because it is an approximant and is problematic in an AL. Ultimately, the consonants available are [f,m,p]. The /p/ is eliminated in a later stage, because it could not be included in a statistically controlled AL[12]. In Table 6, the occurrence of the P and T consonants in initial position is quite balanced. The K-consonants, however, show a higher value. Especially /x/ may be expected to have a confounding initial preference, this is discussed below.

---

[12] [p] was preferred since it is a plosive, not a nasal, which creates a more even pattern of manner of articulation in the AL. Additionally, *px, pk* have a lower average O/E value than *mx, mk*, and the /p/ is more evenly distributed in various word-positions in Dutch than /m/. However, because the /p/ and /f/ cohort density values are overall very low, the combination of the labial consonants {/p,f/} does not yield a language that can be parsed in three ways (PKT, KTP, TPK) without significant difference between the parsings (both transitional probability and cohort density values fail a T-test comparison). The impact of using /m/ instead of /p/ is tested in experiment 3.

	Position 1 labial (P)	**O/E**	**Position 2 velars (K)**	**O/E**	**Position 3 Coronals (T)**	**O/E**
	.m	1.25	.k	1.52	.d	.d: 0.95
	.f	0.98	.x	2.10	.s	.s: 1.05
average		1.11		1.81		1.00

Table 6. Overview of the O/E ratio of the consonants in initial position in Dutch (Celex database)

The selected syllables are analyzed for their occurrence in word positions in the lexicon. Table 7 gives the syllable type counts in three word positions in the lexicon (word-initial, word-medial, word-final)[13].

	word position	**summed observed types**	**average per syllable**	**percentage** (position preference)	**prefers parsing:**
position 1 (P) {me,mu,fa,fu}	initial	131	33	0.45	PKT
	medial	148	37	0.51	TPK
	final	10	3	0.03	KTP
position 2 (K) {ke,ki,xo,xi}	initial	70	18	0.36	KTP
	medial	101	25	0.52	PKT
	final	22	6	0.11	TPK
position 3 (T) {so,sy,da,dy}	initial	77	19	0.30	TPK
	medial	161	40	0.62	KTP
	final	21	5	0.08	PKT

Table 7. Percentage of the AL syllables in various word-positions in Dutch. For type counts per syllable, see appendix B.

Overall, preferences for syllable positions are fairly even: a pattern for medial preference, and a low word-final occurrence is found among all syllables. The preference for the velar consonant /x/, that was observed in Table 6 does not show in Table 7. The cause for the over-representation of the /x/ was the morphological prefix /xə-/. Since the schwa is not used in the AL, the initial preference for the /x/ is not assumed to be a confound. Table 8 sums the type counts which favour a particular segmentation to determine whether a certain segmentation is preferred. This is not seen to be the case.

---

[13] Syllable positions are calculated as follows. The lexicon is analysed for occurrences of the specific consonant in the positions of a segment-frame: .C̲VCV , CVC̲VCV, CVC̲V. (The '.' signifies a word boundary). After summation of the occurrences in all three positions, the percentage of the total number of occurrences can be calculated per position.

syllable type counts that favour a PKT segmentation:	253
syllable type counts that favour a KTP segmentation:	241
syllable type counts that favour a TPK segmentation:	247

Table 8. summation of syllable position counts that lead to a specific segmentation.

Per consonant, two vowels are selected (only one vowel per consonant results in a very small language) from the available Dutch vowels: /a, e, i, o, u, y/, which are subdivided into high vowels: /i, u, y/ , which have a relatively short duration, and [–high] vowels: /a, e, o/, which have a relatively long duration. Each consonant was assigned a short and a long vowel, for symmetry and for equal durations of the CV-syllables.

To control for lexical preferences for certain CV combinations, the cohort density of CVC sequences is examined. Cohort density is calculated as “the sum of the logged frequencies of all words that share the initial three phonemes of the non-word” (Shatzman & Kager 2007). Table 9 shows the cohort density values of the selected CVC sequences.

Position 1 C_1VC_2	**Cohort Density**	**Position 2 C_2VC_3**	**Cohort Density**	**Position 3 C_3VC_1**	**Cohort Density**
mex	18.36	kes	3.43	som	3.56
mek	18.23	ked	0	sof	9.6
mux	0	kis	25.96	sym	1.15
muk	1.66	kid	0	syf	0.7
fax	0	xos	0	dam	31.1
fak	1.98	xod	13.47	daf	0
fux	0	xis	0	dym	0
fuk	0.6	xid	0	dyf	0
Average value	5.10	Average value	5.36	Average value	5.76

Table 9. Overview of the cohort density of the selected consonant-vowel pairs.

By means of a Student’s t-test, the cohort density values for the CVC sequences were compared for a possible preference. Lists in which each of the CVC-syllables occurs eight times (the language consists of 4^3=64 words) were compared, and did not show significant difference (p>0.68).

Additionally the transitional probabilities of the biphones was controlled. When the syllables are combined into CVCVCV non-words, the language consists of in total 64 ((4*4*4)) possible words (see appendix A for the entire word list). The three syllabic words can be parsed in three different ways, however, in this way *mekeso* is a word in the language, *kesome* is a word, and *someke* is be a word. Crucially, then, a particular parsing should not be favoured.

To control for this, the transitional probabilities (TP) of the biphone frequency are compared. In the lexicon, every phoneme has a certain probability of being followed by a certain phoneme. By computing this probability for the entire CVCVCV word, it allows for control over preferred phoneme sequences (as found in the Dutch lexicon). The TP is calculated as the logged product of the biphone transitional probabilities (Shatzman & Kager 2007). A comparison of the individual biphone TP values of non-words, using a Students' t-test showed no significant difference (p> 0.12).

The syllables as given in Table 3 above were concatenated to make 64 CVCVCV words. These words were ordered pseudo randomly (no two identical words after each other) in a list of 896 words (each word occurs fourteen times).

The artificial language stimuli and test items were synthesised using Fluency, with the Dutch male voice NL2 (200 Hz), the syllables were concatenated into fluent speech by MBROLA (Dutoit et al. 1999). A pseudo random order of 896 CVCVCV stimuli was created, with no non-word occurring twice in a row. In this way, every word was heard 14 times. The average syllable duration for all slots was 220 ms., resulting in the total duration of the language of 9 minutes and 54 seconds. Using the program Praat (Boersma & Weenink 2008) the beginning of the AL was faded-in from 30 dB up to 90.8 dB (sound volume was adjusted to comfortable volume using volume control on computer) over the first 5 seconds, and faded out to 30 dB over 7 seconds. This is done to ensure that listeners do not hear a fixed beginning or ending of the language.

Procedure

Subjects were tested in a soundproof booth. The sound was played over a set of Beyerdynamic DT250/80 headphones, at a comfortable volume. (Amplification: computer internal audio card RME digi96/8 PAD set to 17% output over a 2x Mono amplifier Marantz MA6100).

In this experiment, subjects were exposed to ten minutes of an artificial language made up of CV-syllables. The duration of ten minutes is found to be sufficient exposure in artificial language learning studies by Peña et al. (2002), and it is also used in the study by Boll-Avetisyan & Kager (2008) for Dutch AL segmentation. The artificial language is carefully controlled for cues that may influence segmentation. The aim is to leave only phonotactic distribution as a cue. In this fashion, the AL does not contain prosodic cues, no pauses, and no variation in duration (controlling for lexical cues is dealt with below in more detail).

Prior to listening to the AL for ten minutes, subjects had to complete a monosyllable identification test (similar to Bonatti et al. 2005:453). If subjects failed on more than 2 out of the 10 they were excluded (only one subject was excluded).

The subjects were told they would be listening to an artificial language and that they should try to find out what could be words in the language. Before the AL started, the following message was displayed on the computer screen:

(12) 'Je gaat nu tien minuten luisteren naar een denkbeeldige taal. Het gaat erom om uit te vinden welke woorden deze taal bevat. Luister daarom goed. Klik op de knop om te beginnen...' (You will now be hearing an imaginary language for ten minutes. The goal is to find out which words this language contains. Listen carefully. Click button to start…)

The subjects were only required to listen attentively (see Toro, Sinnett, Soto-Faraco (2005) for discussion on how attention to the input influences word-segmentation results of statistical segmentation: diverted attention was seen to lead to impairment of word segmentation performance). In total, the experiment lasted 15 minutes.

The test phase was a forced choice test. All words in the test phase had been heard previously in the AL. The difference between the two words in the test phase was the location of the word boundary. Subjects chose between PKT-KTP words, between KTP-TPK words, and between PKT- TPK words). Subjects were asked to choose, or guess if necessary, which of the two words is more likely to belong to the language they have just heard. After hearing the AL for ten minutes, the language faded out, and the following message was displayed:

(13) 'Je krijgt nu telkens paren van onzin-woorden te horen. Geef door een druk op de knop aan of het eerste woord (1) of juist het tweede woord (2) meer lijkt op een woord dat je in de denkbeeldige taal hebt gehoord. Klik op de knop om te beginnen...' (You will now be presented with pairs of nonsense words. Indicate by pressing the button whether the first word (1) or rather the second word (2) is more like a word you have heard in the imaginary language. Click the button to begin…)

The subjects' response was assumed to be less influenced by the AL language they heard with every presented test-pair. The number of test items was therefore restricted to 32.

Each subject chooses within one category. As a result, subject 1 will only choose between PKT and KTP words, where subject 2 will choose only between KTP and TPK words, and subject 3 chooses between PKT and TPK.

The test items were selected in such a way as to have an equal distribution of manner of articulation for consonants (manner of articulation influenced responses in Onnis et al. 2005); and in the distribution of vowels (equal number of short and long vowels). Also words with identical vowels in adjacent syllables were avoided.

2.2. Results

In test phase comparing PKT to KTP words, there was a 56% ($p<0.05$) preference PKT words. In the comparison of KTP and TPK, there was a 79% ($p<0.001$) preference for TPK. With PKT and TPK, there was a 73% ($p<0.001$) preference for TPK.

The prediction for this experiment was that segmentation would occur where the average O/E ratio of consonant sequence was lowest. As seen in Figure 3 below, the predicted segmentation would be KTP, where the sequence of PK is avoided.

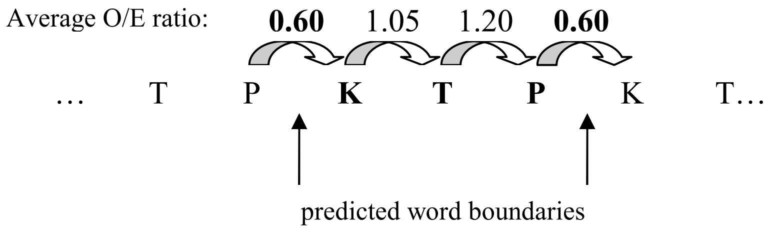

Figure 3. predicted word boundaries on basis of average O/E values of the PKT language.

The results (Table 10 and Figure 4) do not show this pattern. In a binomial test, the responses of the participants were compared for PKT vs. KTP, KTP vs. TPK, and PKT vs. TPK.

	N (response)	Preference	p (asymptotic significance two tailed)
PKT vs. KTP	251- 197	56% - 44%	0.021
KTP vs. TPK	94 – 354	21% - 79%	0.000
PKT vs. TPK	121 – 327	27% - 73%	0.000

Table 10. Results of the PKT experiment.

The means of the subject responses are given in Figure 4.

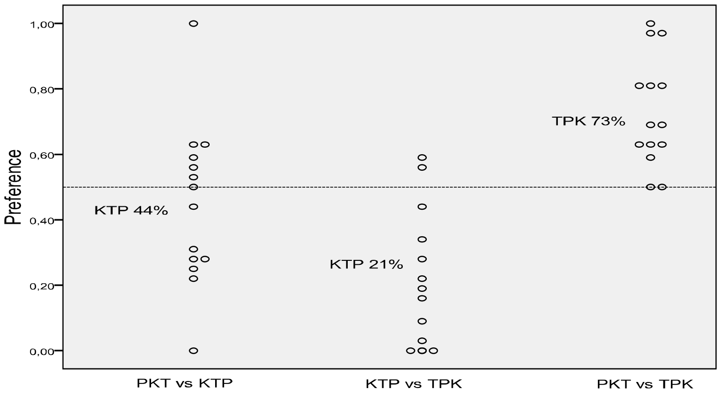

Figure 4. Results of the PKT language, each dot gives the mean preference per test subject.

2.3. Discussion

In the results, the predicted preference for segmentation KTP scores lowest. From Table 10 follows that KTP words have the lowest score (N(total)=291; PKT N(total)=372, TPK N(total)=681). Additionally, although no explicit prediction was made about the TPK-segmentation, the TPK words were predicted to score lower than the PKT words, this on account of the labial-initial preference in Dutch. The reverse seems to be true, and there seems to be a preference for coronals word-initially. Even if there is a preference for T-initial, however, the preference for PKT over KTP words still shows that the low O/E ratio between PK cannot predict segmentation preference.

Since the language contained many words (192 possible words), it may be the case that the words were difficult to recall from memory, and that the test phase stimuli influenced their responses. However, when looking at the responses on the first 16 test items, the same pattern emerges (PKT vs. KTP 59% - 41% (p=.006), KTP vs. TPK 21% - 79% (p=.000), PKT vs. TPK 26% - 74% (p=.000)).

The mean scores per stimulus were inspected for a preference for a particular consonant in word initial position, this is shown in Table 11. The scores of the velars /k,x/ are close together. The scores of the coronals /d,s/ are also close together, there is no indication of a preference for either consonant in word-initial position: the coronals were preferred as a group. The largest difference is found between the scores of the labials /f,m/. This suggests that the labials were not perceived as a group. In the study by Onnis et al. (2005), perceived phonological similarity is seen to influence segmentation. The possibility that the level of dissimilarity between the labials /f,m/ negatively influences segmentation is tested in Experiment 3.

initial consonant	mean preference	initial consonant	mean preference	initial consonant	mean preference
m	0.37	k	0.37	d	0.77
f	0.55	x	0.36	s	0.81

Table 11. The preference scores for test words in Experiment 1 (PKT).

Many subjects reported hearing mainly /s/ and /d/ in the AL, which could indicate an influence of phonological features or acoustical salience of these sounds in the synthetic speech. However, close examination of the intensity of speech sounds of the AL did not reveal any acoustical differences between the T-consonants and the other K and P consonants. Additionally, if segmentation occurred on the basis of lexical properties, certain phonemes are expected to be perceptually more salient, giving the impression of a phonological influence. As a result, even if the /d/ and the /s/ sound ‘louder’, this could still be caused by a lexical preference; in any case, acoustically, the cause for this loudness could not be pinpointed (the same language should be tested when synthesised by a different program).

For a more fine grained analysis, the results were entered in a stepwise regression analysis. The results are summarised in Table 12. Since the AL contains only possible lexical cues, an extensive search for lexical factors was conducted. This included preferences for vowels in word specific (initial, medial, or final) word positions, specific vowel sequences, and consonant sequences (for word-likeness influences). Additionally, though not expected to be an influence, token-frequency was inspected. None showed an influence on the results. In Table 12, the predictors listed were either shown to be relevant in other studies (transitional probability and cohort density were an influence in Shatzman & Kager 2007), or were seen to predict results in Experiment 1, 2, or 3. Due to the varying pattern of influence of the lexical factors, they were all considered in the regression analyses. Additionally, the use of the stepwise could indicate which of the factors had the largest influence. By showing all predictors in one table, it can be seen whether one of the lexical factors influences one or more test phases.

First, to ensure that the test words had not been selected to favour certain responses, the transitional probabilities and the cohort densities of the test-words were entered as predictors. They did not show any influence. Then other possible lexical influences were considered. Since many subjects heard /d,s/-initial words, a lexical influence was suspected, but as the O/E ratio was controlled for in the language, the relevant frequency of occurrence was also entered in the regression. However, for T-initial words, no influence of the initial consonant is found.

For the PKT vs. KTP words, regression analysis showed an inverse influence of word initial relative frequency. An inverse relation means that a low occurrence in the lexicon correlates with a high preference, i.e. K-consonants occur often word initially, but are not preferred word initially in the test results. This can be related to a high occurrence in the lexicon of /x/, caused by the morphological prefix /xə/: the high occurrence in the Dutch lexicon predicts a preference for /x/ initial words, but this does not surface in the results. The final syllable also shows an inverse relation, which is unexplained by examining the lexicon (final P-syllables show a low occurrence in the lexicon).

For the KTP vs. TPK words, no influence is found, other than word-type. At first, the presence of final-/i/ in the TPK-words was assumed to be the cause for the preference of TPK. Especially since the short vowel /i/ only occurred with K-consonants, and the /i/ is a more frequent word-ending than the /y/ or /u/ (which occurred with T and P consonants, respectively). However, the regression results show no influence. In the PKT vs TPK words there is an influence of the final vowel. This suggests that the final /i/ may have had a small influence during segmentation. In addition, there may be a tendency to align velars on the right edge of the word; although a CELEX search does not reveal any preference for velar right, research with infants suggest that there is a tendency to align velars with word endings (suggested by Natalie Boll-Avetisyan). Perhaps these factors in an additive way cause a preference for final K, and thus also promoting an initial T.

Overall, there are no cues that consistently appear to influence the results. The final syllable influence occurs in the results of PKT vs. KTP and PKT vs. TPK, which suggests an influence of the final syllable of PKT; however, the influence is found to contribute positively and negatively. As a result, only the word type is found to influence the results.

test pairs	**Group scores**	**type**	**TP**	**CD**	**.C O/E**	**.C relfreq**	**final V**	**init CV**	**fin CV**
PKT vs KTP	PKT 56% (p=0,021)	p=0,009				p=0,000*			p=0,021*
KTP vs TPK	TPK 79% (p=0,000)	p=0,000							
PKT vs TPK	TPK 73% (p=0,000)	p=0,000					p=0,001		p=0,045*

Table 12. Regression results of the PKT-language (N=42). An asterisk (*) indicates an inverse relation. TP= transitional probability, CD= cohort density, .C O/E= consonant word-initial O/E

ratio, .Crelfreq= consonant word initial relative frequency, finalV= final vowel, initCV= initial syllable, finCV= final syllable.

An unintended confound was the presence of the Dutch word *sofa* in the language. However, regression analysis did not show an influence of the presence of /sofa/ in the words (as the first part of a TPK word, or the last part of a KTP-word). Neither did inspection of test results on words containing the /sofa/-string against words with a 'broken' /sofa/ string (e.g. *sofaxi* vs *faxiso*). As a result, the Dutch word *sofa* is not expected to have an influence on the results (also, only one subject reported hearing the word).

In conclusion, no consistent influence of lexical distribution factors is found to explain the results of the PKT- artificial language. Though the results suggest a preference for T-initial consonants, a regression analysis did not show evidence for a lexical preference for T-initial. Only between the P-consonants a difference in preference was found in the test scores: suggesting that perhaps phonological similarity (as discussed in 1.2.) played a role; however, this pattern is not found with K-consonants, which scored even lower than P-consonants. Ultimately, this experiment showed that the O/E value between segments could not predict segmentation.

3. Experiment 2 (PPT): The Similarity Constraint in Speech Segmentation

The previous experiment tested the influence of under representation in the lexicon on word segmentation. It was found that the O/E ratio between consonants could not predict the segmentation behaviour. Instead of avoiding the under represented string PK, subjects were seen to prefer T-initial words (TPK) over both PKT and KTP words. Moreover, subjects preferred PKT words over KTP words.

In Boll-Avetisyan & Kager (2008) they find an influence of similarity avoidance on speech segmentation in an AL by Dutch speakers. This experiment also intends to find whether there is similarity avoidance, but crucially also enables comparison with the previous experiment by matching the O/E ratios for lexical distribution. As a result it is set up in similar fashion to the previous experiment, but intends to determine whether the similarity constraint influences speech segmentation, such that subjects avoid consonant sequences with similar place of articulation. If subjects place word boundaries between the PP-sequences, that is evidence for the similarity constraint. The similarity hypothesis thus predicts that subjects will prefer PTP-words over PPT and TPP words.

3.1 Method

Participants

Twenty-four Dutch native speakers were recruited for this experiment. They were paid 5 euros for participation. All were students at Utrecht University. None reported having hearing difficulties.

Group 1		Group 2		Group 3	
male:	1	male:	1	male:	2
female:	7	female:	7	female:	6
age (years; months):	20;8	age (years; months):	21;10	age (years; months):	21;11

Table 13. Participants of Experiment 2(PPT).

Materials

The AL was similar to the previous PKT language, but the velar consonants were replaced by labial consonants. The language contains no cues for segmentation, other than phonotactic cues, and as a result, segmentation of the language can occur in three ways, as given in Figure 5.

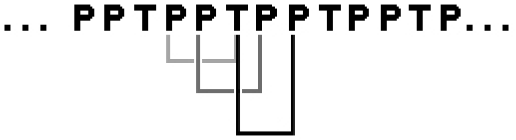

Figure 5. Illustration of the three possible segmentations in a PPT artificial language

The artificial language aims to differ minimally from the previous PKT language, but to introduce consonants that are subject to OCP. For this reason, the velars are replaced by labial consonants that have a similar O/E ratio.

P	**P**	**T**
be	ma	da
bu	mi	du
fo	vo	se
fi	vy	sy

Table 14. The CV syllables of the PPT language.

The same coronal (T), and labial (P) consonants are used as in the previous experiment: the /d,s, f, m/. The resulting selection is given in Table 15a. Below it, in Table 15b, the PKT language from Experiment 1 is given for comparison.

PPT –Experiment 2					
P_1VP_2	O/E	P_2VT_3	O/E	T_3VP_1	O/E
fv	0.13	vs	0.65	sf	1.55
fm	0.62	vd	0.51	sb	1.40
bv	1.53	ms	1.11	df	0.53
bm	0.57	md	1.83	db	0.96
average	**0.71**		**1.03**		**1.11**

Table 15a. Consonant O/E ratios for PPT language.

PKT – Experiment 1					
P_1VK_2	O/E	K_2VT_3	O/E	T_3VP_1	O/E
fx	0.17	kd	0.37	sm	1.81
fk	0.81	xs	1.16	sf	1.55
mk	0.76	xd	2.02	dm	0.91
mx	0.65	ks	0.66	df	0.53
average	**0.60**		**1.05**		**1.20**

Table 15b. Consonant O/E ratios for the PKT language. 15a+15b enable comparison of the average O/E ratio values of the PKT and PPT language.

As can be seen in Table 15a and b, the O/E ratios between the positions are similar. Unfortunately, the labials /f/ and /m/ could not be in the same slot as they were in the PKT language. A PPT language with /f/ and /m/ in the same slot could not be statistically controlled for both cohort density and transitional probability. As a result, the consonants /f/ and /m/ were assigned into different slots.

As noted in Chapter 2, in Dutch labials show a preference to appear in word initial position. This was controlled as much as possible, but the labial consonants show a higher word-initial frequency. In the previous experiment, however, a clear preference was found for the coronal consonants /d, s/ word-initially, though the O/E ratio in the lexicon predicted on the contrary. In this fashion, the over-representation of word-initial labials is accepted in the AL, but will be examined for possible influence in the results using regression analysis.

The syllables in the PPT language, as given in Table 14 (above) are examined for their occurrence in specific word-positions in Dutch. Table 16 shows that all syllables are preferred in word-medial position. The pattern of preference is similar for the syllables in position 1, 2, and 3.

	word position	summed types	average	percentage	favours parsing:
Position 1 (P) {be,bu,fo,fi}	initial	82	21	0.24	PPT
	medial	250	63	0.72	TPP
	final	15	4	0.04	PTP
Position 2 (P) {ma,mi,vo,vy}	intial	125	31	0.22	PTP
	medial	370	93	0.66	PPT
	final	65	16	0.12	TPP
Position 3 (T) {da,du,se,sy}	initial	94	24	0.32	TPP
	medial	168	42	0.57	PTP
	final	34	9	0.11	PPT

Table 16. Percentage of the AL syllables in various word-positions in Dutch. For type counts per syllable, see appendix B.

Additionally, if the syllable position counts that favour a particular segmentation are summed, the PTP-segmentation, which is predicted by the hypothesis, scores lowest: works contrary the prediction (Table 17)

Syllable type counts that favour a PPT segmentation:	486
Syllable type counts that favour a PTP segmentation:	308
Syllable type counts that favour a TPP segmentation:	409

Table 17. Summation of syllable counts that lead to a specific segmentation.

Per consonant, two vowels were selected from the available Dutch vowels: /a, e, i, o, u, y/. Each consonant was assigned a short (/i, u, y/) and a long vowel (/a, e, o/), for an equal distribution. To avoid a lexical preference due to frequent CVC combinations in the lexicon, the cohort density of the sequences is calculated and used to select the vowels. Table 18 shows the cohort density values of the selected CVC sequences.

Position 1 C_1VC_2	**Cohort Density**	**Position 2 C_2VC_3**	**Cohort Density**	**Position 3 C_3VC_1**	**Cohort Density**
bem	5.27	mas	6.5	syf	0.7
bev	19.53	mid	2.02	sef	2.94
bum	9.78	mad	23.67	syp	45.1
buv	8.5	mis	11.05	sep	10.36
fov	0	vys	0.3	dap	0
fom	0	vyd	0	daf	0
fim	0	vod	1.38	dup	0
fiv	0	vos	2.36	duf	0
Average value	5.39		5.91		7.39

Table 18. Cohort density values for the PPT consonant vowel sequences.

The cohort density values for the CVC sequences are compared using a Students' t-test, and were found not to be significantly different ($p > 0.32$).

A final control was a comparison of the transitional probabilities (TP) of the biphone frequency. The comparison of the individual biphone TP values of non-words, using a Students' t-test showed no significant difference between the possible parsings ($p > 0.24$).

An automated word check was run to ensure no Dutch words could be found in the PPT language: no matching CVCV, CVCVC, or CVCVCV words were found in Dutch (see Appendix A for all the words in the PPT language).

The language was synthesised in the same way as in the PKT experiment.

Procedure

The procedure is the same as the previous experiment. Subjects were tested in a soundproof booth, the language stimuli were played over headphones, at a comfortable volume. Subjects listened to an artificial language for 10 minutes, and were then presented with an auditory forced choice test.

3.2 Results

The prediction for this experiment was that segmentation would occur between the PP-sequences, caused by the similarity constraint. A preference for PTP-words over PPT and TPP was expected. The results, however, again show a preference for a TPP segmentation. When entered in a binomial test, the responses show a similar pattern to the previous experiment:

	N (response)	preference	p (asymptotic significance two tailed test)
PPT vs. PTP	147 – 109	57% - 43%	0.021
PTP vs. TPP	92 – 164	36% - 64%	0.000
PPT vs. TPP	90 – 166	35% - 65%	0.000

Table 19. Results of the PPT 1 experiment

Not only does this experiment show a preference for TPP, the predicted segmentation PTP, again scores lowest overall (PTP N(total)=201; PPT N(total)=237, TPP N(total)=330). The means of the subject responses is given in Figure 6.

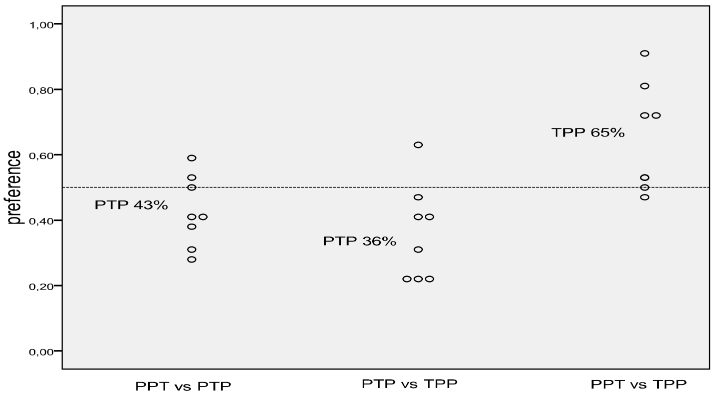

Figure 6. Results of the PPT language.

3.3 Discussion

Contrary to the hypothesis, a TPP segmentation is preferred over both PPT and PTP. If a confound is assumed to cause a preference for T-initial, however, that does not explain why the PTP segmentation scored lower than the PPT: the results show that in direct competition, a PPT segmentation is preferred (57% $p = 0.021$) over PTP.

In the previous experiment, the scores of the stimuli per initial consonant (e.g. sum of all scores on /m/-initial test words) were seen irregular for the labial /f,m/ consonants. The means of the PPT stimuli scores are given in Table 20.

P (position 1)	average score	P (position 2)	average score	T (position 3)	average score
f	0.51	v	0.39	s	0.66
b	0.42	m	0.42	d	0.63

Table 20. Average scores of test words per initial consonant.

It seems that the preference for T-initial consonants cannot be attributed to only the [d] or the [s]. Among the labials, the diversity is slightly higher. In comparison to the previous experiment, the diversity among the labials as compared to the coronals is not as striking, but it still suggests that perhaps the perceived phonological similarity among the labial consonants is lower than among the coronal consonants.

The influence of memory of the words in the AL is assumed to recede during testing, but when examining only the first sixteen (of thirty-two) responses in a binomial test, the same pattern in responses is found: TPP is clearly preferred, and PTP scores lowest (PPT vs.

PTP 59% - 41% (p=.042), PTP vs. TPP 34%- 66%(p=.001), PPT vs. TPP 39% - 61% (p=.017).

Compared to the PKT experiment, the preference for a T-initial segmentation is slightly less pronounced. Perhaps this may be attributable to a small influence of similarity avoidance, but this may also be caused by the high O/E values for the labial consonants /b,v/ used in this PPT experiment. A regression analysis was conducted to search for lexical influences that may explain the results.

test pairs	**Group scores**	**type**	**TP**	**CD**	**.C O/E**	**.C relfreq**	**final V**	**init CV**	**fin CV**
PPT vs PTP	57% vs 43% (p=0.021)	p=0.009							
PTP vs TPP	36% vs 64% (p<0.001)	p=0.000							
PPT vs TPP	35% vs 65% (p<0.001)	p=0.040						p=0.005 *	p=0.006

Table 21. Regression results of the PPT-language. An asterisk (*) indicates an inverse relation. TP= transitional probability, CD= cohort density, .C O/E= consonant word-initial O/E ratio, .Crelfreq= consonant word initial relative frequency, finalV= final vowel, initCV= initial syllable, finCV= final syllable.

In a stepwise regression analysis, no consistent evidence was found for a lexical influence on the data (Table 21). Though participants reported having heard many T-initial words, the occurrence of /d,s/ in the Dutch lexicon in word-initial position did not show an influence. The relative frequency of the consonants is also considered, since both T consonants are very frequent in Dutch; this information is not computed in O/E. However, no influence was found.

In the previous experiment, the final vowel /i/ was suspected to have an influence (it favoured a TPK segmentation), but was not found in a regression analysis. In this experiment, the final /i/ did not favour the TPP segmentation, but a TPP segmentation is still found. Thus, the final vowel /i/ alone does not cause a T-initial preference. However, the final and initial syllable show an influence. In the means of the test-stimuli, syllables with /mV/ score very high, which correlates with the finding for syllable positions. However, /bV/ syllables show a high occurrence in word-initial position in the Dutch lexicon, but correlate with a low preference in the experiment. Possibly this can be explained by the fact that the high /bV/-initial count is largely caused by the morphological prefix /b@-/. The high count for /mV/ is not caused by morphology, and may therefore be more salient to learners. On the other hand, because there is no influence of the consonant-initial count, this is a tricky point to argue. Ultimately, the influence of lexical distribution on the perceived segmentation is not straightforward. Perhaps there are phonological influences such as stress pattern that causes preference for a syllable.

In this experiment, there is no evidence of a similarity constraint influencing word segmentation. However, there is evidence, in both the PPT and the PKT language, of a preference of T-initial segmentation. Since no lexical factors were seen to influence the results consistently, the explanation should be sought outside of lexical influence. The notion of phonological similarity as discussed in Onnis et al. (2005) seems most plausible for explanation. In their experiment, the ability of subjects to detect non-adjacent dependencies depended crucially on phonological similarity. This raises the question whether in this experiment the labial consonants may not have been perceived as phonologically similar, and thus 'derailing' the segmentation rhythm. The odd one out in the consonant set is the nasal [m]. If the [m] is not perceived as phonologically similar to the [f,b,v] consonants, this causes an a-symmetrical language, causing segmentation to prefer the phonological similar set of coronals [d,s] word initially. In this fashion, the absence of the predicted PTP-preference can be explained, and the preference for TPK segmentation.

To ensure a statistically controlled AL, the /m/ was the only possible labial consonant in the PKT language. To keep the PPT language minimally different from the PKT language the /m/ consonant had to be retained. However, this particular consonant may cause an unexpected result (namely a preference for T-initial). As a result, the next experiment tests the hypothesis whether the nasal [m] may have caused a the T-initial preference, and, simultaneously, whether its exclusion will show evidence of similarity avoidance in Dutch.

In this experiment, the consonants with similar place of articulation could not predict segmentation behaviour. Evidence for this constraint in Dutch, however, is provided by Kager & Shatzman (in progress) and Boll-Avetisyan & Kager (2008). The next experiment (Experiment 3) intends to test whether the presence of the /m/, as a nasal, may have influenced segmentation. It is hypothesised that /m/ was not perceived as phonologically similar to the other labial (/v/) in the same position in the AL. As a result of this, the segmentation rhythm is disrupted and leads to a preference for the coronal consonants, which are perceived as phonologically similar. The illustration shows the asymmetrical distribution. This may lead to a preference for initial T consonants, since when three P consonants may be grouped together, excluding the nasal, this may have resulted in a very irregular starting point for segmentation.

$P_{continuant}$	$P_{continuant}$	$T_{continuant}$
$P_{plosive}$	$P_{\textbf{nasal}}$	$T_{plosive}$

Pierrehumbert (1993: 9-10) assumes that, in speech segmentation, to discriminate between phonemes more distinctive features are considered before less distinctive ones. In this way, sonority and major place of articulation features are first checked to be distinctive and then, if necessary, secondary place features are considered. If many phonemes belong to the same class, they will need to be distinctive in secondary features. In Arabic, Pierrhumbert says, there are fourteen coronals, and only three labials. Since the class of the coronals is larger than the class of labials, more features need to be considered in order to distinguish between the coronals, than for the labial class. As a result, secondary features become more important for contrast when the size of a class is bigger.

She explains that the /f/ and /m/, though sharing [labial], differ with respect to sonority. By comparison, the /s/ and the /n/ also share place of articulation [coronal] but differ in many features needed to distinguish them from all the other near relatives in Arabic such as /š/, /t/, /l/. In this way /f/ and /m/ are more similar than /s/ and /n/.

Dutch has six labial consonants /p,b,f,v,m,w/, and the nasal-oral contrast (enabling the contrast in sonority) is necessary to distinguish the stops /b,m/. Frisch et al. (2004) note the influence of the size of the consonant inventory, and show that in their similarity measure (given in (11a), above) the similarity of /f,m/ decreases when the consonant inventory increases. In a larger natural class, smaller sub-classes are needed for contrast between the individual phonemes: in this way, the nasal may be perceived as belonging to a different sub-class of labials due to its nasal characteristic.

Shatzman & Kager (2007) use the labials /p,m,f/ and the coronals /t,n,s/, in this way providing a balance in the features of the consonants. Boll-Avetisyan & Kager (2008) use /p,b,m/ and /t,d,n/, which is also neatly balanced for features. Though the consonants in the previous two experiments shared place of articulation, they may still have shown too much variation in other features, most notably the nasal feature of the /m/.

Following Onnis et al. (2005), phonological similarity is necessary for segmentation of non-adjacent dependencies, and may also be assumed to play a role for the constraint on similarity. The current experiment sets out to determine why the previous PPT experiment did not find a constraint on similarity. The assumption is that, due to a low degree of similarity between the labials, where /m/ was the odd syllable, a similarity effect was not found. In this way, since the labials were an a-symmetrical group (one and three: /m/ - /v,b,f/, segmentation preferred the phonological similar pair of T-consonants /d,s/. The effect of the /m/ on segmentation is tested in Experiment 3.

4. Experiment 3 (PPT): the Role of Phonological Similarity for Segmentation

This experiment tests a second type of PPT language (PPT-2), but with the crucial difference from the previous experiment that the language does not contain an /m/. The /m/ is replaced by the labial plosive /p/, resulting in a more balanced (and thus presumed more phonologically similar) grouping of labials. In this way the labials will be perceived as belonging to the same group. Due to the similarity constraint an avoidance of labial consonant sequences (*PP) is predicted.

4.1 Method

Participants

Forty-two Dutch native speakers were recruited for this experiment. They were paid 5 euros for participation. All were students at Utrecht University. None reported having hearing difficulties.

Group 1		Group 2		Group 3	
male:	5	male:	3	male:	4
female:	9	female:	11	female:	10
age (years; months):	21;5	age (years; months):	21;6	age (years; months):	22;1

Table 22. Participants of Experiment 3 (PPT-2)

Material

The syllables that make up the PPT-2 language are given in Table 23.

P	**P**	**T**
ve	fo	se
vu	fu	si
po	ba	da
py	bi	dy

Table 23. Syllables of PPT-2. The background indicates the distribution of long and short vowels.

The [m] from the previous PPT language is replaced by the [p]. A problem, however, is that labials are over-represented in word initial position. In this experiment, the O/E ratios of the P-consonants in initial position is the double of the T-consonants. However, from the previous experiments the initial O/E ratio was not seen to exert a visible influence: both times the highest initial O/E ratio (/k,x/ and /b,m/ were preferred least in initial position; the T-consonants (/d/ and /s/) were preferred most. Post-test regression analyses did also not indicate influence of initial O/E.

Table 24 shows the O/E ratios of the consonant sequences. The PP sequence is clearly under represented.

PPT2					
PP C_1VC_2	O/E	PT C_2VC_3	O/E	TP C_3VC_1	O/E
vf	0.15	fd	0.49	dv	0.65
vb	0.19	fs	0.91	dp	0.63
pf	0.22	bd	1.64	sv	0.91
pb	0.71	bs	1.11	sp	1.86
avg	0.32		1.04		1.01

Table 24. Overview of the O/E ratios of PPT2-consonant sequences.

As seen in Table 24, the PP sequences have a low mean value. In the previous experiments (PKT and PPT), the crucial comparison was based on a similar O/E ratio. Here, the influence of lexical distribution (as given by the O/E ratio) cannot be compared with the influence of the similarity constraint, since this language contains too many differences with the PKT language (e.g. O/E is a great deal lower than in the previous experiments (0.32 (PP) vs 0.60 (PK) and 0.71 (PP)). As a result, segmentation is predicted between the PP consonants, but the effects of lexical distribution (O/E) and OCP cannot be told apart. In this experiment, the crucial comparison is with the previous PPT language concerning the effect of replacing /m/ by /p/.

The occurrence of the PPT-2 syllables in the Dutch lexicon is examined. Table 25 gives the observed counts (types) of the syllables in three word positions (word-initial, word-medial, word-final).

	word position	summed types	average	pct	favours parsing:
Position 1 (P) {ve,vu,po,py}	initial	92	23	0,32	PPT
	medial	182	46	0,64	TPP
	final	12	3	0,04	PTP
Postion 2 (P) {fo,fu,ba,bi}	initial	90	23	0,32	PTP
	medial	173	43	0,62	PPT
	final	16	4	0,06	TPP
Position 3 (T) {se,si,da,dy}	initial	124	31	0,22	TPP
	medial	377	94	0,68	PTP
	final	55	14	0,10	PPT

Table 25. Percentage of the AL syllables in various word-positions in Dutch. For type counts per syllable, see Appendix B.

The syllables have an apparently similar distribution, except for the T-syllables in medial position (caused by the syllable *si*). This may cause a confound, especially since it results in a preference for the PTP sequence (table 26), which is predicted by the hypothesis. However, since the preference is caused by a single syllable *si*, analysis of the results should be able to detect if this particular syllable had an influence.

Syllable type counts that favour a PPT segmentation:	320
Syllable type counts that favour a PTP segmentation:	479
Syllable type counts that favour a TPP segmentation:	322

Table 26. Summation of syllable counts that lead to a specific segmentation.

The PPT-2 syllables are controlled for a lexical preference by comparing their cohort density values (Table 27).

Position 1 C_1VC_2	**Cohort Density**	**Position 2 C_2VC_3**	**Cohort Density**	**Position 3 C_3VC_1**	**Cohort Density**
veb	1.92	bid	11.59	sip	10.85
vuf	0	bad	10.74	sep	10.36
vub	0	bas	20.78	sev	0
pyb	60.44	fus	0	dav	9.15
pyf	0	fos	0	dyp	8.74
pof	0	fod	0	dyv	5.37
pob	0	fud	2.98	dap	0
average	8.25		7.49		7.08

Table 27. Selected consonant-vowel pairs, with their respective summed log frequencies.

The cohort density values were compared using a Student's t-test, and did not yield a significant difference of cohort density ($p > 0.65$). Comparison of the transitional probabilities (TP) of the biphones, shows that the PPT-2 language can be parsed three different ways without significant difference ($p > 0.16$).

Finally, the CVCVCV non-words were cross checked for existing words in Dutch. When only considering words of more than two syllables, the PPT2 language did not contain any Dutch words.

The language was synthesised in the same way as in the PKT and PPT1 experiment.

Procedure

The procedure was the same as the two previous experiments.

4.2. Results

The prediction for this experiment was that the similarity constraint would cause avoidance of PP sequences. The similarity constraint was expected to show influence, since the labial consonants were balanced in manner of articulation. In a binomial test, however, the results show no preference for a particular segmentation.

	N	preference	p (asymptotic significance two tailed)
PPT vs. PTP	216- 232	52% - 48%	0.479 ns
PTP vs. TPP	222 - 226	50% - 50%	0.887 ns
PPT vs. TPP	234 - 214	52% - 48%	0.369 ns

Table 28. Results of the PPT2 experiment, 'ns' stands for not significant.

When summing the responses on each particular parsing, it can be seen how close together the preference scores are: PTP N(total)=454; PPT N(total)=450;TPP N(total)=440). The means of the subject responses is given in Figure 7.

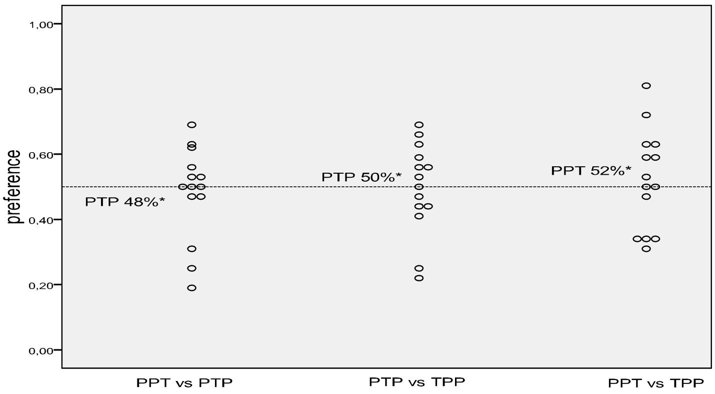

Figure 7. Results of PPT2. Each dot represents the mean score of a subject.

4.3. Discussion

In this experiment, there is no preference for any segmentation. Two things are most striking: (i) there is no PP avoidance, and hence no evidence for similarity avoidance, and (ii), there is no preference for T-initial segmentation TPP.

First the scores of the first sixteen (out of thirty-two) forced-choice tests are inspected. This shows a slight preference for PPT: (PPT vs. PTP 57% - 43% (p=0.052), PTP vs. TPP 51% - 49% (p=0.738); PPT vs. TPP 57% - 43% (p=0.038). This suggests that there may have been a preference for labial initial segmentation in the AL, but that this recedes during testing. A preference for labial initial correlates with their lexical properties in Dutch: labial consonants have higher O/E ratios for initial occurrence than the coronal consonants, the syllable positions predict a labial initial segmentation, and maybe the single high value for the cohort density of /pyb/ (which occurs in *publiek* 'public', *publiceren* 'publish' etc.), may have stood out. Though this seems plausible, it does not explain why the preference would then decline during the test phase. If anything, the lexical properties of Dutch should be less apparent during the first test phases, since the subjects have been listening to a ‘foreign’ language, and may be less inclined to apply Dutch lexical statistics to the test words.

When looking at the scores for test-words with a particular consonant, no clear preference is found. Remarkably, however, there is still a larger margin of difference between the labial consonants than between the coronals. In the previous experiments this was taken as an indication that /m/ caused an asymmetry, but here the same pattern arises without /m/. However, since there was no preference for either segmentation, and did not differ from chance, an element of guessing must be assumed. The scores seen in (28) may thus not be clearly indicative of segmentation behaviour.

P (pos 1)	mean score	P (pos2)	mean score	T (pos3)	mean score
v	0.57	f	0.52	s	0.53
p	0.47	b	0.45	d	0.46

Table 28. average scores of the stimuli beginning with a specific consonant.

Since the only difference from Experiment 2 (PPT) was the replacement of /m/ by /p/ , the results suggest that the /m/ had an impeding effect on P-initial segmentation: there is no more preference for initial T-consonants. Due to increased phonological similarity, the language can be parsed with all sets in initial position; without a preference for any parsing.

However, the results do not show evidence for a constraint on similarity. This may be because it is still overruled by an unknown factor that caused the preference for TPK in the first experiments. On the other hand, if that T-initial preference is taken to be caused by the absence of phonological similarity among labials (since, in this experiment, with phonological similar labials there is no T-initial preference), then it is unclear why there is no OCP effect.

Of course, the subjects may not have heard any segmentation and guessed during the forced choice test. However, when asked after the experiment, many subjects reported to have heard a rhythm, but having difficulty choosing between two non-words in isolation. This indicates that in fact segmentation did take place, but without any preference for a particular segmentation. This also indicates that there may be a problem with the forced choice test paradigm. Some subjects reported (three in the PPT-2 experiment) to have heard many e.g. /b/-initial words, but when their individual responses were examined they showed no clear preference for /b/ words. The words presented in isolation were perhaps too different in this context. From this line of reasoning it follows, however, that test subjects may not have had a strong enough memory trace of the words: that the language may have been too large to store word forms. Either a longer exposure time may be necessary, or a smaller non-word set.

For a closer look into the results a regression (stepwise) is performed. As seen in Table 29, the lexical influences vary. An analysis presents a fragmented picture: for example, in PPT vs. PTP, there is a relation between the response and the final syllable, but simultaneously there is an inverse relation with final vowel. Remarkably, there is no influence of the high O/E ratio for initial labial consonants. Though there are lexical influences found, there is no dominant lexical cue.

Regression analysis PPT 2 (N subjects = 42)									
Group	Group scores	Type	TP	CD	.C O/E	.C relfreq	final V	init CV	final CV
PPT vs PTP	52% vs 48% (p=0.479)						p=0.012 *		p=0.003
PTP vs TPP	50% vs 50% (p=0.887)							p=0.037	
PPT vs TPP	52% vs 48% (p=0.369)		p=0.000	p=0.009 *				p=0.012 *	

Table 29. Regression results of the PPT2-language. An asterisk (*) indicates an inverse relation. TP= transitional probability, CD= cohort density, .C O/E= consonant word-initial O/E ratio, .Crelfreq= consonant word initial relative frequency, finalV= final vowel, initCV= initial syllable, finCV= final syllable.

In conclusion, the PPT-2 experiment showed that when the manner of articulation of consonants is equally distributed, the preference for T-initial segmentation disappears. In this way, the /m/, though belonging to the natural class of labials, was not grouped with the other labials /f, v,b/. This supports the view that features are important during word segmentation. On the other hand, the balancing of consonants for manner of articulation did not result in evidence for the constraint on similarity. On account of the strong preference for T-initial in the previous experiments, there may be a confound (acoustic or phonological) that favours T-initial segmentation, cancelling out the similarity effect. In a comparison of PPT vs. PTP segmentation, however, there is no evidence for similarity avoidance.

5. General Discussion & Conclusion

In this thesis, the assumption was that constraints on learning form, or at least influence, the patterns in language. A constraint on similar consonants occurring together has been attested in several languages, including Dutch, and this constraint has had an influence on the lexicon: the frequency of similar consonants occurring in sequence is low. On the other hand, the frequency of patterns in a language has been shown to influence processing: a pattern with a higher frequency in the lexicon will be easier to process. Especially if constraints are assumed to be based on language input (in either a gradient manner, or by stochastic interaction of constraints), the frequency of patterns in the language in itself give rise to a constraint: a constraint based on frequency of occurrence. As a result, the two constraints conform and can be hard to tell apart.

Two artificial language learning experiments were conducted in order to distinguish between a phonotactic constraint on similarity and a constraint based on lexical distribution. Both artificial languages had matching lexical distribution values, but were crucially different in that the second language contained similar consonants, which could be subject to the phonotactic similarity constraint (i.e. comparing segmentation of an accidental gap with a systematic gap). The first experiment, examining the effect of lexical distribution, found that the lexical properties of non-adjacent consonants C_1VC_2 could not predict segmentation. The second experiment examined the effect of the similarity constraint, but failed to find an influence of the consonant similarity. Subsequently a third experiment was performed to establish whether the dissimilarity of features of the consonants had influenced segmentation behaviour, and also caused the absence of the similarity constraint. In this experiment, segmentation behaviour was more regular, but did not show an effect of the constraint on consonant similarity.

The experiments show no effect for lexical distribution on segmentation, and no evidence for a constraint on homorganic consonants. The low O/E ratio of the PK consonants (0.60) as compared to the KT (1.05) and TP (1.20) consonants was predicted to give rise to a word boundary. Hence, a preference for KTP words was predicted. However, the results showed a very strong preference for TPK words; KTP words were preferred least. This is taken as evidence that lexical distribution did not influence segmentation. In Experiment 2, the similarity of the PP sequence was predicted to give rise to a word boundary. The O/E ratios of the sequences were matched with the PK sequences, so that avoidance of a PP sequence could only be caused by the similarity constraint. In this way a PTP segmentation was predicted. However, the results showed a preference for TPP words, and PTP words scored lowest. As a result, there was no evidence for a constraint on consonants with similar place of articulation.

The strong preference for T-initial words in both languages is remarkable. It cannot be explained by their lexical distribution in Dutch (as shown by multiple regression analyses), nor does the acoustic signal show any irregularities (duration of consonants was balanced). However, many subjects reported having heard many /d,s/ (T), suggesting that they were somehow perceptually more salient. (The quality of the consonants as synthesised by the programme Fluency may have had an influence, and a different synthesis should be tested.) Not only is the T-initial preference remarkable, it is also surprising that there was no effect of the similarity constraint, which is attested in various languages, including Dutch, and which has previously been shown to influence segmentation (Kager & Shatzman in progress; Boll-Avetisyan & Kager 2008). These facts taken together suggested an unknown confound in the languages, and a third experiment was needed to test this hypothesis.

Experiment 3 tested whether the similarity constraint could predict segmentation. The variable under investigation was replacing the /m/ of the PPT-1 language with /p/. The nasal features of /m/ were assumed to cause /m/ to be regarded as separate (as a sub class) from the other labials: in this way causing an asymmetry in the language among the labial consonants and giving rise to a preference for the T consonants. If the labials were perceived as similar, the similarity constraint would apply. A PTP preference was predicted. The results, however, showed no preference for any segmentation. Though the T-initial preference was not found, there was no evidence for the constraint on similarity.

The PPT languages did not show evidence of the similarity constraint. It is important to determine what the differences are with the studies by Boll-Avetisyan & Kager (2008), who did find an effect of the constraint in speech segmentation. In their study, the same labial consonants appear in position 1 and position 2: that is, P_1= /p, b, m/ and P_2 =/p, b, m/. Clearly, this contributes to perceived similarity. Additionally, the features nasal and voice are evenly distributed, yielding only the crucial difference of place of articulation. When compared with the languages in the present study (Table 30), the consonants do not show such a regular pattern.

	PPT (B&K 2008)						PPT						PPT-2					
consonant	p	b	m	t	d	n	b	f	m	v	d	s	p	v	b	f	d	s
place	P	P	P	T	T	T	P	P	P	P	T	T	P	P	P	P	T	T
voice	-	+	+	-	+	+	+	-	+	+	+	-	-	+	+	-	+	-
stop	+	+	+	+	+	+	+	-	+	-	+	-	+	-	+	-	+	-
continuant	-	-	-	-	-	-	-	+	-	+	-	+	-	+	-	+	-	+
nasal	-	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-

Table 30. overview of relevant features of the PPT languages.

In the table it can be seen that, in terms of the given features, the consonants in the AL of Boll-Avetisyan & Kager (2008) had one main contrast, which was place of articulation. In the PPT language, /m/ clearly stands out from the rest of the consonants due to its nasal feature. But also in the PPT-2 language, the contrast between the consonant features may have played a role, they are not as evenly balanced as in the Boll-Avetisyan Kager experiment: features could also contrast (e.g. /p/ - /d/ : voice, place contrast; /v/-/d/ : continuant voice contrast). As a result, in the present study, the contrasting feature between coronals and labials was not purely place of articulation, as it was in Boll-Avetisyan & Kager (2008). This may have caused a decrease in the strength of the similarity avoidance constraint, which is a subtle cue, but it seems unlikely that the total absence of the similarity constraint can be attributed to unbalanced features.

A second difference is the size of the language. In the Boll-Avetisyan & Kager study the AL contains 27 words (81 for different parsings). The language in the current experiments counted 64 (192 for different parsings). During a 10 min. exposure, subjects in the current study will have heard a word a maximum of fourteen times, where in the Boll-Avetisyan & Kager study they will have heard a word a maximum of 33 times: more than double. In this fashion, the memory trace may have been so weak that it disappeared as soon as words in isolation were presented. For instance, two subjects in the PPT-2 experiment reported hearing initial /f, b/ (a PTP segmentation), but their responses did not reflect this preference. In this fashion the forced choice paradigm, even when examining the first sixteen responses may not have reflected segmentation. If there is a possible discrepancy between learning and reproduction, then it remains unexplained why T-initial words were learned. In theory, remembering only the first consonant of words could indicate segmentation preference (e.g. in PPT-2, an initial /f/ yields a PTP preference). As a result, the perceptual salience of the coronals /d,s/ may well have determined the response preference in the first two test phases, regardless of the word structure (however, when the perceptual salience is attributed to the fact that they become salient exactly because they are favoured initially, this argument does not hold). To test whether memory of the word forms influenced the results, either a longer exposure to the AL may be necessary, or a smaller word set.

Another possible explanation for not finding similarity avoidance may be an acoustic factor. Although the intensity, duration, and pitch of the T consonants were matched with the P and K consonants, there may have been an unknown acoustic factor causing preference for initial /d,s/. If this was the case, then it may have cancelled out the similarity effect, which gave a null result. However, the first sixteen test trials in the PPT-2 experiment showed a tendency toward PPT in a PPT vs. PTP comparison. The assumption of an acoustic confound should therefore be verified by using a different programme to synthesize the AL.

On the other hand, it may be that similarity avoidance does not influence speech segmentation. Only the experiments by Boll-Avetisyan & Kager (2008) and an unpublished word-spotting experiment by Kager & Shatzman (in progress) provide evidence for the similarity constraint used for segmentation. Other studies, such as Berent & Shimron (2003), Frisch et al. (2004), Coetzee (2005), Shatzman & Kager (2007), show that similarity avoidance plays a role for acceptability and facilitates processing for of non-words presented in isolation, which shows preference for a word form, or word formation. Logically this preference can be useful in speech segmentation, but it may also only influence processes of word formation, and in this way diachronically give rise to the distinct lexical distribution pattern of under-representation. Coetzee (2005) shows that perception of speech sounds in a continuum are subject to similarity avoidance: in perception there is a bias towards the dissimilar speech sound. Due to processes of diachronic variation, it may be that words gradually evolve toward a maximally contrastive form. In this way, similarity avoidance may play a role in the diachronic process of language formation, and thus leaves a distinct lexical distribution pattern but does not actively contribute to speech segmentation.

5.1. Conclusion

The set of experiments reported in this thesis cannot answer the central question of whether there is a difference between the influence of lexical distribution and the influence of the phonotactic constraint on similarity. It did show the influence of phonological features for perceived similarity. Since similarity depends on the level of contrast, the phoneme inventory of classes influences the level of similarity between consonants. Apparently, the feature [nasal] set the labial consonant /m/ apart from the other labials in this study. Moreover, the notion of phonological similarity does seem to play an active role in language segmentation. In this study, the coronals were preferred in initial position as a group, since they were distinct from the other speech sounds (the labials) which were not grouped together, presumably because they differed too much internally. As a result the perceived similarity between the coronals may have lead them to be perceptually more salient. This process in turn may have facilitated the detection of regularities in the speech input, since the learner tracks a specific element in the input. Clearly, a focus on specific elements in the speech input enables the learner to detect regularities, and to extract any rules more efficiently. In this fashion, the phonological features of the phonemes in the language influence statistical learning, and demonstrates how statistical learning interacts with linguistic properties of language.

References

Albright, Adam (2006). “Gradient Phonotactic Effects: Lexical? Grammatical? Both? Neither?” *80th Annual LSAMeeting*, Albuquerque. 1-14.

Baayen, R.H., R. Piepenbrock, L. Gulikers (1995). *The CELEX Lexical Database (CD-Rom)*. Philadelphia, PA: University of Pennsylvania.

Berent, Iris, Joseph Shimron (2003). “Co-occurrence Restrictions on Identical Consonants in the Hebrew Lexicon: Are They Due to Similarity?” *Journal of Linguistics 39*. 31-55.

Boersma & Weenink (1996). *Praat version 5.1.04*. < http://www.fon.hum.uva.nl/praat/ >

Boersma, Paul, Bruce Hayes (2001). “Empirical Tests of the Gradual Learning Algorithm.” *Linguistic Inquiry, 32*. 45-86.

Boll-Avetisyan, Natalie, René Kager (2008). “The Obligatory Contour Principle in Artificial Language Segmentation.” *Presentation at Manchester Phonology Meeting*.

Bonatti, Luca L., Marcela Peña, Marina Nespor, Jacques Mehler (2005). “Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing.” *Psychological Science, 16-6.* 451-459.

Bryson, Bill (1990). *Mother Tongue*. Penguin Books: London.

Chomsky, Noam, Morris Halle (1968). *The Sound Pattern of English.* Harper & Row Publishers: London.

Coetzee, Andries W., Joe Pater (2006). “Lexically Ranked OCP-Constraints in Muna.” University of Michigan and UMass < http://roa.rutgers.edu/files/842-0606/842-COETZEE-0-0.PDF >

Coetzee, Andries W., Joe Pater (2008). “Weighted Constraints and Gradient Restrictions on Place Co-occurrence in Muna and Arabic.” *Natural Language Linguistic Theory* 26. 289–337.

Cole, Ronald A., Jola Jakimik (1980). “A Model of Speech Perception.” *Perception and Production of Fluent Speech Perception.* Ed. Ronald. A. Cole. Lawrence Erlbaum Associates, Publishers: New Jersey. 133-164.

Coleman, John, Janet Pierrehumbert (1997). “Stochastic Phonological Grammars and Acceptability.” *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology.* Association for Computational Linguistics: New Jersey. 49-56.

Côté, Marie-Hélène (2001). “Consonant Cluster Phonotactics: A Perceptual Approach.” PhD Dissertation, Massachusetts Institute of Technology. 1-345.

Cristià, Alejandrina, Amanda Seidl (2007). “Why Cross-Linguistic Frequency Cannot be Equated with Ease of Acquisition in Phonology.” *U.Penn Working Papers in Linguisitics, 10*. 1-12.

Dutoit, T., V. Pagel, N. Pierret, F. Bataille, O. Van der Vreken (1999). “The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes.” *Proc. ICSLP '96*. Philadelphia. 1393-1396.

Finn, Amy S., Carla L. Hudson Kam (2008). “The Curse of Knowledge: First Language Knowledge Impairs Adult Learners’ Use of Novel Statistics for Word Segmentation.” *Cognition, 108*. 477-499.

Frisch, Stefan A., Bushra Adnan Zawaydeh (2001). “The Psychological Reality of OCP-Place in Arabic.” *Linguistic Society of America 77-1*. 91-106.

Frisch, Stefan A., Janet B. Pierrehumbert, Michael Broe (2004). “Similarity Avoidance and the OCP.” *Natural Language & Linguistic Theory 22*. 179-228.

Hauser, Marc D., Elissa L. Newport, Richard N. Aslin (2001). “Segmentation of the Speech Stream in a Non-Human Primate: Statistical Learning in Cotton-Top Tamarins.” *Cognition, 78*. B53-B64.

Hayes, Bruce, Colin Wilson (2008). “A Maximum Entropy Model of Phonotactics and Phonotactic Learning.” *Linguistic Inquiry, 29*. 379-440.

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

Jusczyk, Peter W., Angela D. Friedrici, Jeanine M.I. Wessels, Vigdis Y Svenkerud, Ann Marie Jusczyk (1993). “Infants’ Sensitivity to the Sound Patterns of Native Language Words.” *Journal of Memory and Language, 32*. 402-420.

Kager, René, Keren Shatzman (in progress). “Consonant Identity Avoidance in Speech Segmentation.” Unpublished manuscript, Utrecht University.

Kawahara, Shigeto, Hajime Ono, and Kiyoshi Sudo (2006). “Consonant Co-occurrence Restrictions in Yamato Japanese.” *Japanese/Korean Linguistics, 14*. 27-38.

Kuo, Li-Jen (2008). “The Role of Natural Class Features in the Acquisition of Phonotactic Regularities.” *Journal of Psycholinguistic Research.* (published online) <http://www.springerlink.com/content/u4206j0n2853227k/>

Luce, Paul A., Nathan R. Large (2001). “Phonotactics, Density, and Entropy in Spoken Word Recognition.” *Language and Cognitive Processes 16*. 565-581.

Mattys, Sven L., Peter W. Jusczyk (2001) Phonotactic Cues for Segmentation of Fluent Speech by Infants.” *Cognition 78*. 91-121.

McCarthy, John J. (1986). “OCP-Effects: Gemination and Antigemination.” *Linguistic Inquiry, 17*. 207-263.

McQueen, James, M. (1998) “Segmentation of Continuous Speech Using Phonotactics.” *Journal of Memory and Language 39*. 21-46.

Newport & Aslin (2004): “Learning at a Distance: I. Statistical learning of non-adjacent dependencies.” *Cognitive Psychology 48*. 127-162.

Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen and Richard N. Aslin (2004). “Learning at a Distance II. Statistical Learning of Non-Adjacent Dependencies in a Non-Human Primate.” *Cognitive Psychology, 49*. 85-117.

Onishi, Kristine H., Kyle E. Chambers, Cynthia Fisher (2002). “Learning Phonotactic Constraints from Brief Auditory Experience.” *Cognition, 83.* B13-B23.

Onnis, Luca, Padraic Monaghan, Korin Richmond, Nick Chater (2005) “Phonology Impacts Segmentation in Online Speech Processing.” *Journal of Memory and Language, 53.* 225-237.

Peña, Marcela, Luca L. Bonatti, Marina Nespor, Jacques Mehler (2002). “Signal Driven Computations in Speech Processing.” *Science, 298.* 604-607.

Pierrehumbert, Janet (1993). “Dissimilarity in the Arabic Verbal Roots.” *Proceedings of the Annual Meeting of the North Eastern Linguistics Society, 23.* 367–381.

Saffran, Jenny R. (2002). “Constraints on Statistical Language Learning.” *Journal of Memory and Language, 47.* 172-196.

Saffran, Jenny R., Elissa L. Newport, Richard N. Aslin (1996). “Word Segmentation: The Role of Distributional Cues.” *Journal of Memory and Language, 35.* 606-621.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, Elissa L. Newport (1999). “Statistical Learning of Tone Sequences by Human Infants and Adults.” *Cognition, 70*. 27-52.

Saffran, Jenny R., Erik D. Thiessen (2003). “Pattern Induction by Infant Language Learners.” *Developmental Psychology, 39-3.* 484-494.

Seidl, Amanda, Eugene Buckley (2005). “On the Learning of Arbitrary Phonological Rules.” *Language Learning and Development, 1.* 289-316.

Shatzman, Keren, René Kager (2007). “A Role for Phonotactic Constraints in Speech Perception.” In: J. Trouvain and W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 1409-1412. Dudweiler, Germany: Pirrot.

Steriade, Donca (1999). “Alternatives to the Syllabic Interpretation of Consonantal Phonotactics.” In: O.Fujimura B.Joseph and B.Palek (eds.) *Proceedings of the 1998 Linguistics and Phonetics Conference*. Prague: The Karolinum Press. 205-242.

Toro, Juan M., Marina Nespor, Jacques Mehler, Luca L. Bonatti (2008). “Finding Words and Rules in a Speech Stream: Functional Differences Between Vowels and Consonants.” *Psychological Science, 19*. 127-144.

Toro, Juan M., Scott Sinnett, Salvador Soto-Faraco (2005). “Speech Segmentation by Statistical Learning Depends on Attention.” *Cognition, 97.* B25-B34.

Vitevitch S. Michael, Paul A. Luce (2005). “Increases in Phonotactic Probability Facilitate Spoken Nonword Repetition.” *Journal of Memory and Language 52.* 193-204.

Vroomen, Jean, Jyrki Tuomainen, Beatrice de Gelder, B. (1998). “The Roles of Word Stress and Vowel Harmony in Speech Segmentation.” *Journal of Memory and Language, 38*. 133-149.

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

Zwitserlood, P., H. Schriefers, A. Lahiri, W. van Donselaar (1993). “The Role of Syllables in the Perception of Spoken Dutch.” *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19.* 260-271.

Appendix

Appendix A. Stimuli lists of the artificial languages

Language PKT

PKT		KTP		TPK	
mexisy	mexody	xisyfa	kedyme	symexi	dymexo
mexosy	mekiso	xosyfa	xodafu	syfaxi	sofaki
mekesy	mukida	xisyme	xisofu	symexo	symuxi
mekisy	fakeso	xosyme	kisome	syfuxi	dafuxo
fakesy	fakedy	kesyfa	kidyme	syfaxo	sofuki
faxisy	faxiso	kisyfa	xodamu	somexi	dameki
fukesy	faxida	xisyfu	kedafu	syfake	dafaki
fakisy	fukeso	kesyme	xisomu	syfuke	symuki
faxosy	fukedy	xodafa	xosofu	symeke	dafuki
fukisy	fakidy	xisofa	kedamu	syfuxo	symuxo
fuxisy	fakiso	xosyfu	kidafu	somexo	dymeke
fuxosy	faxoso	kedafa	xosomu	damexi	somuke
mekeda	fukidy	kisyme	kidamu	syfaki	somuxi
mekida	fukiso	xosofa	xidyme	syfuki	dymeki
mexoda	fuxiso	kidafa	xidafu	symeki	damuke
mukesy	fuxida	xisymu	xidamu	sofaxi	somuki
fakeda	mexidy	xodame	kesofu	damexo	somuxo
mukisy	fuxoso	xisome	xodymu	someke	damuxi
fukeda	faxody	kedame	kesomu	dafaxi	damuki
fakida	mukeso	xosymu	kisofu	sofuxi	dyfaxi
fukida	mukedy	xosome	kedymu	sofaxo	damuxo
mexiso	fuxody	kidame	kisomu	sofake	dymuke
mexida	muxoda	xidafa	xodyfa	dymexi	dyfuxi
muxisy	mukidy	kesyfu	kidymu	dafuxi	dymuxi
mexoso	mukiso	kesofa	kedyfa	sofuke	dyfaxo
faxoda	faxidy	kisyfu	kidyfa	dameke	dyfake
muxosy	fuxidy	kisofa	xidymu	dafaxo	dyfuke
mekeso	muxiso	xidame	xidyfa	someki	dymuki
mekedy	muxida	kesymu	xodyfu	sofuxo	dyfuxo
mukeda	muxoso	xodyme	kedyfu	dafake	dymuxo
fuxoda	muxody	kesome	kidyfu	symuke	dyfaki
mekidy	muxidy	kisymu	xidyfu	dafuke	dyfuki

Language PPT-1

PPT		**PTP**		**TPP**	
bemase	fomase	masebe	vosebe	dabema	sebema
bemasy	fomasy	masebu	vosebu	dabemi	sebemi
bemada	fomada	masefo	vosefo	dabevo	sebevo
bemadu	fomadu	masefi	vosefi	dabevy	sebevy
bemise	fomise	masybe	vosybe	dabuma	sebuma
bemisy	fomisy	masybu	vosybu	dabumi	sebumi
bemida	fomida	masyfo	vosyfo	dabuvo	sebuvo
bemidu	fomidu	masyfi	vosyfi	dabuvy	sebuvy
bevose	fovose	madabe	vodabe	dafoma	sefoma
bevosy	fovosy	madabu	vodabu	dafomi	sefomi
bevoda	fovoda	madafo	vodafo	dafovo	sefovo
bevodu	fovodu	madafi	vodafi	dafovy	sefovy
bevyse	fovyse	madube	vodube	dafima	sefima
bevysy	fovysy	madubu	vodubu	dafimi	sefimi
bevyda	fovyda	madufo	vodufo	dafivo	sefivo
bevydu	fovydu	madufi	vodufi	dafivy	sefivy
bumase	fimase	misebe	vysebe	dubema	sybema
bumasy	fimasy	misebu	vysebu	dubemi	sybemi
bumada	fimada	misefo	vysefo	dubevo	sybevo
bumadu	fimadu	misefi	vysefi	dubevy	sybevy
bumise	fimise	misybe	vysybe	dubuma	sybuma
bumisy	fimisy	misybu	vysybu	dubumi	sybumi
bumida	fimida	misyfo	vysyfo	dubuvo	sybuvo
bumidu	fimidu	misyfi	vysyfi	dubuvy	sybuvy
buvose	fivose	midabe	vydabe	dufoma	syfoma
buvosy	fivosy	midabu	vydabu	dufomi	syfomi
buvoda	fivoda	midafo	vydafo	dufovo	syfovo
buvodu	fivodu	midafi	vydafi	dufovy	syfovy
buvyse	fivyse	midube	vydube	dufima	syfima
buvysy	fivysy	midubu	vydubu	dufimi	syfimi
buvyda	fivyda	midufo	vydufo	dufivo	syfivo
buvydu	fivydu	midufi	vydufi	dufivy	syfivy

Language PPT-2

PPT		**PTP**		**TPP**	
vebisi	pybisi	bisive	fusive	sivebi	davebi
vebise	pybise	bisivu	fusivu	siveba	daveba
vebida	pybida	bisipy	fusipy	sivefu	davefu
vebidy	pybidy	bisipo	fusipo	sivefo	davefo
vebasi	pybasi	biseve	fuseve	sivubi	davubi
vebase	pybase	bisevu	fusevu	sivuba	davuba
vebada	pybada	bisepy	fusepy	sivufu	davufu
vebady	pybady	bisepo	fusepo	sivufo	davufo
vefusi	pyfusi	bidave	fudave	sipybi	dapybi
vefuse	pyfuse	bidavu	fudavu	sipyba	dapyba
vefuda	pyfuda	bidapy	fudapy	sipyfu	dapyfu
vefudy	pyfudy	bidapo	fudapo	sipyfo	dapyfo
vefosi	pyfosi	bidyve	fudyve	sipobi	dapobi
vefose	pyfose	bidyvu	fudyvu	sipoba	dapoba
vefoda	pyfoda	bidypy	fudypy	sipofu	dapofu
vefody	pyfody	bidypo	fudypo	sipofo	dapofo
vubisi	pobisi	basive	fosive	sevebi	dyvebi
vubise	pobise	basivu	fosivu	seveba	dyveba
vubida	pobida	basipy	fosipy	sevefu	dyvefu
vubidy	pobidy	basipo	fosipo	sevefo	dyvefo
vubasi	pobasi	baseve	foseve	sevubi	dyvubi
vubase	pobase	basevu	fosevu	sevuba	dyvuba
vubada	pobada	basepy	fosepy	sevufu	dyvufu
vubady	pobady	basepo	fosepo	sevufo	dyvufo
vufusi	pofusi	badave	fodave	sepybi	dypybi
vufuse	pofuse	badavu	fodavu	sepyba	dypyba
vufuda	pofuda	badapy	fodapy	sepyfu	dypyfu
vufudy	pofudy	badapo	fodapo	sepyfo	dypyfo
vufosi	pofosi	badyve	fodyve	sepobi	dypobi
vufose	pofose	badyvu	fodyvu	sepoba	dypoba
vufoda	pofoda	badypy	fodypy	sepofu	dypofu
vufody	pofody	badypo	fodypo	sepofo	dypofo

Appendix B. Syllable positions

Syllable positions language PKT:

syllable	position	types-tokens	types-types
me	initial	603	81
	medial	210	95
	final	9	8
mu	initial	104	13
	medial	69	24
	final	1	1
fa	initial	197	28
	medial	56	25
	final	1	1
fu	initial	20	9
	medial	6	4
	final	0	0
ke	initial	58	16
	medial	69	34
	final	7	4
ki	initial	88	25
	medial	44	26
	final	17	16
xo	initial	61	17
	medial	121	25
	final	0	0
xi	initial	67	12
	medial	31	16
	final	2	2
so	initial	139	27
	medial	135	59
	final	4	4
sy	initial	59	12
	medial	26	9
	final	2	2
da	initial	116	27
	medial	133	61
	final	13	13
dy	initial	43	11
	medial	170	32
	final	2	2

Syllable positions language PPT-1:

syllable	position	types-tokens	types-types
be	initial	151	17
	medial	72	37
	final	2	2
bu	initial	258	28
	medial	26	20
	final	3	3
fo	initial	103	11
	medial	82	43
	final	0	0
fi	initial	151	26
	medial	458	150
	final	71	10
ma	initial	312	56
	medial	439	174
	final	157	32
mi	initial	200	35
	medial	555	157
	final	75	24
vo	initial	219	22
	medial	101	37
	final	80	8
vy	initial	43	12
	medial	7	2
	final	2	1
se	initial	164	39
	medial	210	76
	final	15	14
sy	initial	59	12
	medial	26	9
	final	2	2
da	initial	116	27
	medial	133	61
	final	13	13
du	initial	45	16
	medial	32	22
	final	6	5

Syllable positions language PPT-2:

syll pos PPT 2	position	types -tokens	types- types
ve	initial	138	39
	medial	101	46
	final	6	3
vu	initial	77	12
	medial	107	30
	final	1	1
po	initial	273	31
	medial	182	83
	final	14	8
py	initial	37	10
	medial	82	23
	final	0	0
ba	initial	256	45
	medial	96	52
	final	4	4
bi	initial	125	25
	medial	259	74
	final	18	12
fo	initial	103	11
	medial	82	43
	final	0	0
fu	initial	20	9
	medial	6	4
	final	0	0
se	initial	164	39
	medial	210	76
	final	15	14
si	initial	227	47
	medial	578	208
	final	180	26
da	initial	116	27
	medial	133	61
	final	13	13
dy	initial	43	11
	medial	170	32
	final	2	2