Phonotactic Illegality and Probability in Speech Perception:
Evidence from second language listeners
Phonotactic Illegality and Probability in Speech Perception:
Evidence from second language listeners

Illegaliteit en waarschijnlijkheid van klankcombinaties in spraakwaarneming:
Evidentie van tweedetaalluisteraars
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op donderdag 22 december 2011 des ochtends te 10.30 uur

door

Tomas Ostar Lentz

geboren op 10 april 1981
te Zaanstad
Promotor: Prof. dr. R.W.J. Kager

Dit proefschrift maakt deel uit van het onderzoeksprogramma Phonotactic Constraints for Speech Segmentation: The Case of Second Language Acquisition, dat mede gefinancierd is door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).
5.1.1 Facilitation and inhibition ................................... 171
5.1.2 Gradience and categoricalness .............................. 173
5.2 Phonotactic knowledge in speech perception ............... 175
  5.2.1 A preparser for speech input ............................... 176
  5.2.2 Formalising the phonotactic grammar .................... 178
5.3 Second language listening ..................................... 186
5.4 Conclusion ......................................................... 187

A ADDITIONAL MATERIAL FOR CHAPTER 3 .......................... 189
  A.1 Stimuli ......................................................... 189
  A.2 Full statistical models ....................................... 189

B ADDITIONAL MATERIAL FOR CHAPTER 4 .......................... 201

REFERENCES ................................................................ 203

SAMENVATTING IN HET NEDERLANDS .............................. 219

CURRICULUM VITAE ...................................................... 225
The first and most read part of a thesis is bound to be the acknowledgements, which is unfortunately also the last and least carefully written part. This thesis has been written on the basis of the research I conducted within the project *Phonotactic Constraints for Speech Segmentation: The case of second language acquisition*, initiated and designed by René Kager. René’s broad interest in phonology and psycholinguistics and his openness to the combination of theoretical insight and empirical evidence has inspired the approach I took in my own research. I wish to thank René for his willingness to take the risk and hire me as a PhD researcher, although I was neither a phonologist nor a psycholinguist. I also greatly appreciate his patience with my often overambitious and vague writings, as well as his explanatory skills. René has been a great project leader and I will remember the inspiring group meetings almost as fondly as the mouthwatering group eatings (the name might be morphosemantically infelicitous, but was picked on its phonological property of rhyming with meeting; for semanticists and morphologists I would like to stress that the group was not eaten).

I am very glad to have worked in a project group with such a high degree of cooperation and friendship among its members. Frans Adriaans and Natalie Boll-Avetisyan have always provided me with the most ‘peer’ review one could wish to receive. Maybe just as important was our friendship: I always looked forward to having lunch, a party, a movie, a drink or a coffee with them, which made it all the more attractive to come to Utrecht every day. Frans, I really enjoyed discussing work practicalities, but even more work impracticalities and grand research plans. Some day, we will improve the whole of science with this one simple paper we have been discussing (I do hope you have kept all those brilliant notes we made on the back of old prints and beermats). Natalie, your relaxed attitude has rubbed off on me; without that, I would still have been thinking about experiments and reading more papers. I thank you for your fun-loving attitude to your life in Utrecht, but also for your helpful hints on relevant articles and books, as you had read much more relevant literature than I at the start of the project.

During the run of the project, many people joined us for a while. Marieke Kolkman left even before I joined, but I still wish to thank her for her legacy of papers and research tips. Keren Shatzman deserves more thanks than she probably thinks, because with her expertise in running experiments, she gave me many helpful hints and the confidence to start my own experiments. Diana, although you joined the project near the end, you managed to become
involved and provide valuable feedback, as well as being excellent company for coffee or beer. You also helped me to make the whole ‘becoming a doctor’ thing less of an issue. I also owe a lot to the students and/or assistants that helped our group over the time. Frits van Brenk, Joost Bastings, Tim Schoof, Bart Penning de Vries, Jorinde Timmer and Hayo Terband did small or large (mostly large) parts of the actual experiments. Xiaoli Dong, Ao Chen, Liquan Liu, Andréa Davis, Marko Simonović and Johannes Schliesser provided valuable hints (or even better: questions) when I presented my plans in the group meetings.

My research was funded by the Netherlands Organisation for Scientific Research NWO, project grant 277.70.001 to René Kager. I would like to thank not only the NWO, but all the institutions that allowed me to develop as a scholar (although I hesitate to call myself, or anyone, a scholar). These institutions rank from my primary school, the Burgvlietschool in Gouda, to my secondary school, the Coornhert Gymnasium in Gouda, to the University of Amsterdam and the Autonomous University of Barcelona. These institutions, and especially of course the teachers there, allowed me to develop my taste for knowledge and research. I know I am privileged to have been able to enjoy world-class education and I hope this kind of privilege will be more, not less, common in the future, for people of all backgrounds.

The Utrecht Institute of Linguistics OTS deserves a special place in the acknowledgements. I have been very rapidly introduced into many fields I did not know much about: into phonology by Silke Hamann, into statistics by Hugo Quené and into eye-tracking by Luisa Meroni, Arnout Koornneef and Iris Mulders. The availability of the UiL OTS lab, not only the equipment but, more importantly, its staff, has been more than helpful. I would like to especially thank Theo Veenker for help with the actual implementation of the new eye-tracking paradigm that I cooked up. I hardly ever had practical difficulties in the execution of my research and visiting conferences and organise workshops was never a problem. One never knows for sure who to thank for the absence of problems, but I know I owe a lot of thanks to Maaike Schoorlemmer, Martin Everaert and all the secretaries for the lack of ripples. Mariëtte Bonenkamp has been very helpful with the publishing process; for her it might be a routine job and it was very helpful for me to have an overview of all the necessary steps.

Even though my research group provided me with a group of people that are both colleagues and friends, I am glad that my ‘less close’ colleagues were the friendly bunch they were, showing interest in the work of others, willing to organise such great things as buying a foosball table and dinners and/or drink sessions. The tricky part in this part of the acknowledgements is that it is very likely that I will forget the name of some of my colleagues. If you are one
of those persons, know that I most likely did not really forget you, I am just hurrying the writing of these acknowledgements. Anyway, here goes: thank you Ana Aguilar Guevara, Daria Bahtina, Elise de Bree, Anneloes Canestrelli, Desiree Capel, Ao Chen, Anna Chernilovskaya, Alexis Dimitriadis, Xiaoli Dong, Lizet van Ewijk, Gaetano Fiorin, Gianluca Giorgolo, Nadya Goldberg, Bettina Gruber, Sander van der Harst, Nivja de Jong, Naomi Kamoen, Brigitta Keij, Annemarie Kerkhof, Cem Keskin, Anna Kijak, Loes Koring, Arnout Koornneef, Huib Kranendonk, Kiki Kushartanti, Bert Le Bruyn, Pim Mak, Sophia Manika, Marijana Marelj, Hannah De Mulder, Emilienn Ngangoum, Sieb Nooteboom, Rick Nouwen, Sandrien van Ommen, Andreas Pankau, Ingrid Persoon, Liv Persson, Eric Reuland and on a completely unrelated note, Sinterklaas (in case this book is read by children), Dagmar Schadler, Rianne Schippers, Marieke Schouwstra (now the guardian of the data this thesis is based on), Marko Simonovic, Giorgos Spathas, Roberta Tedeschi, Sharon Unsworth, Rosie van Veen, Anna Volkova, Marie-Elise van der Ziel, Arjen Zondervan and Rob Zwitserlood. Thank you all, for all the small and large favours you have done me.

Three of the people I have listed as ‘less-close’ colleagues above have actually been my closest colleagues in the literal sense: Rick, Ivana and Dagmar, you have been great roommates, making me feel really professional and/or inspiring me to behave as a professional every time I saw you working diligently and every time I saw your impressive collections of (actual) books. Rick, your idea of not drinking the terrible machine coffee but taking a real break and walk two minutes for a good coffee from Brandmeester’s did wonders for my productivity and probably also for my health. Dagmar, it has been great to share music with you. Ivana, it has already been a while since we were roommates, but I still remember how I enjoyed our quick and short discussions, whether about phonology or just about anything. As a new PhD student in a new field, it was great to have your help nearby.

I also wish to thank Arthur Samuel and the research groups of Paula Fikkert and Jacques Mehler and Marina Nespor for the opportunity to discuss my work in an informal setting. The same goes for the participants in the PPT meetings, especially Hannah for organising so much of it. I also thank Hannah and Frans for being my paranymphs.

Tikitu de Jager, the proofreader formerly known as Samson, and Marijn Koolen have been my academic friends, but also real-life friends, for a long time. In fact, my first experience with the development of theories on second language acquisition came from the attempts Marijn, Frans and me made to force Tikitu to speak Dutch (for the non-initiated reader, this was at his own request). I am afraid Tikitu has also learned the problems a Dutch native speaker has when writing English. If you, other reader, do not notice these
problems, Tikitu is to be thanked, as he proofread the whole book. I also wish to thank all my other friends, who have not seen me very often lately, for the fact that I had a nice social life before starting to write this thesis. Hopefully, I am still welcome to come back to this social life.

Without the support of my family, I might still have been able to write this thesis, but I would never have done it. My parents, Wies and Fred, have always supported and encouraged me in my academic career, both mentally and practically, while not putting any pressure on me, a combination that is a remarkable accomplishment in itself. I also thank my brothers Sven and Ruben and my late grandparents, Nettie, Kees, Hilda and Kas. The stories my grandmothers told me have made me realise that education is an opportunity and a privilege.

Belén, last but not least thank you a lot for making me realise that after all, the academic world is very silly and writing a thesis is the culmination of silliness. On a more practical note, I thank you and I apologise, because you have had to suffer the horrors of being the girlfriend of a thesis-writer, while I was able to rejoice in the pleasures of having a girlfriend who would feed me and make me coffee in the cases I forgot to do so, and force me sometimes to still have fun.
ABSTRACT

Phonotactic knowledge is knowledge of the sound combinations of a language. Previous research showed that frequent sound combinations are perceived more easily than ill-formed ones, while illegal combinations are often filtered out or adapted. Phonotactics can be described in categorical terms (i.e., a sound combination is legal or illegal), or in gradient terms (i.e., a combination has a certain level of wellformedness). In addition, phonotactics can both be described in negative or positive terms, by defining when combinations are legal vs. when they are illegal. This thesis presents experimental research on the mental representation of phonotactics that is used for speech perception.

A cross-modal priming experiment on second language listening showed that consonant clusters that are frequent in the second language are easier to process, in spite of perceptual epenthesis, an illusion caused by illegality of the same clusters in the first language. This disjunction between a categorical and a gradient effect suggest separate representations of both categorical and gradient phonotactics. However, the dichotomy between categoricalness and gradience is confounded with the dichotomy between negative and positive representations. The categorical filtering effects are based on negative knowledge, while the gradient facilitation effect is based on positive knowledge. Therefore, phonotactic effects on speech segmentation were investigated next.

In segmentation, the effect of negative categorical knowledge is facilitatory, instead of inhibitory. Illegal combinations cue word boundaries, because illegal combinations cannot occur word-internally. An eye-tracking method was developed to test segmentation effects independently of word activation. An experiment with Dutch native listeners showed that phonotactic knowledge has an effect on segmentation in the absence of full lexical recognition. A follow-up experiment showed that illegal segmentations are inhibited; there might additionally be a bias toward more wellformed segmentations. In the same experiment, second language listeners of Dutch were more likely to split consonant clusters that are illegal in Dutch, but legal in their native language. Nevertheless, this effect was more categorical for native listeners. When other, more wellformed segmentations fail to yield words, second language listeners make segmentations that leave illegal clusters intact, which native listeners hardly do. These results show the existence of categorical effects of illegality and gradient effects of wellformedness. All second language learners studied before showed transfer of categorical phonotactic illegality from their first language. To test if categorical knowledge is always fossilised, the segmentation of English and Dutch by native Dutch listeners, proficient in English,
was studied. The listeners applied English categorical phonotactics to English, showing that categorical phonotactics can be acquired. The empirical data presented in this thesis suggest that phonotactic representations can be intrinsically positive and negative, while their effects are categorical and gradient, depending on the task and the strength of the representations. A model of phonotactic grammar is proposed. This grammar is based on constraints representing positive or negative knowledge. It is applied to speech before lexical access takes place and can yield both gradient and categorical effects on speech perception.

*Keywords:* Phonotactics; Speech segmentation; Second language acquisition; Phonology; Gradient wellformedness; Markedness; Word recognition; Speech perception
INTRODUCTION

Phonotactics is a term used to describe the properties of the set of sound combinations of a language. In any language, some sound combinations are allowed, but hardly occur, some can only be used in certain situations, others abound and there are also combinations that are never used. This phenomenon has been studied from different angles, either as the distributions of sound combinations or as a system underlying the distributions.

Phonotactics has been shown in various ways to influence speech perception. As speech perception is performed by listeners, the phonotactics that influences perception has to be present and represented in the mind of the listener. Phonotactics is about sound combinations, but it is not a property of the sound combinations, because it is language-specific. Phonotactic influences on perception also depend on the listener’s language.

Listeners might benefit from knowledge of phonotactics in perception; the sounds that are recognised (as phonemes) should be grouped together into legal combinations. If that is not the case, it is possible that something went wrong. Knowledge about the legality of sound combinations thus has the potential to help correct errors and resolve ambiguities in the acoustic signal. In addition, knowledge about the illegality of sound combinations can also provide information that is valuable in speech perception, for instance to detect word boundaries and acoustic reductions.

As the effects of phonotactics on perception provide evidence of a mental representation of phonotactics, these effects can be employed to study the psychological adequacy of different theories of phonotactics. In other words, phonotactic effects on perception reveal the cognitive system underlying phonotactics.

Phonotactic knowledge could in principle correspond to the simple distributions of sound combinations of a language. However, recent research has shown that not all phonotactic knowledge can be directly related to statistics, as some unattested combinations are treated as more wellformed than others (e.g. Berent, Steriade, Lennertz, & Vaknin, 2007; Moreton, 2002). This indicates that listeners have a representation of phonotactics that is more complicated than a simple tabulation of frequency counts.

To address phonotactic effects that cannot be reduced to mere statistics, phonologists have proposed different theories of the phonotactic grammar, for
instance in the form of abstractions or learning biases and predispositions that shape the representation of the grammar. These proposals will be discussed in detail below, but it is important to notice that phonological theories of phonotactics generally assume the existence of a phonotactic grammar, that is either caused by the sound combination distributions (through learning), or alternatively, is the cause of such distributions. Importantly, the grammar is not the same as the distributions. Insofar as the grammar matches the observed phonotactic effects, it might be said to be licensed by it. However, not all properties of the implementations of phonotactic grammar have been empirically shown to be psychologically real, i.e. to be also properties of the representations in the minds of language users. Of course, these implementations are not wild, unfounded assumptions; considerations based on language typology, phonetic factors and theoretical parsimony usually underlie the assumptions made by scholars working on theories of phonotactics. However, following such considerations does not guarantee that the actual mental representations of phonotactic grammar are similar to the theory.

The psychological reality of proposed theories of phonotactic grammars should be tested empirically. There is in fact already a body of experimental research on phonotactics. Unfortunately, this research tends to eschew detailed theoretical frameworks of the mental representations of phonotactics, as it focuses primarily on the effects of phonotactics in speech processing, not on the representations of phonotactics. In many psycholinguistic studies, the phonotactic status of phoneme combinations is equated with their frequency of occurrence. This equation obscures the potential differences between the phonotactic knowledge of language users and the distribution of sound combinations in a language. The equation also implicitly reduces phonotactic knowledge to a positive value ranging from zero (unattested) to the highest possible probability (theoretically one, i.e. all sound combinations are one and the same; in practice, the highest probability is much lower).

The possibility to connect the observed effects of phonotactic knowledge to theoretical assumptions about phonotactic grammars needs to be investigated in a more integrated way. This thesis addresses some of the main issues in phonotactic theory and attempts to connect them to their effects on speech perception. These issues can be placed on two dimensions.

First, the question is how phonotactic knowledge is used in speech perception. It is possible that it guides perception independently, which would mean that perception is influenced by phonotactic cues even when these cues are not coherent with other cues, such as lexical knowledge. It is also possible that phonotactic information is not used to build percepts, but rather to check them, facilitating wellformed ones and/or inhibiting illegal percepts. This question will be addressed in the context of models of speech perception.
The second important issue is the form of the information deduced from phonotactic knowledge. This information is used in the process of speech perception, but it is unclear how: it might cause the rejection of illegal or improbable percepts, i.e. an inhibitory effect, but it might also be used to favour the most probable percept, i.e. a facilitatory effect. Neither option necessarily excludes the other.

If phonotactic knowledge is facilitative, hence employed to push perception towards the most probable percept, it is likely to be a gradient effect, as the phonotactically most likely percept is not necessarily the only acceptable option. Information from other linguistic sources, such as pragmatic, syntactic or lexical knowledge, might make another percept more felicitous. On the other hand, if phonotactics is inhibitory, it might very well be categorical in nature, as illegal sound combinations cannot be in the final percept; no syntactic, pragmatic or lexical cue can make it possible to perceive acoustic input felicitously if the percept contains a phonotactically illegal word, as no cues are expected to point towards such words, for the simple reason that they by definition do not exist. However, this also means that lexical information is possibly enough to restrict perception to legal sound combinations: as long as only words are recognised, the outcome of perception is phonotactically legal. Hence, the fact that perception is biased towards (only) phonotactically legal sound combinations does not have to be attributed to independent knowledge of phonotactics.

As phonotactics are language-specific, there is one notable exception to the idea that words are phonotactically legal. Words in one language are not necessarily phonotactically legal in another language. Therefore, if it is indeed true that phonotactic knowledge of a language is beneficial for perception of that language because it guides perception towards words, it is likely that the phonotactics of one language can be detrimental to the perception of another language. Second language perception can thus offer valuable insights into the effects of phonotactic knowledge on perception without an alternative lexical explanation. In addition, native and second language listeners can be tested on the same tasks. The differences observed on these tasks can then be assumed to mainly reflect the differences in phonotactic knowledge.

Proposals for inhibition or facilitation caused by phonotactic knowledge, as well as for phonotactic gradience or categoricalness, have to be addressed in the context of a theoretical framework of phonotactics in speech perception.

There are two starting points to for such a theoretical framework. First, the place of phonotactics in models of speech perception has to be considered. The value of such models is that they account for many empirical findings on speech processing. Second, theories of phonotactics have to be considered;
these account for phonotactic phenomena unrelated to word recognition and are therefore more elaborate regarding the structure of phonotactic knowledge.

This introduction will first address the architecture of speech perception. Although speech perception is not limited to word recognition, the normal purpose of speech perception is the mapping of speech sounds onto words representations; hence, speech perception will addressed by discussing models of word recognition, in §1.1 (p. 4).

The possible form of phonotactic knowledge will be discussed in §1.2 (p. 21). As this thesis focuses on a perceptual process, these linguistic theories have to be assessed on their psychological reality. The scholars that formulated the different theories of phonotactics were not always attempting to capture the psychological aspects of phonotactics, i.e. knowledge of phonotactics; many scholars have attempted to capture the linguistic aspects of phonotactics. There is no fault in studying phonotactics as a property of language, but for the purposes of this thesis, a cognitive view on phonotactics has to be taken. Linguist approaches thus have to be translated to more psycholinguistic statements that yield predictions about perception.

A cognitive view can be imposed by considering the relation between phonotactic theory and the empirical effects of phonotactic effects on the behaviour of language users. §1.3 (p. 35) inventories the literature on phonotactic effects on speech perception, especially with respect to the theoretical concepts identified in §1.2.

1.1 MODELS OF WORD RECOGNITION

Not all word recognition models include a full-fledged theory of the representation of phonotactic knowledge, but some models do incorporate the use of information that can be derived from phonotactic knowledge. However, there are also models in which phonotactics is explicitly not assumed to be represented separately; these models explain phonotactic effects as an emergent effect of properties of the lexicon (these models are usually interactive models, most notably the TRACE model).

Unfortunately, word recognition models do not agree on the way phonotactic knowledge is used in speech perception. In some models phonotactics directly shapes perception, while in others it merely modulates perception in inhibiting or facilitating percepts that are the result of the application of other knowledge, e.g. candidate words in the lexicon that have been activated. The origin of these different views is not empirical evidence for the place of phonotactics in particular, but considerations of the structure of the perception process in general. These considerations will first be addressed in a general discussion of the architectures of word recognition models.
When perceiving speech, the mind maps a highly variable acoustic signal onto a string of meaningful units, usually words (or morphemes). Most models assume that the acoustic signal is mapped onto low level units, usually phonemes. As lexical entries are also assumed to be composed of phonemes, phoneme recognition is an important step towards words. Lexical representations do not only contain phonemes, but also of the order of the phonemes in the word, which means that the temporal nature of speech is also important. The Cohort model (Marslen-Wilson, 1987) addresses the relation between phonemes, their order and word recognition. In this model, a cohort containing all the words that start with the first recognised phonemes (at the start of the utterance) are activated. As more phonemes are recognised, fewer words match the signal. The words that do not match are deleted from the cohort, until only one word matches the phoneme string. At that point, the end of the word is known and the process can start again at the phoneme after the last phoneme of the word.

The Cohort model successfully captures the temporal ‘left-to-right’ nature of word recognition, but assumes that all phonemes are correctly recognised. The slightest noise or careless pronunciation would make it impossible for the model to successfully terminate with a correct percept. This fails to capture the fact that human listeners can recover from mispronunciations in the speech they are hearing, with the use of lexical knowledge. For instance, a near-word, e.g., ‘christmash’, will be recognised (as ‘christmas’), but a mispronunciation that results in another word will not be corrected, e.g. ‘mash’ will not be recognised as ‘mass’ as both are correct pronunciations of existing words. The employment of lexical knowledge was captured in interactive models such as TRACE (McClelland & Elman, 1986). In such models, a misperceived phoneme or feature can be corrected by a lexical entry, if that lexical entry is backed up (activated) by enough other low level or top-down information.

Another problem for the Cohort model is that it cannot deal with words that end before the cohort converges on them. A word such as ‘can’, perceived in the context ‘can deal’, will not be recognised as the recognition can continue to a cohort containing ‘candy’, which also matches the input. At the point that the word ‘can’ ends, the cohort is not reduced to one entry, while the cohort does not correspond to the correct perception when the cohort does converge on one entry (‘candy’). TRACE suffers from the same problem.

The problem of overlap and embedding has been addressed by adding competition between words to the models. This is a hallmark of the Shortlist A and B models (Norris, 1994; Norris & McQueen, 2008). These models differ from the cohort model and the TRACE model because they also model speech segmentation and its effect on word recognition. The Shortlist models also differ from TRACE in another aspect: they do not allow the activation of units
at a low level to be changed due to higher level information; nevertheless, the model does allow phonemes to activate lexical items when the match is imperfect.

Imperfect matches between two levels should not be discarded, as the speech signal is prone to noisy transfer and imprecise production. Hence, a lot of information has to be used to resolve ambiguities, while errors and noise have to be filtered out. The information used in this process comes mainly from knowledge of the lower level units, as the mapping from acoustic input to phonemes or features makes it possible to match the signal to words. Lexical knowledge is also important, as it reduces the number of possible percepts to the lexical entries.

Phonotactic knowledge also has the potential to reduce the number of possible percepts, but on a lower level than lexical knowledge. Phonotactic knowledge has the potential to correct illegal sound combinations, just as lexical knowledge can correct non-words. A sound combination is phonotactically illegal if it cannot occur within a syllable or word. However, lexical knowledge has the same potential, as illegal combinations can never be part of a word and can thus also be corrected by mapping a non-word containing an illegal combination to the nearest word. Nevertheless, illegal sound combinations can actually occur in continuous speech, but only if they straddle word boundaries. Illegal combinations should therefore not always be corrected and can be used to identify possible word boundaries. Still, as the Shortlist models propose, lexical competition can also be used to identify word boundaries.

Phonotactic knowledge is not an essential part of word recognition models, as many phonotactic effects are also possibly explained as the effect of lexical knowledge. An important question is thus whether phonotactics should be included in word recognition models. The way in which it is included can only be answered after the role of phonotactic knowledge has been identified, but both questions can only be formulated after a discussion of the architecture of word recognition models.

1.1.1 The architecture of word recognition models

Bottom-up processing

At least two types of mental representations are normally identified in models of speech perception. On the lowest level, acoustic input is mapped onto a limited number of sound categories. These are traditionally phonemes, although features and syllables have also been proposed. Phonemes are the temporally smallest units that can distinguish meaning. Acoustically different realisations of the same phoneme do not convey different meanings, while two sounds that differ only slightly, but are still placed across a phoneme
boundary, can make the difference between two different words (Trubetzkoy, 1939; Liberman, Harris, Hoffman, & Griffith, 1957; J. J. Jaeger, 1980, but cf. Savin & Bever, 1970). If the listener ignores the acoustic differences between different versions of the same phoneme, no meaningful information is lost. Knowledge of phonemes thus helps word recognition. In a similar way, other basic units, such as features and syllables, can help to recognise spoken words.

Phoneme recognition is not a straightforward reduction of the information in the speech input; it is a task that the listener can only perform with knowledge of phonological processes such as allophony. The same can again be said for perception of other basic units, such as syllables or features, although syllables might be less variable and features more. Syllables are nevertheless also hard to recognise, as their boundaries in time are not always marked in the speech signal. A non-word of two syllables, e.g. ‘ansta’, can be recognised as two syllables in many different ways, e.g. ‘an’ + ‘sta’, ‘ans’ + ‘ta’ or even ‘anst’ + ‘a’ in the case of Dutch; see e.g. Quené (1989, 1993) for an overview of cues used in the perception of boundaries. For the remainder of this thesis, the speech input is assumed to be mapped to phonemes, although alternatively theories exist and are acknowledged. However, it is too confusing to phrase every theory in terms of phonemes, features and syllables.

When the low-level units are recognised, they can be linked to words. Words are represented as strings of phonemes and strings of phonemes can thus be similar or identical to words. This kind of bottom-up processing is the general architecture of many word recognition models. The TRACE model, proposed by McClelland & Elman (1986), links features – activated by acoustic input – to phonemes and phonemes to words. On each of these levels, units that are consistent with the lower level (or the acoustic input) are activated. Units that are inconsistent with each other, because they are placed on the same level and activated simultaneously, inhibit each other. This ensures that a stretch of speech is not recognised as two different units.

The TRACE model thus assumes that a match between features and words, through phonemes, can be established. If a perfect match is not possible, the information in the lexicon helps to reach recognition of the nearest word, i.e. a word that is not strictly present in the input. Activation from the higher level can also activate nodes at lower levels; this so-called lexical feedback can help correct errors in the feature or phoneme recognition.

Lexical knowledge is essential for speech perception, as speech perception is word recognition. However, the contribution of lexical knowledge is more profound. Knowledge of word forms can provide information that fills gaps in the information that was extracted at the phonemic level. If a listener hears a word in which a phoneme is replaced by white noise, i.e. ‘legi#latures’, the missing phoneme is not needed to activate the form ‘legislatures’; the word
with this form is activated. Listeners do not notice the absence of the phoneme (Warren, 1970; Samuel, 1987; Frauenfelder, Segui, & Dijkstra, 1990).

A similar effect occurs for phonetic categorisation of sounds that are ambiguous between two phonemes; perception of phonemes that make an ambiguous string a word is preferred, meaning that the same ambiguous sound is interpreted as different phonemes depending on lexical knowledge. This is called the Ganong effect (Ganong, 1980).

Note that this top-down use of lexical knowledge to solve problems at lower levels is not the lexical feedback that Norris, McQueen, & Cutler (2000) vehemently argued against. They were referring to lexical feedback that changes the actual perception of phonemes and the discussion focused on the definition of 'perception'; the usefulness of lexical knowledge for speech recognition is not under debate. With the Shortlist models, Shortlist (later renamed Shortlist A) and Shortlist B (Norris, 1994; Norris & McQueen, 2008), the same authors propose a model in which phoneme perceptions are indeed not influenced by lexical knowledge. Nevertheless, lexical knowledge is still important in these models, as the final perception largely depends on it. The phoneme string is, as in TRACE but with a different mechanism, mapped to the best fitting string of words; if a perfect fit is not possible, mismatches between recognised phonemes and perceived words have to be tolerated. The Shortlist models were developed to deal with the speech segmentation problem, i.e. the mapping of the continuous speech input to a string of discrete lexical items; the TRACE models cannot elegantly handle words that are not isolated. Although continuous speech contains some cues to word boundaries, these are not enough to determine where all word boundaries are; this makes it necessary to use different sources of information to accomplish segmentation (see Mattys, White, & Melhorn, 2005, for an overview).

The Shortlist B model differs from the Shortlist A model in that it does not consider phonemes and words to be units that receive activation and pass activation on. Rather, it considers phonemes and words to be events, each associated with a probability. Word recognition is modelled as a Bayesian process in which the probability of possible perceptual outcomes is calculated. The most likely events are assumed to be the actual outcome of the process. Although the shift from units and activation to events and probability is made on well-argued grounds, they do not matter for the present discussion. In the remainder of this introduction, Shortlist B will be discussed, rather than Shortlist A.

In both Shortlist models, the first level of recognition is a string of phonemes. In Shortlist B, a time slice of acoustic input is not categorised as one phoneme; the probabilities of all phonemes given the acoustic evidence are calculated. These probabilities are, for practical reasons, derived from phoneme confusion.
rates, but represent the certainty with which the input could be categorised. The phonemes are used to find words that match the input at consecutive points in time (for practical reasons, at every new phoneme). Bayesian probability derivations are used to reach the most probable parse of the input, which can be assumed to be the most likely perception. The a priori probabilities of words are derived from their frequency. After the identification of the set of candidate words, a process starts that finds segmentation paths through the speech stream. Such paths are given a probability based on the probability of the words they consist of, but also based on the felicitousness of the segmentation. A segmentation hypothesis in which a sequence of words occur is assigned the probability of the words. Hence, if a stretch of speech is not assigned to a word, the path is less probable, although the possibility of an unknown word is left open.

Thus, the Shortlist B model uses knowledge of phonemes, namely how likely they are given acoustic input. It also uses lexical knowledge in the form of the set of words. As all phonemes are in principle possible, the most likely phoneme given the acoustic input does not have to correspond to the phonemes in the most likely word. The Shortlist B model, just like TRACE and even Cohort, thus employs lexical information to resolve problems in the recognition of phonemes and aid speech segmentation.

In sum, both TRACE and Shortlist employ bottom-up processes informed by phonemic and lexical knowledge. The discussion on lexical feedback is not relevant for the questions discussed in this introduction, but note that many authors disagree with the rejection of lexical feedback. See e.g. Luce, Goldinger, & Vitevitch (2000), who propose the Adaptive Resonance Theory (ART) architecture that uses feedback and resonance as a feature, not a burden, and claim that the inclusion of feedback is actually parsimonious, because feedback and resonance are seen in many other perceptual domains. The ART architecture will be discussed below. It suffices to note that problems at the phonemic level can sometimes be resolved by lexical information.

**Exemplar models**

There is an alternative to the view of word recognition of the compositional bottom-up models. It is the exemplar-theoretic view of perception. Every episodic encounter of a word is kept in memory as a trace, according to exemplar-based models. A trace is a pattern of activation of all the features recognised in the input, including non-linguistic features such as the speaker’s vocal characteristics. New input is compared to all memory traces and the input is perceived as the same as the closest trace. Abstraction only occurs during retrieval, as the input is assigned to the same category as the closest
exemplar. The representation of words in an exemplar-based model is not abstract; it is not a composition of phonemic representations (Goldinger, 1998).

Exemplar models have been applied to perception in general, not just to speech perception. An important aspect of these models is that the episodes that are subsumed by a category are not generalised or normalised. However, it turns out that the recognition of a stimulus as belonging to a newly learnt category depends on its closeness to the centre of the category, the prototype (Smith, 2002; Homa, Hout, Milliken, & Milliken, 2011). These authors presented participants in an experiment with tokens of a category of dot patterns. None of these tokens was in the centre of the category, the centre being the mean position of the tokens. New tokens were then judged on their similarity to the tokens; distance to the centre, not to previous tokens, predicted the perceived similarity. The authors conclude that this proves that the representation is an abstraction over the example tokens.

Exemplar theory can thus not account for every aspect of perception. It turns out that it also cannot explain every aspect of speech perception. Goldinger (1998, 2003) argues that although there are episodic effects on (repeated) word recognition, there are also effects of normalisation, i.e. of representations containing abstractions. The same is argued by Pierrehumbert (2001a, p. 139), who states that “…the correct model must describe the interaction of word-specific phonetic detail with more general principles of phonological structure.”

Ernestus (to appear) also advocates a hybrid model with both episodic and abstract word representations, to account for both the frequency effects for acoustically reduced pronunciations on the one hand and the privileged status of the unreduced form in recognition out of context on the other hand. The first phenomenon cannot be adequately explained by purely abstractionist models, as these would recover the unreduced (canonical) form after normalisation. If this is the case, reduced forms cannot have properties, such as frequency effects, that differ from the canonical form. The privileged status of canonical forms, on the other hand, cannot be accounted for by exemplar-based recognition alone, as a frequent reduction should be easy to activate as compared to an infrequent canonical form. There is thus still a normalisation process that makes the recognition of both reduced and unreduced forms as the same word possible.

**ART models**

Both Goldinger (1998, 2003) and Ernestus (to appear) refer to dynamic models, for instance the Adaptive Resonance Theory (ART) framework, to solve the puzzle posed by exemplars and abstract representations. ART was formulated by Grossberg and colleagues (Grossberg & Stone, 1986; Grossberg, Board-
Both abstract representations and episodic effects are incorporated in ART models. For speech recognition, this model has been formulated as the ARTWORD model (Grossberg & Myers, 2000; Grossberg, 2003), which is a synthesis of earlier ART models of speech perception.

The ARTWORD model contains phoneme units, that are present in the short-term memory if activated by the input by so-called iconic sensory features provided by peripheral neuronal activation, typically coming from the auditory nerve. The phonemes provide bottom-up input to higher-level list chunks, that represent concatenations of lower level units; list chunks compete with each other and provide top-down feedback to the units they consist of, similar to the interactive activation in the TRACE model.

If a chunk can provide enough feedback to the lower-level units that support it, the activation resonates temporarily. The resonant state is tantamount to recognition of the word. The representation of the word is thus an attractor state, an activation pattern into which the network is likely to resonate if it comes close enough, i.e. given enough input in the right direction. The ARTWORD model features abstract representations, but the activation of these representations is not completely independent of previous experience or knowledge at higher levels. Importantly, ART models are dynamic and their activation patterns develop over time. This allows for the modelling of the influence of temporal sequencing and durational information from the input.

1.1.2 Phonotactic cues in word recognition

The use of phonotactic cues in word recognition

The architecture of almost all speech recognition models, except the pure exemplar-based models, is based on the existence of representational units on different levels. There are representations for features (and/or phonemes) and for words. These units are thus entities in the model and their existence explains their influence on word recognition, together with the connections in which they are involved. The terminology is slightly different for Shortlist B, as the words and phonemes are modelled as events in that model.

If the word recognition process relies on a bottom-up, compositional mechanism that yields correct and complete mappings, word recognition models can be defined in the most straightforward way without assuming the use of phonotactic knowledge. The lexical look-up of a correctly recognised string of phonemes is simple if the string is indeed a word; normally, one would expect this to be the case as the input is produced by a speaker who intends to convey a message expressed in words. However, perception is quite
stable even when disturbed by noise. The word recognition process has to be assumed to be more complicated than sheer phoneme-to-word mapping.

The complications start at the lowest level: phonemes and/or features lack reliable invariant acoustic properties (but cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and syllable recognition suffers from the problem of identifying the correct boundaries. The string of recognised phonemes is thus not guaranteed to be in accordance with the speaker’s intention. As described above, lexical knowledge can help to correct imperfections. Nevertheless, the knowledge contained in the lexicon is not guaranteed to solve all problems that percolate from lower levels. Even if the recognised phoneme string was perfectly in accordance with the speakers intention, the speaker might have used words unknown to the listener. As listeners can actually perceive words they have not encountered before, the actual recognition system should be assumed to have a way of dealing with unknown words.

In addition to the possible insufficiency of lexical and phonemic knowledge to decode speech input, it is possible that it is too hard to identify all lexical forms that are possibly present in the speech stream, if the search for such forms has to initiated at every phoneme. The calculations needed for the Shortlist B model include a lot of very improbable events. To keep the time needed to run a simulation of a Shortlist B recognition process within reasonable boundaries, the set of events for which probabilities are calculated is pruned; events that are very unlikely are dropped from the process. This is especially relevant for segmentation paths, as these have to be identified before the set of words is known. This process is left unspecified, in the sense that the psychological reality of this part of the process is not addressed.

It is an empirical question whether the computations of Shortlist B, that might require exhaustive lexical searches at every phoneme, are too complicated for the human mind (for instance because they take too long or because they lead to a number of activated units that surpasses the mind’s capacity). Unfortunately, it is not possible to satisfactorily answer this question yet, as the current knowledge of the neural substrate of word activation is too limited to make the counterarguments sufficiently precise and falsifiable.

In sum, it is possible that a mental process that uses only low level recognition and a bottom-up compositional process and top-level knowledge cannot converge on word recognition. If this is indeed the case, word recognition cannot be such a process, as word recognition does normally converge and yield an output. Nevertheless, this argument is too speculative, as it might be the case that clever solutions to the problems sketched above exist in the actual mental process of word recognition.

There is a more convincing argument for the idea that lexical and phonemic knowledge are not enough to decode speech. If the phonemic and lexical
knowledge do not suffice to decode speech, listeners should use other types of knowledge or fail to decode speech in many circumstances. As speech perception is assumed to be possible for even degraded or erroneous speech, there is indeed evidence that additional knowledge is employed.

Phonotactic knowledge is among these types of additional sources of recognition cues. It affects the recognition process: Massaro & Cohen (1983) found that sequences of two sounds that were both ambiguous between two phonemes were more likely to be recognised as a legal phoneme sequence, even if an illegal combination was also possible. This effect is very similar to phoneme restorations based on lexical knowledge, as found by Warren (1970) and Ganong (1980) (discussed above).

Phonotactic knowledge also helps to identify boundaries between words, as illegal sound combinations can in principle only occur across word boundaries. An illegal sequence is thus likely to contain a word boundary. McQueen (1998) found that listeners are quicker and more accurate when spotting words with a boundary within an illegal sound combination.

In addition, phonotactic legality is used to assign probabilities to segmentation paths, as segmentations of speech are less likely to be made if they contain parts that cannot be legal words. The word ‘apple’ in ‘fapple’ is harder to recognise than in ‘vuffapple’, as ‘f’ is not a possible word, while ‘vuff’ is, even if neither ‘vuff’ nor ‘f’ are actual words (Norris, McQueen, Cutler, & Butterfield, 1997; Norris & McQueen, 2008, see also McQueen, 1998).

Incorporating phonotactic cues in word recognition models

Phonotactic knowledge has been incorporated in both the Shortlist and the ART models, but in different ways. Vitevitch & Luce (1999) propose to add representations of sound combinations as units; this would mean that phonotactic knowledge is represented in a similar way to phonemes and words. For the Shortlist models, incorporation of phonotactic knowledge usually takes the form of adaptation of the activation of other units (in Shortlist B, the probability of other events).

TRACE does not readily model phonotactic effects on speech recognition. One of the reasons for this is that it was not designed for speech segmentation. The phonotactic effects it would have to explain are the effects on non-word sequences of phonemes, such as the ambiguous biphones used by Massaro & Cohen (1983). McClelland & Elman account for the influence of phonotactic (il)legality on the classification of acoustically ambiguous phonemes by assuming that frequent transitions are reinforced through lexical feedback. When many lexical items contain a certain phoneme sequence, these lexical items are activated by the sequence, which in turn activate the phonemes in the
sequence through lexical feedback. This feedback is less strong for phoneme sequences that are infrequent in the lexicon and absent for illegal sequences.

As TRACE does not really address the problem of the complexity of converging on the right segmentation without evaluating huge numbers of hypotheses, the remainder of this subsection will be dedicated to the place of phonotactic knowledge in Shortlist B and ARTWORD.

The designers of Shortlist B did consider that phonotactic knowledge could be represented in the form of representations of sound combinations, as they mention that “[a] more complete model would (...) include modulation of the computation of \( P(\text{PhonemeString}) \) as a function of transition probabilities” (Norris & McQueen, 2008, p. 363). In other words, the phoneme string ‘event’ has a probability that does not only depend on the probability of the phonemes given the acoustic evidence, but it can also have an independent (a priori) probability. This is similar to the a priori probability of words, that is based on their frequency.

The same approach, be it in a very different framework, is taken by Vitevitch & Luce (1999), namely within the ART framework. The ART model incorporates frequency effects on other units than words and phonemes. Frequency effects for lexical items where already proposed by Grossberg & Stone (1986), but according to Vitevitch & Luce (1999), the same effects are present at the sublexical level, because frequent phoneme combinations are represented as chunks (in the experiments, these frequent combinations are biphones). These chunks aid the recognition of frequently combined phonemes, because they offer more connections between the acoustic input and the representations of words.

Words that contain the frequent phoneme combinations have a representation that is connected to the representation of the biphone, as well as the representations of the phonemes. When the acoustic input activates the biphones and phonemes, the word thus receives extra activation and can reach a resonant state more quickly. This architecture can explain a number of findings on word and non-word processing reported by the same authors and colleagues (Vitevitch & Luce, 1998, 1999, 2005). The assumption that some biphones are represented in a similar way as phonemes is actually not recent; Trubetzkoy (1939) already suggested that complex sounds, consisting of what in some languages are two phonemes, are monophonemic in other languages (see also Hintze, 1950; Morciniec, 1961). However, these researchers avoided considering a biphone as monophonemic if the components of the biphones can also occur separately. This makes \([sf]\) monophonemic in Spanish, which has no \([j]\), but not in many other languages, including English. The proposal by Vitevitch & Luce (1999) is different, as they consider the represen-
tation to depend on the frequency of the biphone, irrespective of the separate occurrence of its parts.

In ART, input activates items in working memory, that can resonate with (list) chunks in short term memory. These chunks can thus be phonemes and highly-frequent phoneme combinations, but also syllables and words. The chunks can resonate with words. Resonance equals activation, or recognition in the case of a word. By virtue of the mathematical structure of the model, resonance implies mutual reinforcement, but of a transient nature. A temporary equilibrium occurs, in which the best matching item is active and resonates with its component lists chunks. The resonance is due to bottom-up processing alone, but the presence of higher level representations is essential for the possibility of resonance. As the sublexical chunks in short term memory include both phonemes and phoneme combinations, the phoneme combinations allow for a second route into the word, thus facilitating its recognition. This gives phonotactic structures a representation quite similar to phonemes. Figure 1.1 shows the gist of this idea.

The ART-based model offers a view of phonotactics in which high probability is special. The role of highly frequent sublexical units’ role in word recognition is as follows. To reach a resonant state corresponding to the perception of a word, there have to be connections between the input and the word’s representation; the wellformed units’ representation provide extra connectivity. For example, if /kæt/ ‘cat’ is assumed to be composed of such highly frequent biphones and /fɪʃ/ ‘fish’ not, the first word will reach a resonant state much faster, given the right input. This is schematically shown in Figure 1.1. Words that do not contain wellformed structures, such as ‘fish’, mostly resonate when their phonemes are supported by the input, but if a word contains one or more of the separately represented units, there are more routes for it to resonate with the input.¹

The Vitevitch & Luce (1999) proposal thus also incorporates sublexical information by assuming separate knowledge of phoneme combinations, just like the suggestion of incorporating transitional probabilities in Shortlist B. The difference is that the information in the lexicon contains connections between the lexical items and the sublexical representations of biphones. In fact, the authors propose no strict formal properties of the sublexical units involved; any chunk that can be encountered in processing can be represented and used for the processing of speech (Grossberg et al., 1997, p. 482). In ART, the sublexical units thus influence words, while in Shortlist, they are proposed to influence phonemes.

¹ Note that in the original proposal by Vitevitch & Luce (1999), the units for the biphones of the word ‘fish’ are shown and assumed to exist, but with a very low connectivity to the word and the input. These units are suppressed for ease of explanation.
Figure 1.1: Sublexical representations of highly-frequent phoneme combinations explaining facilitatory phonotactic effects. Simplified version, based on Figure 2 from Vitevitch & Luce (1999, p. 378). Direct links from input to words are not shown and the infrequent bigraphemes /fi/ and /sj/ are not shown, as they are assumed not to be represented.
It should briefly be mentioned that phonotactic knowledge is not the only type of knowledge that can modulate the speech recognition process. Cues for word recognition can also come from supralexical levels, for instance from the shared nature of discourse. The possible topics of every new sentence are limited and interlocutors tend to align the use of lexical items and the manner of referencing (Pickering & Garrod, 2004). This supralexical information can be treated in the same way as word frequency, which is a lexical effect. The main difference is that frequency is relatively static and can thus be stored with the lexical entry, while discourse is dynamic.

This is acknowledged by the designers of Shortlist B. They posit that the a priori probability of a word is for practical reasons derived from the word frequency: “...we assume that P(Word,) can be approximated by the word’s frequency of occurrence in the language. However, P(Word,) will also be influenced by factors outside the scope of the present model, such as semantic or syntactic context ” (Norris & McQueen, 2008, p. 362).

For dynamic models, e.g. ARTWORD, the same can been proposed (cf. Grossberg & Stone, 1986; Grossberg et al., 1997; Grossberg, 2003); a semantic stratum of words might receive some activation from the topic of conversation, which accounts for easier resonance for words in accordance with the topic.

Other cues, mainly sublexical, are derived from phonetic or phonological knowledge. One important sublexical cue is derived from knowledge of assimilation processes. When a phoneme is adapted to a phoneme preceding or following it, it is usually no longer contrastive in the position it is in, while it is contrastive in other contexts. E.g., ‘garden’ can be pronounced as ‘gardem’ before ‘bench’, but not in all contexts. Knowledge of this phonological process helps the listener to notice that the ‘m’ in ‘gardem bench’ can also be an ‘n’ (Mitterer & Blomert, 2003), as place of articulation can be assimilated regressively.

Speech segmentation cues in word recognition models

Shortlist A does not allow the phonotactic information to calculate possible word boundaries; it only uses phonotactic information to inhibit words that cannot be recognised without segmenting the speech stream in an illegal way, at least in McQueen’s discussion of his results (McQueen, 1998). This discussion predates the formulation of Shortlist B. However, it invokes the Possible Word Constraint, which has been incorporated in Shortlist B. A stretch of speech that could not be a possible word and thus violates the Possible Word Constraint is assigned a very low probability, virtually guaranteeing that segmentations violating this constraint will not be the outcome of the perception process. The phonotactic knowledge expressed as a constraint is thus an external source of cues, but not represented as an event itself. Rather,
phonotactic evaluation is evidence, in this case against a ‘segmentation event’ (a segmentation path containing an impossible word).

Apart from phonotactic cues, there are other sublexical cues that are employed in speech segmentation (Mattys et al., 2005). Such cues include subphonemic cues (Quené, 1989; Salverda, Dahan, & McQueen, 2003). In the Salverda et al. study, segment lengthening made listeners favour a perception of the monosyllabic embedded word, against the longer enveloping word. The authors suppose that listeners interpret the lengthening as the effect of an upcoming prosodic boundary, thus allowing the lengthening to function as information about boundary positions. In Shortlist B, such effects have to be assumed to modulate the probability of segmentation paths.

In the ARTWORD model, the problem of embedded words is solved differently. The model moves from word to word by building up resonance states of a temporal nature, that bind different units together. The architecture is in principle biased towards longer words, as it allows the largest chunk that can resonate to dominate over shorter ones (‘candy’ wins over ‘can’), as more input resonates with larger chunks, making this resonance stronger. However, this bias is only effectuated if there is no competing stronger resonance with the surrounding context. The end of a large chunk (e.g. ‘dy’ in ‘candy’) can belong to the next word and resonate with it. If it does (e.g. if the context is ‘can deal’), this next word (‘deal’) will inhibit the longer chunk (‘candy’). This will favour the shorter word in its competition for the stretch of speech (the acoustic input ‘can’ will resonate with the representation of the word ‘can’, as this word can resonate, having beaten its larger brother ‘candy’ for the stretch of input with the help of ‘deal’).

Subphonemic cues in the form of durational differences have an effect in the ART model as a long pause (‘can . . . dy’) leads to the culmination of the activation of the shorter chunk before the pause, as it has full support from the phonemes in the input and is not inhibited by the larger chunk. Activation of the larger chunk cannot reach the highest possible level, because the support from the first part of the word has already decayed before the next part of the word is processed. Therefore, the smaller word wins the competition with the larger word when the pauses are long enough. Along similar lines, it is likely to be possible to incorporate the findings of Salverda et al. (2003) mentioned above in the model. A number of effects of duration and silencing are already incorporated, as mentioned by Grossberg and colleagues (Grossberg et al., 1997; Grossberg & Myers, 2000; Grossberg, 2003).

Another sublexical cue that guides word recognition is the metrical pattern of a language. In English, strong syllables are more likely to be at word onsets than at offsets. This information can thus be used to find the onsets of words and Norris, McQueen, & Cutler (1995) indeed found that listeners
use this information, as they prefer to maximise strong syllables. E.g., the word /stæmp/ ‘stamp’ was recognised slower and less accurately in the string [stæmpdʒ] than in [stæmpədʒ], as the second syllable is strong in the first string, thus making it more likely that the [p] is grouped with the following context than with the word ‘stamp’.

Such cues are incorporated in the Shortlist A model by increasing the activation of segmentation candidates (or the words in these segmentations) that follow the metrical regularity. This solution can easily be transferred to Shortlist B.

Metrical cues have not been explicitly modelled in ARTWORD, but their effects might be analysed as the result of the prominence of stressed syllables. These are in general longer in duration and thus gather more activation for resonance. Few lexical items will resonate with weak syllables that contain a reduced vowel, thus making it less likely for this syllable to enter in a resonant state. This allows the boundary consonant to group with the preceding word more easily. However, simulations and modelling are necessary to see if the ART model behaves similar to human listeners in this respect.

In sum, Shortlist B models most sublexical cues, including phonotactic cues in word segmentation, as information that is added from an external source and that dynamically updates the a priori probability of events (words or segmentation paths). In ART, many sublexical cues are incorporated in the architecture of the model, either because the dynamics of the model are sensitive to subphonemic cues, or, as in the case of sublexical information, because the sublexical knowledge is explicitly represented. The latter is also a possibility in Shortlist B, but it is in contrast to the way in which additional cues are normally incorporated.

The placement of phonotactic information in word recognition (or speech perception in general) can now be made more specific. The first option for representations of phonotactic knowledge is that sound combinations are actually represented, in a form similar to words or phonemes, including frequency or probability, as advocated explicitly by Vitevitch & Luce (1999) and discussed favourably by Norris & McQueen (2008). This can account for biases towards wellformed sound combinations. Wellformed would then be defined as ‘well-known’; the more a sound combination occurs, the higher its probability; this knowledge is then used as a cue for word recognition.

The second option is that phonotactic cues come from an external source (which can be best called ‘phonotactic grammar’, to avoid confusion). This phonotactic grammar parses sound combinations and modulates the activation of other representations to the extent that they are compatible with the parses assigned by the grammar. This option can account more easily for the constraining phonotactic influence on speech segmentation, such as the
rejection of impossible words and the detection of boundaries in illegal sound combinations. Capturing phonotactic cues as being generated by a phonotactic grammar locates phonotactic knowledge outside of word recognition proper. Shortlist B includes the employment of external cues, but it is not specified how phonotactic cues are generated.

ARTWORD, on the other hand, does not really incorporate information coming from a phonotactic grammar. The idea is not rejected by its designers, but this is due to the fact that it is not discussed and that the segmentation problem is not treated as extensively as for the Shortlist models.

The only phonotactic knowledge that can be incorporated into ART is positive probability information, for the sound combinations that a listener knows. It is still to be seen if the constraining effects of phonotactics can be incorporated into an ART model and explained as the result of only positive knowledge. Nevertheless, it is conceivable that this is possible. Along similar lines as the use of metrical information, an impossible word, e.g. the /f/ in the non-word ‘fapple’ can resonate with the ‘a’ and the ‘p’ and thus impede the recognition of ‘apple’; the syllable ‘vuf’ in ‘vufapple’ will be more likely to be grouped together, leading to decaying activation of ‘vuf’ (hence the ‘f’) before the word ‘apple’ is processed. This assumption does not necessarily follow from ARTWORD, even after adding the sublexical units proposed by Vitevitch & Luce (1999). A more explicit account of the way sublexical units become involved in processing is needed, because sublexical units such as ‘fa’ and ‘uf’ then have to be assumed to be always more frequent than the impossible words such as ‘f’. This is trivially true exactly for those cases in which the ‘word’ is impossible, but this only means that there is no a priori reason to assume that ART models such as ARTWORD cannot incorporate phonotactic cues for speech segmentation. However, the negative influences of phonotactic illegality have to be translated to positive influences of alternatives in order to model the effects in an ART model.

To resume, the models of word recognition do not unequivocally answer all questions about the representation of phonotactic knowledge, but they offer two options: phonotactic knowledge comes from outside and modulates units (or events) in the model, or sound combinations have separate representations that directly influence the recognition process. When phonotactic knowledge is modelled with representations of sound combinations, facilitatory effects on speech recognition can be modelled. The view that phonotactic knowledge is something more or less outside the word recognition process, that provides modulating input, can address the effects of phonotactics on speech segmentation. It involves the assumption that a phonotactic grammar parses the phonemic input. This grammar contains constraints on sound sequences and on possible words; if a possible perception threatens to violate the constraints,
the perception is modulated negatively, i.e. inhibited or attenuated. As the concept of a phonotactic grammar is still vague, the next section will address the possible architectures for a phonotactic grammar.

1.2 PHONOTACTICS IN LINGUISTIC THEORY

This section inventories theories of the representation of phonotactic knowledge, in order to understand the possible formalisations of phonotactic knowledge. In this thesis, the formalisations should account for the cues that phonotactic knowledge provides for speech perception. However, linguistic theories of phonotactics are not necessarily designed for this purpose. In linguistics, the term phonotactics is used to refer to the study of the restrictions and/or allowances on the combinations of sounds in a language. The latter entails a shift from the language user to language proper.

The structure of words in a language is only partially defined by its phoneme inventory; this inventory only provides the elements to form words in a language. Phonotactics in its linguistic conception provides an account of the way these elements can be combined. For instance, in Dutch the sequence /mr/ is never used (within a syllable), but /rm/, as in ‘arm’, is. In addition, some combinations are used frequently, while others only occur marginally. There are two ways to describe this. One is by assuming the language is constrained by a grammar, e.g. it has a phonotactic grammar that does not allow or generate /mr/ sequences. The other way is by tabulating the attestedness and frequencies of sound combinations. The table could simply include a ‘legal’ mark for every entry that is attested anywhere in the language, or an ‘illegal’ mark for all unattested combinations. Alternatively, the table could indicate the probability of each combination. However, most linguists consider such a tabulation of distributions not as the phonotactics of a language, but as the explanandum of phonotactic theory.

The explanation of the distributions is then given by a system, the grammar, that assigns wellformedness to sound combinations, or generates sound combinations. The grammar does not have to explain all the entries in the tabulation, as some might be accidental. In addition, phonotactic knowledge, as stored in the mind of a listener, need not necessarily include the whole phonotactic tabulation of a language. E.g., Moreton (2002) showed that American English listeners have a bias against /dl/, but not /bw/, although neither sequence is attested.

Contemporary phonologists often implicitly try to capture the cognitive aspects of phonotactics, when they try to formalise it in a grammar. The implicit assumption is that the mental representation of phonotactics is more abstract and systematic than the tabulation.
Still, the conception of phonotactics as a theoretical concept in linguistics can differ from the conception in psycholinguistic approaches. Psycholinguistic research often only addresses the probability of a sound combination, i.e. the wellformedness value from the tabulation (usually the frequency), as the effects of phonotactic wellformedness are studied, not the underlying explanation of the wellformedness.

Nevertheless, there is a potential role for a phonotactic grammar in speech perception, next to, or even in place of, sound combination frequencies, as it can address empirical evidence that experimentally observed phonotactic knowledge diverges from the surface distributions of sound combinations.

1.2.1 Phonotactic grammaticality

Categorical legality

Phonotactics has traditionally been regarded as grammatical knowledge that defines which sound combinations a language uses to construct words (Chomsky & Halle, 1965, 1968). Regardless of the architecture of the system responsible for allowing or banning sound combinations, phonotactics is in this view a function that assigns a binary value to every sound combination, legal or illegal. Chomsky & Halle (1965) connect the linguistic concept of phonotactic legality to cognition, as they claim it can be observed in the intuitions of language users. Native speakers of English know that /brk/ is a word of English, /blk/ is not, but could be, while */bnk/ could not be. The differences between attested and unattested (/brk/ vs. /blk/) do not suffice to identify phonotactic legality, as unattested words can be in the legal stratum, while other unattested words (e.g. */bnk/) are illegal. The most important conclusion made by Chomsky & Halle is that there is a phonotactic grammar that makes this difference, not the entries in the lexicon.

Given this conclusion, phonotactics is studied as a grammar in the phonological – linguistic tradition, i.e. as a system that distinguishes legal and illegal structures. The most important property of grammars is that they contain abstractions. Chomsky & Halle (1965) followed Jakobson & Halle (1956/1971) by using an abstract format for phonological theories; they assume his format to be innate. The format comprises a view of phonemes as composed of distinctive features and a view of morphemes as having a lexical representation containing a phonological structure.

Chomsky & Halle (1965, 1968) propose this abstraction because it provides the necessary format to state generalisations about regularities and restrictions in the lexicon. Among these statements are the so-called lexical redundancy rules, that specify how some features do not have to be specified in the lexicon because they can only take a particular value due to the context. E.g., the
Dutch /η/ before a /k/ is actually not specified for the features [−coronal, 
−anterior, +back]; these features (the velar place of articulation) are always 
predictable from the following consonant, the /k/.

In this approach, sound combinations are made to build word forms that 
differ from each other. Some combinations are employed, while others are 
simply not in the inventory of the grammar. Many of the structural character-
istics of words turn out to be related to the syllable. Hence, researchers such 
as Fudge (1969) and Selkirk (1982) define the legality of sound combinations 
with reference to their place in the syllable. They argue that a syllable has 
to contain a core, usually a vowel, preferably an onset of consonants and 
optionally (in some languages) a coda of consonants. Sequences of sounds 
can have different legality depending on their position in the syllable. E.g., 
‘rm’ is legal in a Dutch coda, but not a Dutch onset. The underlying reasons 
for legality of sound combinations is derived from abstract rules, for instance 
the sonority hierarchy, that assigns a different sonority to different kinds of 
phonemes; in many languages, onsets have to rise in sonority and codas have 
to have falling sonority.

With a phoneme inventory and abstract phonotactic rules, the syllable 
inventory is defined, together with its side-effect, the inventory of sound 
combinations. As the inventory of legal syllables must be at least as large as 
the inventory of different syllables in the lexicon, the difference in phonotactic 
legality between the unattested syllables ‘bnick’ and ‘blick’ can be explained.

Gradient wellformedness

In the 1960s, concurrently with the work of Chomsky & Halle (1965), a 
more psychologically-driven approach to the description of phonotactics took 
the gradient nature of listeners’ wellformedness judgements to be part of 
the phonotactic knowledge of these listeners, as well as categorical legal-
ity/illegality. Greenberg & Jenkins (1964) and Scholes (1966), argued that 
there is more to phonotactics than binary wellformedness judgements, i.e. 
judgements that take the values ‘legal’ and ‘illegal’ only, as native speaker intu-
itions reveal. As Greenberg & Jenkins, p. 579 put it, “we are likely to be dealing 
with a dimension more complex than a three-point scale consisting of existing 
sequences, generatable sequences, and non-generatable sequences”. The three 
researchers just mentioned have all used rating scales and magnitude estima-
tion tasks and found a consistent correlation between linguistically derived 
wellformedness and participants’ gradient wellformedness judgements.

The ‘linguistically derived’ notion of wellformedness of Greenberg & 
Jenkins (1964) is based on the number of substitutions needed to arrive 
at an actual word; thus, wellformedness is linked quite directly to lexical 
frequency, because when a non-word contains many phoneme combinations
that are frequent in the lexicon, it is likely that only a few substitutions are needed to get to another word. Scholes (1966) advocates a formalisation of possible theories of phonotactic grammaticality, which assigns a level of grammaticality to any string of phonemes. This grammaticality can either be derived from native speaker judgements and explain linguistic generalisations, or alternatively, linguistic generalisations can explain speaker judgements.

### 1.2.2 Legality and gradient wellformedness in theories of phonotactic grammar

As described above, gradient wellformedness has been analysed as being the direct effect of frequencies of occurrence of sound combinations, which entails that there is no grammatical system behind these effects. Traditional phonotactic grammars define whether a sound combination is legal or not, which does not directly yield any gradience.

There are two approaches to the relation between gradient wellformedness and phonotactic grammar. One is to describe gradient wellformedness as a difference in markedness, as assigned by a phonotactic grammar. Markedness is a theoretical notion, expressing that a linguistic structure is, in opposition to another structure, likely to be avoided. It depends on the grammar of a language to what extent markedness is acceptable; this depends the type of structure as well. This entails that the correlation between the probability of a sound combination and its wellformedness is driven by wellformedness: the more wellformed a sound combination it is, the more likely it is to occur.

The other approach is to assume gradient wellformedness is derived from the frequency of occurrence. Anttila (2008) refers to this as the ‘lexical explanation’ of gradient wellformedness effects: the more a combination occurs in the lexicon, the more wellformed it is. If one follows this reasoning, lexical frequencies and probabilistic phonotactics are isomorphic.

Probability differences are mostly gradient, as probability has a continuous scale. If gradience is exclusively caused by probability differences, it is accidental that probability differences follow phonotactic regularities, as the probability itself is the cause of any phonotactic wellformedness. This means that the regularities in phonotactic wellformedness and probabilities are not explained and presumably need not be. Following this logic, there is no reason why categorical legality is not accidental as well; phoneme combinations could just accidentally not occur, i.e. have a zero probability. This thwarts the effort of explaining any phonotactic wellformedness, as legality of sound combinations is not random, but shows many general tendencies that are to be reviewed below. If this explanatory goal is accepted as an argument for a categorical grammar, it should also be accepted as an argument for a grammar that explains the regularities in gradient wellformedness. This is especially
relevant as there are many indications that the same abstract properties of sound combinations can sometimes show categorical effects and sometimes gradient effects.

For instance, lexical distributions of English syllabic nuclei are more tightly connected to the coda than to the onset (Kessler & Treiman, 1997). The vowel of a syllable predicts the coda much more than it predicts the onset, suggesting that sound combinations are more restricted within the rhyme than between nucleus and onset. In many languages, such restrictions are categorical, but in English, the number of categorical restrictions between nucleus and coda and between nucleus and onset happen to be the same. The tighter connection between coda and nucleus that often shows in categorical restrictions shows in gradient restrictions in English.

There is more evidence for the relation between gradient and categorical effects. They often invoke the same linguistic factors, such as phonetically grounded phenomena. This can be revealed both within one language and across languages. Frisch, Pierrehumbert, & Broe (2004) argue that phonotactics causes certain properties of the lexicon that are gradient in some languages but categorical in others. They refer to the pattern of similarity avoidance described as the Obligatory Contour Principle (OCP), that was first proposed as a categorical phenomenon by McCarthy (1986). The OCP shapes the lexicon, according to Frisch et al. (2004) because “lexical items that avoid repetition will be easier to process, and so will be favoured in acquisition, lexical borrowing, coining novel forms, and in active usage.” In this way, “once a similarity avoidance pattern becomes established, it will be further reinforced by the grammars of the speakers that learn the pattern, since grammar influences borrowing and novel word formation as well.” Frisch et al. prefer to model the OCP as a gradient effect, depending on a metric of similarity to be avoided, as this leads to a better fit with the data they wish to explain, namely the Arabic lexicon.

The proposal of Frisch et al. entails that phonotactic probabilities are at least partially explained by linguistic factors. The proposed functional motivations for the OCP are possibly phonetic; similarity avoidance, as in the case of the OCP, might be caused by tiring of the articulators or insensitivity of the auditory system to repeated sounds, a factor that in turn eventually, in the course of language change, shapes the lexicon. Nevertheless, as not all languages obey the OCP to the same extent (some not at all), the OCP is language specific knowledge. Its application in a language is to a certain extent arbitrary and hence in the domain of phonology and phonological knowledge. Frisch et al. (2004, p. 182) propose “(...) that the realization of similarity avoidance constraints in the worlds languages falls on a continuum of strength from the gradient to the categorical” and go on to explain that the
OCP is categorical in some languages and gradient in others. The observation that sound combinations that are categorically illegal in some languages tend to be marginally well-formed in other languages is also made by Berent et al. (2007), as will be discussed in §1.3.3 (p. 40).

An even more compelling argument against severing probabilities from phonotactics is that a sound combination might be avoided in a language depending on the phonological context, but in a completely predictable way. Dutch does not generally allow voiced obstruents in codas, but regressive voicing assimilation can optionally change a coda obstruent into a voiced one (Wetzels & Mascaro, 2001; Zonneveld, 2007). It is thus possible that structures that are avoided in most cases, e.g. voiced obstruents in codas, are only allowed if there is a good reason, namely that the alternative, e.g. conflicting voicing values within a consonant cluster, is even worse in the language concerned. This makes it possible to fully explain the data with a grammatical explanation, but the same is not possible with a probabilistic explanation.

In comparison to Dutch, note that German does feature categorical final devoicing across phonetic contexts, since it has no voice assimilation. This is another example of the cross-linguistic observations that show that a phenomenon can be gradient in one language and categorical in another. Although the probability of voiced obstruents in Dutch codas is not actually zero, as in German, it is not necessary to reject applying the same explanation based on categorical final devoicing for Dutch and German.

Devoicing is a phonological principle that is grounded in articulatory phonetics and observed in many other languages, including German; in addition, final devoicing is observed categorically in those contexts that are unaffected by contact with other phonemes. A theory incorporating final devoicing and voice assimilation can thus perfectly explain the legality and thus probability of obstruent devoicing in both Dutch and German codas, while the phenomenon is only strictly categorical in German. The idea that linguistic regularities are observed in interaction with other regularities is the hallmark of constraint-based theories, in which constraints can be violated if needed to avoid the violation of more important constraints. Constraint-based theories can be expressed within a number of different frameworks, the most well-known being Optimality Theory (OT; Prince & Smolensky, 2004 (1993)). These will be discussed in §1.2.2 (p. 27).

In sum, it is theoretically unattractive and moreover not necessary to assume that avoidance of a combination is caused by phonotactics when it is categorical, but not when it is almost categorical. Avoidance of a sound combination is not only witnessed by its complete absence, but can also show in a low frequency of occurrence. Attributing these factors to a phonotactic
grammar can avoid the unwanted consequence that language typologies are accidental and that many regular interactions have no explanations.

For these reasons, many phonologists assume that gradient wellformedness is not directly caused by frequency of occurrence, but that it should be explained by a phonotactic grammar as well. There are two ways of achieving this goal. One is to assume that a phonotactic grammar causes both categorical and gradient wellformedness and that gradient wellformedness causes differences in probability distributions. The alternative is to assume that probability distributions shape the grammar. These two possibilities will be discussed below.

Gradient and categorical wellformedness in a phonotactic grammar

The lexical redundancy rules proposed by Chomsky & Halle (1965, 1968) might informally be seen as a type of constraints on the structure of words, but they were proposed when phonology was conceived formally by means of a set of rewrite rules operating on, amongst others, sound sequences. These rules were driven by the input. However, phonologists had increasing difficulties describing with rules the phenomena they discovered on the output. They often had to resort to multiple rules on the input to avoid certain structures in the output.

Kisseberth (1970) proposed that a rule has to transparently describe its output function, if it has one. Rules should not conspire to achieve one output goal. In general, he advocated the use of statements that call for avoidance of certain structures, because these structures are marked. This idea was one of the principal motivations behind the formulation of Optimality Theory (OT) by Prince & Smolensky (1993/2002). In OT, constraints were made independent and rules were abandoned. Constraints come in two types: markedness and faithfulness.

Markedness is a concept that was already present in earlier work on language (e.g. Jakobson & Halle, 1956/1971). In phonology, a sound or sound combination can be considered as more marked than another for different reasons, for instance, it can be harder to pronounce or perceive. Markedness shows in the avoidance of the marked structure in favour of an unmarked one. This avoidance can show in patterns of acquisition (when unmarked structures are acquired earlier) or in typology (languages that use a marked structure also use the unmarked counterpart; the unmarked structure is never avoided if the marked structure is present). A markedness constraint expresses the force away from marked forms to unmarked forms. It thus departs from ordinary illegality in that it only indicates that a structure might not be ‘good’ enough (in other words, might be too marked) to be legal.
Faithfulness, on the other hand, describes the preservation of (lexical) contrast by requiring that surface and lexical representations must match in some respect. Markedness constraints in phonology are independent of the lexical representations. It would not even be problematic if the lexicon contained marked illegal sound combinations (nevertheless, these tend to disappear due to a process known as lexicon optimisation; see Prince & Smolensky, 1993/2002). However, marked forms in the lexicon do violate the markedness constraints. In principle, markedness constraints require the removal of all marked structures. This pressure is countered by faithfulness constraints. If there were no faithfulness constraints, lexical contrast would be impossible to achieve, as words would be reduced to the (same) most unmarked structure.

A form can thus be marked, but still be legal due to a faithfulness constraint. This means that markedness and faithfulness constraints are in conflict. The OT framework addresses this conflict. The constraints act on a mapping, normally between lexical representations and their realisation. An example taken from Kager (1999) is the realisation of the word /bed/. Coincidentally, this word means ‘bed’ in Dutch as well as English. It is supposed to be represented with the voiced obstruent /d/ in coda position in the lexicon. However, voiced obstruents in codas are marked; languages that have them, also have voiceless codas, while there are also languages that have no voiced codas. Dutch is such a language, featuring final devoicing. The pronunciation of /bed/ in Dutch turns out to be [bet]. This means the markedness constraint is more important than the faithfulness constraint. The Dutch words /bed/ ‘bed’ and /bet/ ‘dab’ cannot be distinguished, as they are both pronounced as [bet]. Only the contrast between the inflected forms [bedə] bedden ‘beds’ and [betə] betten ‘to dab’ shows that the underlying form of ‘bed’ in Dutch contains a /d/. The markedness constrains against voiced codas does not apply to the plural forms, as the /d/ is now intervocalic.

The differences between Dutch and English show that Dutch prefers to obey a markedness constraint, while English prefers to obey a faithfulness constraint. This allows the contrast in the pronunciation between the English words /bed/ ‘bed’ and /bet/ ‘bet’.

The resolution of the conflict between constraints is captured in OT by ranking the constraints. The highest ranked constraint is obeyed first. Formally, the process takes the following form. A word, represented in the lexicon, is the input to a generator that is unlimited; it can in principle generate all logically possible output forms. These output forms are all candidates to become the actual output. The output forms are subjected to the constraints; they might violate markedness constraints, but also faithfulness constraints if they do not correspond to the input. The candidate set is first evaluated against
the highest-ranked constraint. All candidates that violate this constraint are
eliminated. The process is repeated for the next constraint, until one candidate
is left. This is the optimal candidate. The optimal candidate is thus the least
marked candidate that the grammar allows, given the required faithfulness.

In the example of the pronunciation of the word /bed/, the constraints
are traditionally labelled NoVoicedObsCod, ‘obstruents in coda position are
voiceless’, and Ident-IO(voice), ‘segments have the same voicing in input and
output’. Simplifying the candidate set to [bet] and [bed], the evaluation for
English proceeds as follows. First, the faithfulness constraint Ident-IO(voice)
is evaluated. Candidate [bet] violates it, while [bed] does not. Hence, [bed] is
the only remaining candidate and thus legal. The lower-ranked markedness
constraint NoVoicedObsCoda does not have to be evaluated, as only one
candidate is left.

For Dutch, NoVoicedObsCoda is ranked above Ident-IO(voice) and
therefore evaluated first, eliminating [bed] and leaving [bet]. This example is
simplified, as other candidates are possible. Nevertheless, these are assumed
to be even less faithful or even more marked, due to constraints that are not
included in the example.

In OT, legality is thus the same as optimality: a candidate has beaten all
its competitors, all other surface realisations of the same lexical material. As
optimality depends on both input and output, legality cannot be inferred just
from markedness constraints. Returning to phonotactics, this means that a
grammar assigning phonotactic constraints contains not only the knowledge
of markedness in the form of constraints, but also knowledge about the im-
portance of markedness, in the form of the ranking of markedness constraints
against the faithfulness constraints with which they conflict.

However, OT can also be employed to describe gradient wellformedness.
One option to generate gradient wellformedness with an OT grammar is to
take the order of elimination of candidates into account. The later a candidate
is eliminated, the closer it was to being optimal. This is called the harmony
of the candidate, and was already taken into account by Prince & Smolensky
(1993/2002). Every candidate is harmonic relative to another candidate, given
the input; if a candidate A violates a constraint that a candidate B does not
violate and candidate B does not violate any higher ranked constraints that
candidate A does not violate, candidate A is less harmonic. It will therefore
never be optimal and hence is illegal. Nevertheless, candidate B does not have
to be optimal. Hence, illegal candidates do have different harmony values
given the same input. This does not hold for legal candidates, as two legal
candidates normally do not have the same input and if they do, they have the
same harmony value.
Coetzee (2009) proposed that an OT grammar can still be used to assign different wellformedness scores to legal forms. He proposes that wellformedness be derived from the harmonicity of the output, as defined by the markedness constraints it violates. His proposal can (slightly simplified) be explained as follows: to compare two legal candidates, they are evaluated without considering faithfulness, as they do not necessarily have the same input. The comparison employs only the violations of markedness constraints. The highest markedness constraint that is violated by only one of the candidates decides which of the candidates is more wellformed. Note that the same principle can also be used to evaluate illegal forms.

Another option to capture gradience in a phonotactic grammar based on constraints is given by Boersma (1997) and Boersma & Hayes (2001). They argue that probabilistic knowledge, wellformedness and phonotactic legality are to be explained by one grammar, that is learnt from probabilities and can assign both categorical legality and wellformedness. Boersma (1997) proposes Stochastic OT. In this framework, grammars are a set of constraints, as in Optimality Theory. However, the ranking is enhanced with a stochastic value. This allows the grammar to be different in every instance it is consulted, as the place of every constraint in the ranking varies around their mean ranking value. Unlike in standard OT, the same input can sometimes lead to a different output. If a markedness constraint is ranked high, the marked sound combination is usually avoided. However, if the markedness constraint is ranked relatively close to a lower-ranked faithfulness constraint, the marked combination can sometimes still occur. The cause of this is that the ranking position is based on a value that is drawn from a probability distribution. In some cases, the higher-ranked markedness constraint is given a lower ranking than normal and the lower-ranked faithfulness constraint a higher one. This opens the possibility that they flip ranking. Boersma & Hayes (2001) propose that the degree of wellformedness be derived from the probability of occurrence. Categoricalness is actually a limit case, that technically only occurs when a constraint is ranked infinitely far away from a conflicting constraint. Practically, however, near-categoricalness is easily generated with the Stochastic OT grammars proposed by Boersma & Hayes.

The OT variants introduced above can produce gradient and categorical effects with a grammar consisting of categorical constraints, but constraints can also be gradient. An example is the OCP constraint as proposed by Frisch et al. (2004). This constraint can be violated to different degrees, depending on the distance (measured in features) between the homorganic consonants. The larger the violation, the stronger the force to avoid the sound combination. Classic OT is not endowed with a mechanism to cope with gradient violations of constraints, but the framework of Harmonic Grammar
& Smolensky, 1990) is. In this framework, the harmony of an output candidate is based on a sum of the ranking values of the constraints it violates. The highest ranked constraints thus contribute more to the harmony (note that the contribution is negative); hence not violating these constraints results in a better score. The most harmonic candidate wins. The difference with OT is that lower ranked constraints can team up against a higher ranked one.

Wellformedness and legality in grammars derived from probabilities

The nature and direction of the relation between phonotactic wellformedness and gradience might also be reversed. High probability might cause high wellformedness. Assuming that frequency of occurrence has been stored in some form, wellformedness might simply reflect the familiarity of the listener with certain types of sound combinations. In fact, many effects of phonotactics on speech perception, as found in the psycholinguistic literature, have been discussed from the viewpoint that phonotactic wellformedness equals probability of occurrence. To give just one example, Vitevitch, Luce, Charles-Luce, & Kemmerer (1997) state that “(…) phonotactics accounts for the probability that a given phonetic segment will be followed or preceded by another particular segment. In addition, phonotactics refers to the probability that a given segment will occur in a specific position within a syllable or word.” Interestingly, later psycholinguistic work, even from one of the same authors (Luce), refined the definitions, mentioning that “[w]ithin linguistics, phonotactics typically refers to a system of rules or constraints that govern the legality of the occurrence of segments and sequences of segments within the syllables and words of a given language” and that “[w]ithin the category of phonotactically legal configurations, segments and their sequences occur in the linguistic environment with varying frequencies” (Auer & Luce, 2006).

The latter definition seems to suggest that phonotactics is a categorical system that defines the sound combination inventory, while there are also, separately, differences in frequency of occurrence. In this subsection, this option will be explored. Importantly, as Auer & Luce (2006) already discussed, the existence of probabilistic knowledge does not imply that there is no grammatical knowledge in the form of constraints.

In fact, some kind of abstract knowledge has to be imposed on the representation of probabilistic knowledge in any case for it to explain phonotactic wellformedness. For instance, the wellformedness of syllables cannot be derived from their occurrence, as witnessed by the ‘blick’–‘bnick’ difference; neither occurs, but the first is legal and the second is not. At least rudimentary grammar is needed to interpret ‘bl’ as attested in its position (syllable onset), while ‘bn’ is not attested in the same position. In other words, phonotactic wellformedness of pseudowords cannot be calculated from attestedness if
the components of the pseudoword are not identified. As ‘ibl’ is not a legal
syllable, the attestedness of ‘bl’ cannot just be assumed to be enough to make
‘blick’ legal. The rudimentary grammar thus has to include knowledge about
syllable structure and the components of which the attestedness has to be
assessed. The grammar is therefore, in fact, not extremely rudimentary.

Coleman & Pierrehumbert (1997) propose the addition of knowledge of
syllable structure, to arrive at a mechanism to derive a probabilistic parse of a
word or non-word (see also Pierrehumbert, 2001b, 2003). Each combination is
assigned a probability and probabilities of the constituents of the word are
combined. The product of the composition of the constituents, the word, has
a probability that is the product of the composition of the probability of the
constituents. In spite of the addition of this elaborate but not complicated
compositional mechanism, these proposals are still highly probabilistic.

In addition to the issue of assigning probability to pseudowords based
on their components, anyone proposing a grammar of phonotactics based
on frequencies also has to pick how abstract the combinations are that it
learns wellformedness over and what the granularity of these abstractions is.
Pierrehumbert (2001b, 2003) argues that only certain patterns can and will
actually be learnt from the input, as the input is not limitless in size. The
general problem is that a phenomenon of a low but not zero frequency might
not occur in a small sample. However, if the phenomenon is very improbable,
the sample size needed for it to occur more than once might be larger than the
available sample size. Very complicated patterns might thus accidentally have
a zero frequency in the input of one language user, but not for another user.
Pierrehumbert proposes that wellformedness knowledge is not exactly the
same as the simple knowledge of all probabilities: probability is then merely
the input to a derivation process and phonotactic knowledge is the result of
the derivation.

Another proposal for a model of probabilistic knowledge of phonotactics
is suggested by Hayes & Wilson (2008). These authors also propose that con-
straints are induced from positive evidence (lexical frequencies), but shaped by
knowledge of features. This knowledge allows to generalise the probabilities
of attested structures to the probability of similar unattested structures. In
the article by Hayes & Wilson (2008), the generalisations are captured by
constraints that are assigned weights based on a maximum entropy principle;
these weights are then used to combine the sum of constraints violated by
a possible word; this sum is subsequently used to calculate the probabil-
ity of the possible word. The constraint set thus reflects the probability of
sound combinations in the lexicon and the resulting grammar, although it is
constraint-based, is not constraint-based in the typical OT sense, as it assigns
probability, not legality.
Albright (2009) takes a roughly similar approach, in which attested sound combinations are also generalised to other combinations that differ on a few features. Albright proposes that the wellformedness of unattested biphones depends on the frequency of a more general class, that is not specified for all its features.

In sum, to describe knowledge of phonotactic probabilities, some assumptions have to be made regarding the structure of the grammar. Nevertheless, such a grammar does not represent phonotactic knowledge as a set of constraints on the language, but rather, as an inventory of the occurrence of sound combinations, together with a composition mechanism. It could be enhanced with a measure of the distance of each illegal sound combination to legal combinations, for instance based on the difference in phonological features. Nevertheless, a probabilistic grammar does not assign markedness, but probability, to sound combinations. To derive categorical legality from probability, the simple assumption that any probability above zero entails some form of legality is enough.

1.2.3 Phonotactic grammars: markedness and probability

It has been shown above that a constraint-based grammar can model the representation of illegality and yield both categorical and gradient wellformedness. This wellformedness is actually turned upside down; the less markedness constraints are violated, the more wellformed a combination is. When faithfulness is ignored, all sound combinations can be compared on their wellformedness. Still, to determine legality with constraints, the interaction with faithfulness has to be considered, as this is where certain forms of markedness can turn out to be tolerated to preserve faithfulness, while other forms are completely illegal.

A probabilistic grammar, on the other hand, assigns wellformedness in a positive way. The more probable a certain sound combination is, the more wellformed it is. Probability can be assigned on the abstract class the sound combination belongs to, but this does not matter for the outcome.

In other words, an account of phonotactic knowledge has to include both the gradient and categorical wellformedness of sound combinations. This can be achieved by a constraint-based account that assigns markedness. This markedness can be interpreted as negative wellformedness, but it can also be used to arrive at categorical legality when the input to the process is known. Knowledge of the probability of sound combinations can also be used to arrive at wellformedness, when probability is assumed to be positive wellformedness. In addition, this knowledge of probabilities needs to be
more abstract than simple tabulations and needs to contain abstractions and knowledge of grammatical structure.

To study the representation of phonotactic knowledge in speech perception, there is another important fact to take into account. In OT grammars, sound combinations are judged both on their markedness and on their correspondence to the input. Nevertheless, in speech perception, the input is not formed by lexical forms. The mapping runs from an input of sound combinations to an output of recognised words, possibly via an intermediate level at which phonetic cues are categorised (Boersma, 1998, 1999). Therefore, candidates containing marked structures should be completely omitted from consideration, but only if there is no faithfulness that can save them; this is only true for illegal combinations. However, as phonotactic knowledge alone cannot perform perception on its own, it should not yield just one candidate. In this case, the optimal candidate is not the only legal one. Still, faithfulness constraints are relevant; the input, a sequence of sounds, should not be altered too much, as it probably closely matches the correct perception. Rather, the most likely alterations should be indicated; these include word boundaries in marked sequences or changes of marked combinations to less marked ones, to correct for possible misperceptions.

The role of a phonotactic grammar in perception would thus be to detect markedness in the input and order the set of possible percepts based on their markedness. The most marked perceptual candidates should be the least likely, with as possible consequence that a really marked structure (that violates a high-ranked markedness constraint) is changed or split by a word boundary, rather than kept faithfully intact. The action of the phonotactic grammar thus takes place at marked structures. Unmarked structures will not be changed, as changing them would only violate faithfulness constraints.

It is not directly clear how constraints can provide a bias towards more probable or wellformed combinations. If two combinations are both unmarked, they should both be preserved.

Conversely, a probabilistic phonotactic grammar derived from probabilities, as described above, can explain the bias of perception towards the most likely sound combinations. These combinations are more probable; as already mentioned in the discussion of word recognition models, the more probable a perception candidate is, the more likely it is to become the outcome of perception. Shortlist B takes this idea very literally; phonotactic probability can be incorporated as high a priori probability. ARTWORD can be adapted to contain representations of wellformed combinations that aid words containing these combinations to reach the resonant state corresponding to perception.

However, the problem for probabilistic knowledge of phonotactics is that it does not directly address biases against marked structures. It might be
possible to explain the bias to perceive legal sound combinations over illegal ones by a bias for legal combinations, as legal combinations by definition have higher probabilities. Still, it is unclear how a probabilistic grammar could provide cues for word boundaries or how it can correct illegal combinations, but not do the same for legal but improbable combinations. If highly probable combinations call for attention, they might not only impede the recognition of illegal combinations, but also improbable combinations, as long as the distance in probability is high enough. In other words, a probabilistic grammar cannot capture the categorical difference between illegal and just legal. If it predicts perceptual illusions for illegal sound combinations, because the illusion is more probable, it also predicts illusions for marginally legal sound combinations when there are very wellformed alternatives.

As constraint-based grammars and probability-based ones capture different aspects of phonotactic wellformedness, there are four logical options. Either neither of them accurately captures the effects of phonotactics on word recognition, or one of them, or both of them. The principle of parsimony makes it necessary to provide empirical evidence to justify added complexity. The next section of this introduction will inventory the existing empirical evidence and discuss its relation to the two types of grammar.

The most parsimonious option of the four given above is the strawman theory that neither constraint-based grammar, nor probabilistic grammar are licensed by empirical evidence. Nevertheless, this option is only a mock-up; it will be clear that there is abundant evidence for both. Only an even better theory can make the two conceptions of phonotactic grammar both unnecessary. However, it is very well possible that only one conception is needed. As mentioned above, both probabilistic knowledge and constraint-based grammars can yield gradient effects; the first gradient wellformedness and the other gradient markedness. Potentially, both types of gradience can be reduced to one type; high markedness is low probability, or high probability is low markedness. It is possible that apparent markedness knowledge is actually caused by probabilistic knowledge, or vice versa. Therefore, the possibility of unification has to be studied as well. Still, if one of the types of grammar cannot explain an observed effect, it does not have to be rejected immediately. Only if all effects can be attributed to only one type of grammar must the other kind be rejected immediately.

1.3 PsycHOLINGUISTIC evidence for phonotactIC knowledge

From the above discussion, it appears there is a choice for a representation based on grammatical constraints or one based on probabilistic phonotactic wellformedness. Clearly and preferably, the choice should be made based on
empirical evidence. One way of obtaining such evidence is the conduction of speech perception experiments that do not involve word recognition. This section discusses the interpretation of experiments (both on word recognition and speech perception) that are taken as evidence for either constraints or probabilistic knowledge.

It is important to distinguish between phonotactic grammars and phonotactic knowledge. As argued above, it is most parsimonious to explain both gradient and categorical phonotactics with one grammar, and it is neither practically nor theoretically impossible. However, it remains to be seen if language users’ knowledge corresponds to one type of grammar, i.e. if their knowledge corresponds to any of the linguistic descriptions. This has to be empirically tested.

To perform empirical tests, one could use the intuition that a purely probabilistic grammar only yields effects relative to differences in probability, but not purely binary distinctions between illegal and legal. Illegality only emerges as low or zero probability. As discussed above, voiced obstruents in coda position, for instance, are illegal in German and their probability is thus virtually zero. In Dutch, they are also illegal, but due to voice assimilation, they might still have a non-zero probability. For Dutch, probabilistic knowledge would thus never match the categoricalness of a constraint against voiced obstruents in codas. Still, an empirical test to investigate if the knowledge of Dutch final devoicing is categorical is virtually impossible; experimental data is always noisy, which could lead to false counterexamples. Even worse, categoricalness would not even be proven if no counterexample was found, as this could be due to the size of the sample.

On the other hand, there is evidence that probabilities affect speech perception, but also evidence that has traditionally been ascribed to illegality. This evidence will be discussed in depth below, but briefly, a distinction between probabilistic phonotactic effects and categorical effects can be drawn, as follows. Effects of phonotactic probability effects have been proposed to emerge in wellformedness judgements (Pertz & Bever, 1975; Ohala & Ohala, 1986; Coleman & Pierrehumbert, 1997; Frisch, Large, & Pisoni, 2000; Bailey & Hahn, 2001; Pierrehumbert, 2001b, 2003), but also in word recognition, as mentioned above (Vitevitch et al., 1997; Vitevitch & Luce, 1998, 1999, 2005; Luce & Large, 2001; Auer & Luce, 2006), where the more frequent a sound combination is, the easier it is found to process. This also applies to phoneme detection (McQueen & Pitt, 1996).

Categorical knowledge has been proposed to explain effects of perceptual illusions in phoneme detection in illegal clusters (Brown & Hildum, 1956; Massaro & Cohen, 1983; Hallé, Segui, Frauenfelder, & Meunier, 1998; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi, & Mehler,
which occur frequently in second language listening (Weber & Cutler, 2006; Davidson, 2011), and also result in incorrect second language speech production (Davidson, 2006).

It is important to note that the researchers referred to above do not always explicitly choose a categorical explanation over a probabilistic one, or vice versa. Moreover, there is also literature on phonotactic effects that cannot be directly classified as due to probabilities or categorical knowledge alone. This mainly applies to phonotactic effects on speech segmentation (McQueen, 1998; Gaygen & Luce, 2002; Mattys et al., 2005; Weber & Cutler, 2006; Adriaans & Kager, 2010).

1.3.1 Wellformedness judgements and wordlikeness

The existence of phonotactic knowledge derived from probabilities has been proposed as an explanation for gradient wellformedness judgement. Typically, participants in wellformedness experiments are asked to judge a pseudoword. If participants are asked to give an estimate of the wellformedness on a gradient scale, or when binary judgements are averaged, the wellformedness usually has a consistent gradience in it, as shown by Greenberg & Jenkins (1964) and Scholes (1966).

The gradience can often be related to the probability of the sound combinations in the pseudoword. However, some authors argue that a better explanation is found by looking at how much a pseudoword resembles an existing word. This is the approach taken by Greenberg & Jenkins (1964). Ohala & Ohala (1986) compared lexical neighbours and sequence probability and found a better match for the first. On the other hand, Coleman & Pierre-humbert (1997) argue that their probabilistic parser provides a better fit with wellformedness judgements than an account in which the probability is not derived by parsing, but by summing the violations incurred by illegal sound combinations in the pseudo-word. They suggest that their contribution provides a better explanation of gradient wellformedness than lexical neighbours and also than an OT grammar based on constraints, as OT grammars only add markedness violations.

Unfortunately, there is a problem with the direct association of wellformedness judgements with phonotactic wellformedness. The argument that phonotactics is derived from lexical probabilities is circular if the judged wellformedness of a pseudoword is equated with phonotactic wellformedness. This becomes apparent if the procedure of the experiments is put under scrutiny. The participants in the study of Greenberg & Jenkins (1964) were presented with non-words that are supposedly from another language, and...
asked how far that language is from English. In the experiments mentioned in Coleman & Pierrehumbert (1997), participants were even asked to judge whether a word could be a word of English. Ohala & Ohala (1986) also report that the participants were asked to approximately judge how far a word is from the pattern of English words. The participants might thus very well have explicitly assessed the pseudowords on their similarity to existing words.

Wellformedness judgements might thus as well be labelled wordlikeness judgements. Frisch et al. (2000) argue that wordlikeness judgements are still correlated with the probability of sound combinations in the lexicon. Bailey & Hahn (2001), on the other hand, argue that lexical neighbourhoods have more influence on wordlikeness judgements than probabilities of sound combinations, although they also acknowledge that phonotactic factors have an influence. Nevertheless, their method to derive predictions from lexical neighbours is arguably more sophisticated than the probabilistic phonotactic factors to which they compare it, because abstractions over probabilities are not taken into account; the comparison might thus not be fair for phonotactic knowledge.

Hayes & Wilson (2008) and Albright (2009) do use abstractions to show that wellformedness judgements cannot always be explained by the lexicon, as the abstractions account for differences for unattested sound combinations. Still, it is always unclear whether a participant has made a wordlikeness or a wellformedness judgement. One might even argue that participants always tap into lexical knowledge, even when asked to judge the wellformedness of a non-word, as non-words are quite similar to words in length and most likely also in the prosody they are presented with. In other words, as non-words are almost words, they might be assessed with lexical knowledge. This lexical knowledge might obscure the effects of phonotactics, as probabilities in the lexicon are in general highly correlated with phonotactic wellformedness as predicted by any theorisation.

In sum, wellformedness judgements might have to be discarded as not strictly representative of phonotactic knowledge, as suggested by Bailey & Hahn (2001). Nevertheless, the issue is undecided; wellformedness judgements also reveal knowledge of probabilistic phonotactics that can only be inferred from the lexicon through a complicated learning mechanism. In addition, the learning mechanism itself might be biased towards certain sound combinations (Pertz & Bever, 1975). The issues sketched above are problematic for the current study into phonotactic effects on speech perception. As long as there is no definitive understanding of the lexical influence on wellformedness judgements, it cannot be corrected for. One can therefore never be sure that an effect found with wellformedness judgements is not actually lexical. A better
approach to look for the gradient effects of phonotactic knowledge is to look at tasks that are more natural than wellformedness judgments.

1.3.2 Facilitation of wellformed sound combinations

The representation of phonotactic knowledge in a probabilistic way is argued for by Vitevitch, Luce and colleagues on the basis of effects they found of phonotactic probability on speech processing (Vitevitch et al., 1997; Vitevitch & Luce, 1998, 1999, 2005; Luce & Large, 2001; Auer & Luce, 2006).

These authors identified a sublexical level at which phonotactics influences speech perception after a series of experiments on the facilitation and inhibition exerted on speech processing. The experiments in Vitevitch et al. (1997) and Vitevitch & Luce (1998) involved shadowing, i.e. repetition of an auditorily presented item. This was harder for the participant for a word with highly frequent phoneme combinations than for one with low probability phoneme combinations. This pattern was reversed for non-words, but non-words in general took longer to repeat. The authors argue that for non-words, phonotactic (sublexical) knowledge speeds up the repetition, but that for words, there is a larger negative effect due to the larger lexical neighbourhood density for the high probability items. In Vitevitch & Luce (1999), the lexical and sublexical levels are teased apart. In an experiment that involved same-different judgements, the pattern mentioned above was again found when non-words and words were presented separately. However, when they were mixed, sublexical processing was emphasised, as the participants did not know if they could expect a word or a non-word. In this case, the lexical neighbourhood effect is attenuated. The lexical part of processing can also be emphasised; in a lexical decision task, the lexical neighbourhood of the non-words strongly influences their processing, such that non-words with highly frequent phoneme combinations are now processed slower than those with low frequency phoneme combinations.

In a follow-up experiment reported by Luce & Large (2001), the lexical neighbourhood was manipulated as well, independently of the manipulation of phonotactic probability, unlike in the experiments reported above. In this case, the phonotactic probability of words positively affected their processing, while lexical neighbourhood density negatively affected it.

The effects of probability are explained by Vitevitch & Luce (1999) as the result of the involvement of sublexical representations of frequent phoneme combinations. These representations were already mentioned before. Speech recognition is assumed to be best modelled as in the ART model by Grossberg et al. (1997) (a precursor of ARTWORD). The sublexical units facilitate reso-
sance for input that contains these units with words that also contain them, by providing an extra pathway for activation.

Sensitivity to frequency is a general property of the mind. Humans are sensitive to probabilities of occurrence of all kinds of stimuli in their environment (Saffran, Newport, & Aslin, 1996, show this for auditory material), regardless of awareness; patterns are also learnt when they occur in the background, i.e. when participants are asked to pay attention to another aspect of the stimuli (e.g., Watanabe, Náñez, & Sasaki, 2001; Seitz & Watanabe, 2005, show this for visual material). There is no reason to suppose that this type of general statistical learning does not occur for sound combinations, as otherwise this would entail a specific exception for phonotactics.

Recall is higher for non-words that are composed of highly frequent sound combinations, as Gathercole, Frankish, Pickering, & Peaker (1999) showed for 7 and 8 year old children. In addition, Boll-Avetisyan (forthcoming) found indications that this probability effect is modulated by the syllable structure. So, phonological structure is very likely to be subject to statistical learning.

The importance of probabilities for processing is also confirmed by the findings of McQueen & Pitt (1996), who also report an effect of the probability of phoneme combinations, but now on phoneme recognition. They found that, for CVCC syllables, the first consonant of the coda can be spotted more easily when the transitional probability of the cluster is high. This effect was not found for CVC syllables, for which phoneme monitoring was easier in general.

The literature reviewed in this subsection indicates that there is a correlation between phonotactic probability and facilitation. This correlation is quite transparent, as both probability and facilitation can in general be defined as gradient but positive, i.e., they apply only to legal sound combinations. The analysis of these effects usually does not require constraints or the concept of markedness, but it does seem to hinge on certain abstractions.

1.3.3 Correction of illegal sound combinations

There is also evidence that knowledge of markedness, i.e. of constraints that affect the legality of sound combinations, affects speech perception. Speech perception seems to avoid illegal sound combinations; listeners often perceive them as the nearest legal combination, thus making perception in fact unfaithful to the acoustic input.

Polivanov (1931) found anecdotally that Japanese listeners could not perceive Russian syllables like ‘tak’, as Japanese does not allow the coda ‘k’; in fact, it does not allow codas with the exception of nasals homorganic with
the following consonant and consonant gemination of the onset of the next syllable.

Brown & Hildum (1956) found that a phoneme is recognised more easily when it is in a legal phoneme combination than when it is in an illegal one. An illegal cluster can be mapped onto a nearby legal cluster and thus perceived as such: Massaro & Cohen (1983) found that sequences of two phones that were made ambiguous between two phonemes were more likely to be classified as a legal combination. In one experiment, the first phone was ambiguous between /b/ and /d/ and the second between /l/ and /r/. The four logically possible combinations are /bl/, /br/, */dl/ and /dr/, the illegal combination /dl/ was however perceived less than the others; for instance, the phone ambiguous between /l/ and /r/ was more likely to be perceived as an /r/ when the first phone was perceived as a /d/ than when the first phone was perceived as a /b/.

Perceptual illusions do not even require an ambiguity in the acoustic input: Hallé et al. (1998) exposed French listeners to increasingly large parts of clusters that were illegal in French and found that listeners can have trouble recognising a consonant due to the following one. If the first consonant is followed by a second one that could not have followed in French, the perception of the first can be adapted. For example, */tl/ is illegal in French. The closest legal alternative is hypothesised to be /kl/. Therefore, it is harder to perceive a [t] faithfully as /t/ when the stimulus is [tl]. The same [t] was nevertheless perceived correctly as /t/ when the [l] was not presented afterwards.

Dupoux et al. (1999) showed that language-specific restrictions on syllable structure can reduce sensitivity to auditory contrasts. Japanese listeners have trouble distinguishing [ebzo] and [ebuzo], as the first is illegal in Japanese; French listeners have no trouble with [ebzo], but they do not accurately perceive the difference between [ebuzo] and [ebuzo], as vowel length is not contrastive in French; the Japanese listeners did not have trouble with the last contrast. This study is the precursor of the fMRI study by Jacquemot et al. (2003), that shows that different brain areas are activated when the brain processes a contrast that exists in the native language of the listener.

It is interesting to note that the behavioural effects described above have been corroborated by studies using visually presented stimuli. Different brain activation was found in PET scans by Petersen, Fox, Snyder, & Raichle (1990) for phonotactically legal pseudowords and actual words than for illegal pseudowords and words composed of pseudo-letters. Moro et al. (2001) also found that phonotactically illegal structures are processed with more activation in certain brain areas, as detected with PET, namely “Brocas area pars opercularis (BA 44) and the left inferior parietal lobule (BA 40); on the
right hemisphere, the lateral premotor area (Ba 6), the cuneus (Ba 18) and the middle occipital gyrus (Ba 19 and 18)” and “[b]ilateral activations included the superior parietal lobule (Ba 7), the precuneus (Ba 7), the fusiform gyrus (Ba 18/37), the cerebellum and the cerebellar vermis”. A network consisting of these areas was activated when Italian-speaking participants read out quasi-Italian sentences that contained phonotactic, syntactic or morphosyntactic anomalies. The participants had to detect the anomalies. A phonotactic anomaly is for instance present in the sentence “Il gulco gianigzleva le brale”, because ‘gzl’ does not occur in Italian. Compared to a baseline of syntactically, morphosyntactically and phonotactically correct pseudo-Italian, there was more blood flow to the areas mentioned above. However, as all these effects are actually based on written words, they might also be explained as orthotactic instead of phonotactic (see Bailey & Hahn, 2001).

Dupoux et al. (2001) assessed the possibility that perceptual illusions are caused by lexical information. They employed the fact that the epenthetic vowel that Japanese listeners perceive in a consonant sequence depends on the direct sequence. E.g., the illegal consonant sequence /kd/ is always perceived with the epenthetic vowel /u/: /kud/. Lexical information, provided by words, might have caused perception of another epenthetic vowel. However, the perceptual epenthesis turned out be induced by phonological factors alone, as it results in the same vowel (the most unmarked one), even if a lexical resolution was also possible. For instance, the non-word [mikdo] did not activate the word /mikado/, presumably because the cluster [kd] is always resolved by the illusion of an epenthetic vowel /u/, resulting in the perception /mikudo/.

Effects of categorical illegality on speech perception can thus be found by presenting listeners with a sequence that is not legal in their language, as listeners tend to have trouble correctly recognising illegal combinations. This is most likely beneficial in their own language, where these illegal combinations should not be present in the perception output and can thus be filtered out of the input. As mentioned above, this filtering is problematic in a second language with different phonotactics, as it can accidentally filter out possible words. Second language perception thus offers an opportunity to assess knowledge of illegality.

Davidson et al. (2007) found that illegal clusters are distinguished from epenthesised repairs more easily when language learners are presented with minimal pairs, e.g. when the word /z'gamo/ is learnt and contrasted with [zgamo]. This finding can be interpreted in many ways, including one in which the phonotactic information is used to reduce all phonetic contrasts that are not lexical contrasts. The cluster /zg/ is not in the lexicon and therefore does not have to be recognised. In addition, it is marked, so the listener might
prefer not to recognise it. As soon as the lexicon starts containing contrast between a previously illegal sound combination and its perceptual illusion, the perceptual illusion starts being contrastive and the illegal sound combination stops being mapped onto the illusion, hence losing its illegality.

This explanation is backed up by earlier results showing phonotactic effects on speech production. Davidson (2006) reports that the acquisition of illegal clusters is governed more by phonological factors than by lexical factors, showing that the contrastiveness that is acquired by the learners in Davidson et al. (2007) is not caused by the newly acquired lexical items themselves, but by reorganisation of the mapping of input to lexical items through phonological knowledge.

As mentioned above, Moreton (2002) found that there is more perceptive bias against [dl] than against [bw] for American English listeners, showing that wellformedness can be gradient even for unattested sound combinations. This again suggests that mere lexical probabilities, without any abstraction, cannot explain all effects of phonotactic knowledge on speech perception.

Filtering effects are not restricted to categorically illegal sound combinations. Categorical illegality in one language can correspond to gradient illformedness in another, as shown by Berent et al. (2007). They investigated the ban on onsets with a sonority fall like */lb/: These are illegal in many languages, but not in Russian; still, listeners with a knowledge of Russian are slower to classify syllables starting with these onsets as monosyllabic, as compared to onsets with a sonority rise. Failure to classify a syllable as monosyllabic is most likely due to perceptual repair; the illegal cluster is perceived as containing an epenthetic vowel between the two consonants. English categorically bans onsets with a sonority fall, but allows onsets with a sonority rise. Hence, the difference that is categorical in English is gradient in Russian. Berent et al. explained their results as a universal preference for rising sonority in onsets. This preference thus has to be gradient, as the preference means that the closer more the onset approaches the ideal, the easier it is to perceive as one cluster. On the other hand, a language can impose a categorical boundary on the gradient scale, as witnessed by English.

Perceptual illusions might seem to be an unnatural application of knowledge of the phonotactic constraints of a language; illegal clusters are supposed to be avoided in the input. However, this is not completely correct. Illegal sound combinations by definition fail to occur in the canonical forms of words, but in spite of this, illegal sequences of phonemes are plentiful in continuous speech, namely at word boundaries and also in acoustic reductions. Perception can thus potentially benefit from knowledge of phonotactic illegality, because when the listener knows a combination to be illegal, he can hypothesise a
The repair of illegal clusters by perceiving an illusory epenthetic vowel is not the only possible way illegal clusters are resolved. Lexical knowledge can also help. Acoustic reductions can result in illegal phonotactics on the surface. Ernestus (2000, p. 131, ex. 21b–c) mentions a large number of acoustic reductions, some of which result in illegal clusters, like the following two:

- /vɛrsxil/ende → [ʃsxIl]ende (verschillende, ‘different’)
- /bei'vor/beeld → [ˈbvor]beeld (bijvoorbeeld, ‘for example’)

(The examples are only transcribed for the syllable with the reduction.) As Ernestus notices, the clusters /ʃsx/ and /bvr/ are not legal in Dutch. Although an empirical study of the phonotactic knowledge employed in the perceptual resolution of acoustic reductions would be interesting, it would not be free of a confound: listeners are likely to recognise the canonical form of the word and use this top-down information to resolve the acoustic reduction. As noted later by Pluymaekers, Ernestus, & Baayen (2005), reductions occur more often in more frequent words, showing that properties of the word are at least of influence for reductions, whilst it remains to be seen if this also leads to easier perception of reduced forms when the canonical form to be recognised is a frequent word. Ernestus (to appear) argues in addition that both the storage of pronunciation variants and phonetic decoding are necessary to account for the recognition of reduced words.

The use of phonotactic knowledge in the normalisation of reduced forms has nevertheless not been conclusively established. The possibility of a perceived contrast between an illegal non-word and the nearest actual word, as reported by Dupoux (2001) (discussed above), suggests that lexical knowledge alone cannot explain all perceptual illusions, but the recognition of reduced forms suggests that phonotactics is also not always the only explanation. This interplay will be left aside in this thesis, as the perceptive resolution of acoustic reductions is still not completely understood.

The literature reviewed in this subsection suggests a special position for phonotactic illegality in speech perception. The perceptive system takes special care to eliminate illegal combinations, through a variety of repairs. Again, the relation between knowledge of illegality and its effect is quite straightforward: illegal sound combinations are filtered out, even when there is no problem with the recognition of the individual sounds. Such effects have not been reported for legal items, although processing legal but less wellformed items might take more time and effort, according to Berent et al. (2007).

Representations of illegality are easily captured as markedness constraints. The effect of these constraints does not have to be categorical, as discussed
in §1.2.2 (p. 27). On the other hand, probabilistic knowledge in its simplest form cannot capture any differences between unattested combinations. This implies that even if perceptual illusions have to be attributed to probabilistic knowledge, this knowledge is represented as an abstraction of the actual probability of sound combinations.

1.3.4 Segmentation of continuous speech

Illegal clusters occur in the aforementioned acoustic reductions, but also across word boundaries in continuous speech and thus provide potential cues for word segmentation. As discussed above, Norris (1994) and Norris & McQueen (2008) suggest that the words that can be possibly found in the speech stream are activated and then compete; overlapping words inhibit each other, thus leading to recognition of a segmentation that uses every phoneme once. Nevertheless, sublexical cues are also relevant for speech segmentation. Phonotactics is one of these cues, but see Mattys et al. (2005) for an informative overview of all sublexical cues and their relation to speech segmentation.

Phonotactic cues have been shown experimentally to affect the recognition of words in continuous speech (Norris et al., 1997; McQueen, 1998; Gaygen & Luce, 2002; Mattys et al., 2005). McQueen found that a word that is phonotactically aligned is easier to find than a word that is phonotactically misaligned. He embedded words, e.g. the Dutch word /rɔk/ rok, ‘skirt’, in two different contexts. In one, e.g. [fimrɔk], the illegality of a consonant cluster, e.g. */mr/, signals a boundary aligned with the word. In the other context, e.g. [fidrɔk], the word boundary cannot be seen as a phonotactic boundary, as it leaves a voiced coda which is illegal in Dutch. The word is thus assumed to be misaligned with a phonotactically induced boundary before the voiced obstruent. The word is the first syllable and thus aligned or misaligned at its offset, instead of its onset, e.g. the word [dal] dal ‘valley’ was easier to spot in [dalrœyp] than in [dalmrœyp]. The phonotactic effect also occurred there, albeit in a weaker form. This is most likely the result of the fact that after all phonemes of a word are already processed, the word might already be identified and provide information on the way it ends, making the phonotactic cue less necessary to reach recognition of the target word. Phonotactic information and lexical information both contribute to word segmentation, with lexical information dominating phonotactic information in case of a conflict, as shown by Mattys et al. (2005), who pitted both cues against each other.

The effects reported by McQueen (1998) can be interpreted as a facilitatory effect of phonotactic illegality, but also as the effect of the higher wellformedness of the legal onsets. In the example pair [fimrɔk]–[fidrɔk], the recognition
of the target word /rak/ can be hindered by the higher probability of /dr/ vs. the unattested onset /mr/, or helped by the illegality of */mr/ vs. the legal /dr/. Segmentation is thus an interesting testing ground to compare a phonotactic grammar based on markedness constraints or one based on probabilities. Nevertheless, both types still explain the results of McQueen (1998).

Weber & Cutler (2006) found that in speech segmentation in a second language, listeners use knowledge of illegal sound sequences that originates from their first language, as well as markedness knowledge from their second language. German phonotactics forbids e.g. /sl/ as an onset, while this is legal in English. On the other hand, English forbids e.g. /l/ in onsets, while this is legal in German. The participants in the experiment by Weber & Cutler were Germans who were highly proficient in English, as well as English native listeners that did not know German. They had to spot words in nonsense strings and the illegality of clusters provided a benefit to word recognition, because such clusters cue a boundary aligned with the word. E.g., in the case of the nonsense strings [darslidzn] and [darSlidzn], containing the English word /lidzn/, ‘legion’, the German learners used both English and German illegality knowledge and were able to benefit from it and spot the word with greater ease. English listeners did not consider /sl/ illegal, as it is not illegal in English; therefore they do not benefit from facilitation provided by the German illegality of /sl/ that cues the segmentation /dars.lidzn/. The English listeners only spotted words with greater ease in the case of [darlidzn], as /l/ is illegal in English. This suggests that the knowledge that the Germans acquired of English is indeed knowledge of markedness, namely that they added /l/ to their inventory of marked (illegal) clusters. They would be predicted to have greater trouble, not less, if they would have acquired the higher probability of /sl/ in English instead of the markedness of /l/ in English.

Speech segmentation, especially in a second language, thus indeed provides an interesting process with regard to investigation of the effects of phonotactic knowledge. Phonotactic markedness constraints are predicted to have a facilitative effect on perception, guiding perception towards a correct segmentation, in contrast to the inhibitory effect of perceiving marked clusters as shown by perceptual illusions. In addition, the use of phonotactic knowledge in segmentation can theoretically also consist of a bias against breaking up wellformed clusters; hence, probability might have an inhibitory effect instead of its usual facilitatory effect.
1.3.5 Unification of phonotactic effects

The psycholinguistic literature discussed so far shares an important aspect: the effects of phonotactic knowledge were never observed as completely categorical in an experimental setting. This even holds for the correction of illegal sound sequences, where the confusion between an illegal sequence and the nearest legal one was never 100%.

Although it is by definition impossible for any effect to be “partially categorical”, the above does not mean that none of the effects is actually the effect of categorical knowledge. In fact, the effect itself could be categorical, but the observation might be noisy. In addition, the speech recognition process has many aspects and is not dominated by phonotactic knowledge. A sound combination might be completely illegal, which might cause the phonotactic system to reject it categorically, but the phonotactic system may simply not be influential enough to cause this sound combination not to be recognised. Speech perception is too flexible and complex to make it possible to isolate the effect of phonotactics. For this reason, it is rather difficult to interpret the evidence from psycholinguistic experiments equivocally as showing gradient or categorical phonotactic knowledge. This makes it harder to prove the existence of constraints, which are in principle categorical representations of markedness.

Regardless of the formalisation of categorical illegality by means of markedness constraints, gradience or categoricalness of the effects of phonotactic knowledge do not directly follow from the categoricalness of the representations. To classify the representations as categorical, theoretical considerations are needed, as experimental evidence will not give the definitive answer. It is more fruitful to classify effects as restrictive or facilitating. Marked combinations are often avoided in perception, i.e. the perceptive system has a bias against them and tries to filter them out. On the other hand, legal or highly wellformed and/or frequent combinations are facilitated; the perceptive system is biased towards recognising these combinations.

Keeping in mind that gradience and categoricalness in effects can be extracted from one phonotactic system, containing categorical or gradient representations, the main question to answer is whether restriction is the counterpart of facilitation in the sense that the absence of restriction entails facilitation (or vice versa). The gradience of the effects of phonotactics on speech perception nevertheless makes it hard to establish if an effect is inhibitory or facilitatory.

In addition, there is as yet no convincing empirical argument to assume phonotactic markedness to be represented differently from phonotactic wellformedness. The existence of the two concepts in phonotactic theory can
be attributed to two separate scientific traditions, theoretical phonology and psycholinguistics. Phonologists and psycholinguists have separately fulfilled the principles of parsimony in their descriptions, as they normally propose only representations of markedness or of probability.

Phonological theory has capitalised on the avoidance of illegal combinations and the finding that certain combinations are never preferred, even if they are sometimes allowed. A theory based on markedness, i.e. a set of constraints, is an inventory of markedness. As gradient markedness can be assigned by a phonotactic constraint-based grammar, gradient effects can be explained with a constraint-based grammar. This in theory allows for facilitatory effects to be seen as the absence of inhibitory effects.

On the other hand, psychologists that study word recognition have focused on the effects of phonotactic probability instead of markedness. Probability is positive and gradient by definition. These effects include the facilitation of speech perception that occurs for highly frequent combinations. This has forged psycholinguistic descriptions of phonotactics into positive terms, by separating the wellformed combinations from the less-wellformed ones. If phonotactic grammaticality is based on probability, probabilities also have to account for the inhibitory effects normally ascribed to markedness constraints. Inhibitory effects then have to be analysed as the absence of probability and competition by other, more probable perceptual candidates.

Parsimony demands combining descriptions of phonotactic knowledge in a unified theory, as stated above. However, any unification that consists only of listing all theoretical descriptions and all empirical evidence together contains the possibly spurious dichotomy between negative and positive knowledge and effects. This issue can be addressed by looking for empirical differences between the effects of probability and markedness. If these do not exist, the dichotomy is not needed to explain any observations and it should be removed from theory; if it happens to be a relevant distinction, the two different types of knowledge representations have to be investigated and described.

Before discussing the methodology for performing the tests and investigations described above, it is necessary to take stock of the repercussions of the literature and theory discussed above in relation to phonotactic knowledge in speech perception.

Existing models of speech perception can be straightforwardly made to incorporate probabilities of phoneme combinations. However, the question is whether probabilistic knowledge subsumes a phonotactic grammar that generates markedness. Pure probabilistic knowledge cannot explain all effects ascribed to phonotactics; hence, at least some form of abstract knowledge has to exist. Hence, even if phonotactic knowledge is represented in the form of the probability of sound combinations, this probability is not isomorphic to
raw frequency counts. Nevertheless, this does not matter for the incorporation of probabilistic phonotactic knowledge into a word recognition model; if probability can be modelled as connectivity or a priori probability, it can just as well based on frequency as on an abstract derivation from probabilities. Hence, grammaticality based on probabilities can easily be incorporated into a speech recognition model as well.

If phonotactic grammar is only probabilistic, its incorporation into a word recognition model is elegantly modelled in the way Vitevitch & Luce (1999) propose, namely as a representation for sound combinations, either weakly or strongly connected to acoustic input on the one hand and lexical items on the other hand. The stronger the connections are, the easier the sound combination resonates. Shortlist B can also capture the influence that phonemes have on each other. The a priori probability of a phoneme string can be incorporated in the model, which gives the model a bias towards perceiving high probability strings.

The empirical counterpart of such incorporations of probabilistic phonotactics in word recognition models would be a pulling effect from the more probable percepts; they are more likely to be in the output, whatever the input is. This pull does not capture the origin of the movement; marked input does not lead to more pull by probable percepts. Probability thus does not explain perceptual illusions, as these are not related to a pulling but a push effect, away from marked sound combinations.

As both effects, facilitation and inhibition, are part of the explanandum, their independent existence might license a theory in which there are both pulling effects from probability and pushing effects from markedness knowledge. However, as it is theoretically possible to unify markedness to very low probability, or high probability to the absence of markedness, a unification of the phonotactic status of a sound combination is most parsimonious. A sound combination is then not both marked to a certain extent and probable to a certain amount, but only one of them.

To investigate the phonotactic status of a sound combination, and hence of the knowledge that generates that status, markedness and probability have to be tested in conjunction, but manipulated separately. In addition, the effects of either markedness and probability on word recognition have to be related to the knowledge that generates them. Both markedness and probability knowledge can be connected to word recognition models, as discussed above.

Still, it is also an empirical question if phonotactic effects indeed originate from a separate representation of phonotactic knowledge, i.e. if phonotactics is represented separately or if it emerges from other sources. The most parsimonious theory is that there is no independent phonotactic knowledge. Alternatively, there might not be knowledge of either markedness or of prob-
abilistic wellformedness. If one of these types of knowledge does not exist, the supposedly phonotactic effects on word recognition have to be ascribed to other sources of knowledge. The most likely candidate is the lexicon, as it has phonotactic properties. Another source of apparently phonotactic effects might be the acoustic properties of sound combinations. It might just be harder to perceive allegedly phonotactically marked sound combinations because they are physically (acoustically) hard to process. In fact, there is usually a relation between phonological markedness and phonetic grounds for such markedness.

In sum, the effects of markedness and probability on speech perception have to be contrasted and phonotactic effects on speech perception in general have to be contrasted with lexical knowledge, as well as with intrinsic properties of sound combinations. However, the required contrasts cannot easily be made. It is normally not the case that a sound combination is both marked and probable enough to be both pull perception toward it and push it away. It is also not normal for lexical knowledge to not interfere with speech recognition or interfere in another direction to phonotactic knowledge. Lastly, if a sound combination is phonotactically marked, it is likely to be phonetically marked as well.

Nevertheless, the fact that phonotactic knowledge is language-dependent implies that a sound combination can be more or less marked in one language than another. Differences of this type cannot only be attributed to phonetic properties and have to be ascribed to phonotactics, if parsimony is taken seriously. In addition, it is not possible for a monolingual to have phonotactic knowledge not strongly correlated with lexical knowledge. However, a second language learner might acquire words in a new language that do not follow the phonotactics of the native language, allowing the two types of knowledge to diverge. For these reasons, second language listening is an important part of the research presented in this thesis.

Second language acquisition also allows comparing language users with different phonotactics in the same task, e.g. a native listener and a second language listener can be assessed when processing the same word in the target language, while the word’s phonotactic wellformedness differs as this is decided by either the target language or the native language of second language listeners. The latter depends on the assumption that second language listeners transfer phonotactic knowledge from the first language. Transfer occurs for many phonological phenomena (e.g. Escudero & Boersma, 2004), and has been found for phonotactics (Weber & Cutler, 2006) as well, but it remains to be seen if phonotactic knowledge of all types is transferred.
1.4 METHODOLOGY

1.4.1 Research questions

Phonotactics has above been argued to show in speech recognition. However, there is no consensus on the type of knowledge that causes the phonotactic effects. The approach of this thesis is to look for phonotactic effects in a range of speech perception tasks that make it possible to establish whether effects of phonotactic markedness and probability are signs of independent phonotactic knowledge and if they can be reduced to a unified notion of phonotactic wellformedness.

The first question is thus if phonotactic knowledge indeed consists of two parts: markedness and probability. More precisely, the question is if it is indeed impossible to tie the effects ascribed to phonotactic markedness and phonotactic probability together, in accordance with a unified account of phonotactic wellformedness.

The second research question is exploratory. The question is to what extent phonotactic knowledge operates independently of other recognition processes when it shapes perception. One extreme is that it just modulates activations or probabilities of perceptions at higher levels, notably the lexical level. The other extreme is that phonotactics fully organises the acoustic input into a string of pseudowords, that are then matched to the lexicon.

The third question is whether phonotactic effects are language dependent. Language-dependency would comply with the theoretical arbitrariness of phonotactic knowledge; markedness might be explained on phonetic grounds, but does not have to result in illegality in every language.

Second language listening experiments also allow to look for the effects of language mode; if it is possible that a listener applies phonotactic knowledge to the language he is listening to, the knowledge includes the language to which it belongs. This can also indicate that phonotactic knowledge is not mere probabilistic knowledge, as probabilities would be predicted to be added together over all languages, if phonotactic probability is simply isomorphic to frequency of occurrence.

1.4.2 Outline

Three types of observations have been used for the research reported in this thesis: the phonotactic effects on word recognition, perceptual illusions and word segmentation. This choice was made because it allows contrasting facilitation and inhibition in different ways.
This thesis tries to establish whether the theoretical differences between phonotactic probability, phonotactic legality, gradient wellformedness and phonotactic markedness can be established as separate forces with their own effects on speech perception.

**Probabilistic wellformedness vs. illegality**

The first question that will be addressed is if a second language learner, presented with the probability distribution of a non-native language, will first lose markedness effects and then acquire probability effects for sound combinations that are frequent in the new language but illegal in the first. This order of acquisition would be predicted by a unified representation of phonotactic wellformedness. If the learner derives phonotactic grammaticality from probabilities, grammaticality shifts with the change in probability, towards the target language. If markedness is low probability, it should shift together with the probability.

In the experiments presented in Chapter 2 (p. 55), the question stated above was addressed by observing native Dutch listeners and learners of Dutch with either Spanish or Japanese as a native language. The observations concerned two effects typically explained by illegality (markedness) and probability, combined into a single experiment: perceptual filtering of illegal combinations and facilitation of frequent sound combinations. The experiments of Chapter 2 show that these effects operate at different levels in the speech process. These levels are interpreted as a categorical filter that only allows legal percepts and a gradient enhancement for the most frequent percepts. In other words, there are two differences in the effects of markedness and probability. First, first language illegality seems to categorically filter out a contrast in a second language, while probability of occurrence of a subset of the marked structures also causes a bias. However, the probability is enhanced for the unmarked perceptual illusion, that cannot be assumed to actually be present in the input.

These findings suggest that perception is shaped by markedness before it continues to lexical recognition, where probabilities are taken into account. Hence, the findings also suggest that both markedness and probabilistic knowledge exist independently. The types of knowledge are then studied in speech segmentation. As markedness can facilitate segmentation, while high probability can inhibit it, this can provide a convincing corroboration of the effects found with perceptual illusions and word facilitation (in Chapter 2).

**Illegality of clusters vs. legality of segmentations**

Speech segmentation can offer a new window on phonotactic knowledge, as knowledge of illegality can actually facilitate perception, by giving informa-
tion about word boundaries in the speech stream. It first has to be established that phonotactic effects also independently shape perception when it comes to speech segmentation. In Chapter 3 (p. 95), the place of phonotactic knowledge in the process of speech segmentation is explored. The question is whether the speech segmentation can be driven by phonotactics alone. Alternatively, lexical knowledge segments speech and phonotactics facilitates wellformed segmentations, or inhibits illegal segmentations. As speech is always segmented into words, the fact that a phonotactically illegal structure is a cue for a word boundary can be alternatively explained as the absence of other word onsets that compete with the cued boundary. By applying eye-tracking, the effect of phonotactics is observed in the absence of words. Hence, it can be assumed that the effects that are observed with the new method indicate actual phonotactic knowledge, not side-effects of lexical recognition.

**Categorical and gradient wellformedness**

The method therefore allows to establish whether illegality and/or wellformedness are behind the phonotactic effect on speech segmentation. Experiment 4a shows that the phonotactic effect on segmentation cannot completely be explained by a preference to chunk wellformed combinations; there is at least an influence of illegality at the level of the segmentation. The results are compatible with the view that the whole phonotactic effect on segmentation can be explained by markedness knowledge. The results of studies by Weber (2001), mentioned above, also suggested that illegality, not wellformedness, is transferred by German highly proficient listeners of English, from their native language German to English, but their study was not designed to reject wellformedness. In Experiment 4b, second language learners have been tested that cannot have transferred illegality knowledge, as the combinations under scrutiny are legal in their native languages. These native languages were all Slavic languages. These participants showed a more gradient effect, which matches the fact that the clusters that are either legal or illegal for the Dutch listeners of Experiment 4a are all legal in the source languages of these listeners, but the ‘illegal’ ones still more marked. Nevertheless, the effects shown by these second language listeners does not exclude markedness as the only type of phonotactic knowledge that affects speech segmentation.

The results of Chapter 3 suggest at least that constraints on sound combinations have a gradient effect on segmentation, but probabilistic and categorical knowledge is not excluded.
Acquisition of second language phonotactics: illegality and legality

The Spanish listeners of chapter 2 and the Slavic listeners of 3 seem to have trouble switching off categorical markedness, as this type of knowledge filters out its own counter-evidence. Hence the question arises whether categorical knowledge of another language can be acquired at all. In Chapter 4, the language-specificity of phonotactic knowledge is assessed with the same eye-tracking methodology as in Chapter 3. Dutch learners of English show that the acquisition of a categorical cue for segmentation in English is possible and that the cue is applied more to English than to Dutch. Phonotactic knowledge is thus represented as part of one particular language. This means that the simplest probabilistic acquisition cannot explain phonotactic knowledge; if the learners had just kept track of frequencies in all of their input, they would not have different sets of phonotactic knowledge, but one set skewed towards the language most frequently encountered.

1.4.3 Theoretical unification

The results of the experimental chapters call for a theoretical account that can as parsimonious as possible explain the effects of phonotactic knowledge on speech perception. The last part of this thesis is an inventory of the theoretical implications of all findings and a proposal for a grammar that can generate the appropriate effects of phonotactics on word recognition and speech perception.
PROBABILITY AND ILLEGALITY IN WORD RECOGNITION

ABSTRACT

Phonotactic wellformedness facilitates word processing, while phonotactic illegality can cause perceptive illusions. Theoretically, illegality could be negative wellformedness, unifying these two types of representations. This is tested empirically by combining facilitatory and illusion-inducing effects of phonotactics within one cross-modal priming experiment with native and second-language listeners of Dutch. Due to the first languages of the second language listeners, they have a perceptual illusion of epenthesis for the /s/+consonant (sC) combinations, illegal in the first language, allowing a test of the effect of illegal sound combinations on word processing. Dutch listeners recognised words starting with wellformed sC-clusters faster than words with sC-clusters that were less wellformed, but legal. This advantage disappeared when the clusters were epenthesised. Wellformed clusters thus have a sublexical representation without the epenthetic vowel, facilitating mapping of input to lexical forms. Two L2 groups, one with Japanese as L1, one with Spanish as L1, participated. Both L1’s do not allow /sC/ clusters and repair them with epenthesis. The Spanish group did not differentiate between the presence or absence of epenthesis, showing a perceptual illusion; in contrast, the Japanese L1 group did. For the Spanish group, recognition of words containing well-formed clusters improved with increasing proficiency in Dutch, but only when presented with epenthesis. L1 illegality thus filtered the input from which the Spanish listeners learned Dutch wellformedness, showing that illegality operates in a way different from wellformedness. Illegality knowledge filters irrelevant details from speech, wellformedness knowledge makes matches between input and lexical entries more efficient.

Keywords: speech perception, phonotactics, word recognition, speech processing, cross-modal priming

2.1 INTRODUCTION

Representations of native language knowledge heavily influence speech processing. Non-linguistic speech perception is sensitive to details that linguistic perception ignores; with increasing demands on processing as posed by natural speech, listeners resort to linguistic knowledge to abstract their perception
towards the categories of their native language, for instance by remembering
a sound as a phoneme, stripping it of linguistically irrelevant phonetic details
(Werker & Tees, 1984). Observations of linguistic processing allow insight
into the psychological reality of theoretical concepts such as the phoneme
(J. J. Jaeger, 1980; Morais & Kolinsky, 1994; Dehaene-Lambertz, 1997). This
is especially evident in the misapplication of native language knowledge to
a second knowledge (see e.g. Escudero & Polka, 2003; Escudero & Boersma,
2004; Escudero, 2005).

Phonotactics, the topic of this thesis, also influences speech processing; this
influence indicates the existence of representations of phonotactic knowledge
similarly to how the phonemic effects on language processing indicate the
psychological reality of phonemes. Different researchers, using different
methods, have gathered a number of observations of phonotactic effects in
speech processing, but they have also chosen different types of theoretical
descriptions of the psychological representations underlying the observations.
These researchers have made an effort to be parsimonious, but when their
theoretical assumptions are lumped together, the result is a set of different
types of representations, mainly differing in the use of categorical negative
knowledge (phoneme combination AB is illegal) or of positive probabilistic
knowledge (combination CD is more wellformed than AB). However, the total
set of phonotactic observations does not license the theoretical complexity;
many accounts can be combined in a simplified way, for instance by assuming
the absence of negative knowledge to entail positive knowledge or vice versa.

Such a unification is theoretically viable and could capture existing ob-
servations on phonotactic effects on speech processing. Nevertheless, the
unification has to be empirically tested to discover if it is also psychologically
real, or alternatively, if there is a cause for the assumption of multiple types of
representations of phonotactic knowledge.

Descriptions of phonotactic knowledge can be classified as either cat-
egorical or gradient. Categorical knowledge represents which (phoneme)
combinations are illegal, or alternatively, which are legal. Illegal combinations
are processed differently: linguistic perception steers away from illegal struc-
tures, even if this makes the percept unfaithful to the actual input (Polivanov,
1931; Brown & Hildum, 1956; Massaro & Cohen, 1983; Hallé et al., 1998; Kabak
& Idsardi, 2007). Figure 2.1 schematically shows the place of psychological
evidence for categorical phonotactics: perceptual illusions transform illegal
into legal phoneme combinations.

The second way to describe phonotactic knowledge is to make statements
about gradient wellformedness (Moreton, 2002; Pierrehumbert, 2003; Albright,
2009). Gradient wellformedness usually correlates with lexical frequency and
is sometimes equated with it. Whether or not the source of gradient well-
perceptual illusions

\[
/\text{lb}/ /\text{bz}/ /\text{tl}/ \ldots /\text{lb}h/ /\text{buz}/ /\text{kl}/ \ldots
\]

Figure 2.1: Legality in phonotactics. The phoneme combinations are taken from Berent et al. (2007); Dupoux et al. (1999) and Hallé et al. (1998), but are merely meant as an example. Perceptual illusions are directed from illegal phoneme combinations to legal combinations and evidence for perceptual illusions can be used to posit the legality or illegality of a phoneme combination.

formedness is the lexicon, gradient wellformedness knowledge is represented separately: perception of ambiguous input can be biased by lexical entries (Ganong, 1980; Samuel, 1981), but there is also an influence of wellformedness on non-lexical speech processing. Gradient differences in sublexical wellformedness affect accuracy and speed of processing: items containing wellformed phoneme combinations are easier to repeat than items with infrequent (but legal) combinations (Vitevitch & Luce, 1998). Luce & Large (2001) found that this also holds for words, when the lexical neighbourhood’s contrary effect is strictly controlled. More wellformed phoneme combinations are also more likely to be registered in short-term memory (Gathercole et al., 1999). This empirical evidence licenses the assumption that there is a scale of wellformedness, on which legal phoneme combinations occupy positions relative to each other, as shown in figure 2.2.

\[
/\text{fi}/ /\text{kæ}/
\]

memory, lexical processing

Figure 2.2: Gradience for legal phonotactics. The biphone placement on the scale is an example from Vitevitch & Luce (1999); biphones with high probabilities have a facilitating effect on lexical processing.
Not only legal phoneme combinations, but also illegal ones have gradient wellformedness — or better, illformedness — as shown by Moreton (2002), Berent et al. (2007) and Coetzee (2009). The source for this kind of gradience might be a generalisation from knowledge about legal clusters (Kabak & Idsardi, 2007; Albright, 2009) and as such require grammatical generalisations; lexical frequency counts for the actual illegal structure will certainly not explain any gradient illformedness, as they are all zero. For this reason, the picture for illformedness, as given in figure 2.3, has to be made based on experimental evidence and cannot be based on lexical frequencies alone.

Figure 2.3: Gradience for illegal phonotactics. The biphones’ placement on the scale is taken from Berent et al. (2007). Clusters with high illformedness are more likely to suffer from a perceptive bias.

Illformedness and wellformedness are intuitively quite easy to connect, as they seem to be the same thing on two sides of the categorical legality boundary. The categorical difference between legal and illegal structures (figure 2.1) can be assumed to be simply a position on a wellformedness scale that can be applied to all phonotactic structures, as shown in figure 2.4.

However, even though this unification is visually quite straightforward, it includes an important generalisation that is not directly licensed by empirical evidence. If gradient wellformedness, illformedness, and categorical legality, are essentially a result of phonotactic knowledge, constituting values on a

Figure 2.4: Phonotactic gradience and legality combined. The biphone placement on the scale is only meant as an example.
2.1 introduction

wellformedness scale, evidence for the three types of phonotactic status has to show the same type of underlying knowledge. However, evidence for wellformedness, illformedness and legality, as mentioned above, comes from different types of observations. In addition, hypotheses about their origins in phonotactic knowledge diverge. The validity of the unification of descriptions of phonotactic knowledge thus does not follow from the theoretical description of phonotactics.

Empirically, unification entails that the phonotactic values assigned to different phoneme combinations for different reasons are intrinsically comparable. This means that the empirical observations on gradient and categorical wellformedness have a single underlying cause, namely degree of wellformedness. This entails that relations on the wellformedness scale are transitive; if \( A < B \) and \( B < C \), then \( A < C \). In a unified account of phonotactic representations, it does not matter if evidence for \( A < B \) is found in perceptual illusions, while the evidence from \( B < C \) comes from facilitated speech processing. Falsification of the unified account would occur if \( A < B \) and \( B < C \), but \( C > A \), or if \( A < B \) and \( B > A \), when different observations commonly associated with phonotactic wellformedness are made.

This chapter investigates the psychological reality of the unification of categorical and gradient wellformedness, by looking at both perceptual illusions and wellformedness-driven facilitation within one experimental design. The prediction of a unified model is that there is a correlation between the effects of categorical legality and the effects of a value on the wellformedness scale, as the same phonotactic knowledge underlies both. Note that illformedness and wellformedness are (for now) considered as the same phenomenon; the focus is on the distinction between categoricalness and gradience. For this investigation, effects of both categoricalness and gradience need to be identified. First, existing literature on the effect of categorical and gradient phonotactic wellformedness/legality is discussed.

2.1.1 Categorical phonotactics in speech processing

Perceptual illusions from an illegal to a legal structure have been repeatedly found. Polivanov (1931) reported anecdotal evidence from a Japanese listener who could not faithfully perceive Russian words that were illegal in Japanese, instead adding epenthetic vowels to their perception of consonant clusters. Polivanov’s findings have been supported by research under laboratory conditions. Halle et al. (1998) used a gating experiment to show that perception of French consonant combinations is biased towards phonotactic legality when there is a plausible alternative, using the illegality in French of \(*/t\Delta/** and \(*/d\Delta/**. The first consonant is perceived faithfully when the gate does not include the
/l/, but perception changes when the whole cluster is presented to listeners. Dupoux et al. (1999) found a double dissociation: phoneme combinations illegal in French but not Japanese are hard to distinguish for French participants, but not for Japanese ones, and vice versa for combinations illegal in Japanese. This shows that phonotactically induced perceptual illusions are language-specific, reflecting illegality in the native listener’s language. Segui, Frauenfelder, & Hallé (2001) even found a neural correlate: different brain regions used for the processing of legal and illegal structures (see also Moro et al., 2001).

Phonotactic knowledge also affects ambiguity resolution. Massaro & Cohen (1983) showed that when the perception of a sound that is phonetically between English /l/ and /r/ is biased by the phonotactic context it is more likely to be perceived as the phonotactically correct phoneme, i.e. more as /l/ after /s/ and more as /r/ after /s/. The phonotactic effect is larger when the phoneme is not close to either /l/ or /r/. When two consecutive sounds are both ambiguous, participants are less likely to identify the combination as an illegal cluster. A sequence of a sound token ambiguous between /b/ and /d/, followed by one ambiguous between /l/ and /r/, was presented to participants. The four possible percepts (and answers) are /dl/, /dr/, /bl/ and /br/, of which the first is illegal. The answer /dl/ was given less; Massaro & Cohen argue that this can only be explained because the listeners know that /dl/ is an unlikely combination, because the same tokens were more likely to be perceived as /d/ and /l/, respectively, when this percept resulted in a legal combination.

However, it is not the mere absence of a structure in a language that causes a listener not to perceive this structure. Some unattested structures can be perceived. Whether this is the case depends on the phonological principles the structures might violate, as Kabak & Idsardi (2007) showed for different consonant clusters unattested in Korean. In addition, Albright (2009) noticed that the absence of the sound combination [esp#] (# denotes word offset) in English is no reason to consider it illegal; it is perfectly acceptable to most English listeners and one would not expect [esp#] to be filtered out by perceptual illusions like [isp#] or [esk#].

Representations of phonotactic illegality can thus not be inferred from the mere absence of a certain combination; a phoneme combination can be absent from a language without the speakers of that language actually ‘knowing’ this. Representations of categorical differences thus have to be detected experimentally.
2.1.2 Gradient phonotactic knowledge

Facilitating effects of frequent phoneme combinations have been found in a number of speech perception tasks. Vitevitch & Luce (1998) identified two levels in speech processing: lexical and sublexical. The lexical level consists of words, typically competing for recognition. The sublexical level has among its ‘representational units’ phonemes, that provide support to lexical items when the units are activated due to their presence in the acoustic input. The two levels each have a different impact on recognition: inhibitory and facilitatory, respectively. On the competitive lexical level, higher wellformedness is correlated with a larger neighbourhood density for words, as wellformedness was defined as the lexical frequency of the biphones in the words. This makes wellformed words harder to process on the lexical level, as there is more competition. However, wellformedness has a positive influence on the facilitatory sublexical level, as observed in non-lexical tasks. Luce & Large (2001) later disentangled the effects and showed that the lexical neighbourhood effect can be controlled for; in such a case, high phonotactic probability has a facilitatory effect only. The Luce & Large (2001) study does not make explicit which highly-frequent combinations are represented sublexically; every experimental item is in general composed of frequent biphones and phonemes that are frequent in their position in the word, but which of these biphones or phonemes actually cause the effect — and thus are actually represented sublexically — is not the topic of that study, nor of its predecessors, Vitevitch & Luce (1998, 1999). These experiments thus indicate that sublexical, phonotactic representations exist, but not what combinations they concern. Unfortunately, different types of frequency counts can give different results; it is for instance not necessarily the case that phonotactics is learned from lexical statistics. Adriaans & Kager (2010) argue that continuous speech can be the source of phonotactic knowledge, which sometimes leads to different results than extracting phonotactics from the lexicon. It is therefore also necessary to obtain empirical support for assumptions about representations of wellformedness.

Frisch et al. (2000) showed that phonotactic wellformedness (in various measures) also has a positive effect on wordlikeness judgements; Bailey & Hahn (2001) confirmed this effect, but also found that lexical neighbourhood density was a more important factor for participants deciding whether a non-word was wordlike. In short, phonotactic wellformedness is correlated with lexical neighbourhood density, which can sometimes obscure its facilitatory effects, but wellformed phoneme combinations facilitate speech processing, making it easier for wellformed units in the input to be perceived as wordlike and/or to be linked to lexical entries.
In light of the above, there are ample indications that higher phonotactic wellformedness leads to a facilitation of processing, but there is no consensus on the calculations or derivations that can be used to decide if a sound combination is wellformed enough to facilitate processing.

2.1.3 Towards unified observations

This study measures the effects of both illegality and wellformedness within the same experimental paradigm, which allows comparing them without confounds due to experimental differences. Combining categorical legality and gradient wellformedness in one experiment cannot be done in a straightforward manner, due to practical problems. Empirical evidence for categorical phonotactics comes mainly from perceptual illusions, i.e. incongruencies between acoustic input and percepts, while evidence for gradient phonotactics comes mainly from facilitatory effects on speech processing, i.e. differences in processing speed or accuracy for faithful perception. These measurements are of a different type.

In addition, while it is not often controversial whether a phoneme combination is illegal in a language, there are different ways to calculate the wellformedness of a phoneme combination. It can be assessed on its distribution, as was done by Vitevitch & Luce (1998); it can be assessed on phonetic properties, as proposed by e.g. Fleischhacker (2001) or on phonological properties, as advocated by Albright (2009) and Moreton (2002).

This study avoids possibly controversial theoretical justifications for the wellformedness value of phoneme combinations by comparing only empirical evidence, not frequency counts or theoretical derivations.

As mentioned above, the hypothesised unification of phonotactic effects to one phonotactic status makes one important and falsifiable prediction: facilitation of wellformed combinations and perceptual illusions do not combine, because perceptual illusions go towards more wellformed, namely legal, structures, not away from them. If a perceptual illusion is observed from a certain sound combination to another, the direction is from an illegal to a legal combination. The illegal combination is predicted to not be wellformed and hence not to show a facilitatory effect.

Perceptual illusions that change legal input occur in the start state of second language (L2) acquisition. L2 learners might represent a sound combination as illegal if it is illegal in the first language (L1), due to transfer. If the combination is actually wellformed in the second language, the L2 listeners might learn that this combination is wellformed, but the unified account predicts that the combination can then not be also represented as illegal. Whether L2 listeners consider a combination illegal can be revealed by a perceptual illusion. At the
moment that L2 learners acquire the wellformedness of a combination in the
second language, which can be observed by increased facilitatory effects, the
perceptual illusion should have disappeared.

Wellformedness in the Target Language can be inferred if a difference in
facilitatory power between two phoneme combinations is observed, first by
native listeners of the Target Language to make sure there is indeed an actual
difference in wellformedness in the Target Language. If L2 learners of the
Target language show a larger facilitative effect with increasing proficiency,
they are acquiring Target Language wellformedness.

Figure 2.5 shows the transitivity that is predicted if by the unified account
of phonotactics: if there is a perceptual illusion, this is caused by a difference
in legality and illegality in the Source Language of the non-native listener,
giving empirical evidence for the representation of illegality by the listener. In
this study, phoneme combinations are grouped into Neutral and Good ones
(details follow below); these labels refer to the status in the Target Language,
where Good clusters more wellformed than Neutral ones. Figure 2.5 shows
the empirical evidence that is needed for to place a sound combination on the
wellformedness scale.

When beginning L2 learners of the Target Language are exposed to Good
and Neutral phoneme combinations that are illegal in the Source Language,
they likely start with a perceptual illusion against both Good and Neutral
combinations, that transforms both types of phoneme combinations of the
Target Language to combinations legal in the Source Language. This illegality,
revealed by the illusion, allows the conclusion that these combinations are ille-
gal. If there is acquisition of the wellformedness of these combinations in the
Target Language, the combinations move into the region of legal combinations
(as indicated by the dotted lines in Figure 2.5). In this region, gradient well-
formedness knowledge can be acquired, which can be revealed by different
facilitation for Neutral vs. Good combinations.

The most wellformed, i.e. Good, structures are more likely to be learned,
assuming that they tend to have a higher frequency, but it is also possible that
all illegal structures move towards legality together. This again means that
only empirical observations can confirm the represented phonotactic status
of phoneme combinations. Unified accounts predict that more wellformed
combinations should cross the boundary into legality before or together with
the Neutral ones, becoming contrastive with their repairs, i.e., the perceptual
illusion should disappear quicker for more wellformed combinations. As
combinations are supposed to be more wellformed if they are easier to process,
the combinations that are easier to process should thus be less susceptible to
perceptual illusions at the point in the acquisition trajectory that they become
easier to process.
The comparisons mentioned above require one set of experimental items and a common experimental methodology, because otherwise any observed difference between conditions could also be explained as a task or item effect. As speech processing ultimately tries to map input to actual words, a word recognition task is preferable over other paradigms. Perceptual illusions in word recognition would entail the absence of contrastiveness between the actual input and the perceived repair. Contrastiveness is linguistically most relevant when a contrast is used to distinguish word forms (Trubetzkoy, 1939). There is one potential problem: a perceptual illusion involving an illegal form close to a word can lead to perception of the (legal) word, but the lexical Ganong-effect explains the same illusion: ambiguous input is perceived with a bias towards the word (Ganong, 1980). The Ganong effect does not depend on illegality, so a perceptual illusion is no evidence for illegality if it moves towards words. However, in a second language, a perceptual illusion can be directed away from an actual (Target Language) word, towards a repaired version of it, assuming this word is stored faithfully in the L2 listener’s lexicon. If it is not stored faithfully, that already shows that the perceptual illusion is not a Ganong effect, as it has then preceded the existence of a lexical entry.

Perceptual illusions might be observed in priming. Word activation due to unfaithful tokens can be measured using priming, as shown by Broersma &
Cutler (2008), who used priming to show that voicing in obstruents at the end of English words is not contrasted by Dutch listeners, as English words with voiced and voiceless codas are primed both by a faithful version of those words as well as by a version in which the coda voicing is changed; Dutch has no voicing contrast in codas, as Dutch coda obstruents are always voiceless. The present study uses cross-modal priming to measure the activation of words by auditory primes. The recognition of the word is reflected in the latency of lexical decisions that the participants had to make on orthographically presented targets.

The targets are activated (or not) by the auditory primes, in which manipulations occurred. If there is no difference in priming of a word by its correct pronunciation and the expected illusionary perception of the word for L2 listeners, the existence of the illusion has been shown. The sound combinations used in this study are legal consonant clusters of Dutch, that start with an /s/. They will be referred to as /sC/ clusters. Illegality and thus potentially observable perceptual illusions were added by also presenting the clusters to L2 learners with a native language that does not allow /sC/ clusters, namely Spanish (Lloyd & Schnitzer, 1967) and Japanese (Itô & Mester, 1995).

Spanish and Japanese listeners are expected to perceive an illusionary vowel in these clusters (Polivanov, 1931; Dupoux et al., 1999; Dupoux, 2001; Jacquemot et al., 2003). This means that the cluster and its epenthesised version are not contrastive and primes for words containing /sC/ cluster should thus activate the word equally well, irrespective of the presence of an epenthetic vowel. Spanish and Japanese are expected to repair /sC/ clusters as [EsC] and [suC], respectively (Dupoux et al., 1999).

The clusters in the auditory primes were manipulated to contain these types of epenthesis. If L2 listeners do not distinguish the clusters from the epenthesised versions, both types of primes (with and without epenthesis) should activate the word equally well. For example, if Spanish listeners perceive [stat] stad ‘city’ as /Estat/, then primes of both the faithful form [stat] and the epenthesised form [Estat] should activate the word ‘stad’ to the same extent.

In addition to the detection of perceptual illusions, phonotactic wellformedness is expected to be revealed in the priming results, namely in the facilitation of word recognition, unless the lexical neighbourhood is not influential enough to obscure this effect. The wellformedness contrast used in this study is between two sets of legal Dutch /sC/ clusters. If phonotactic wellformedness facilitates speech processing, words containing more wellformed clusters should be activated more easily or strongly and thus be recognised more
easily. The clusters used are hypothesised to have different wellformedness based on corpus analysis, which will be described below.

If there is evidence for negative phonotactic knowledge in the form of illusions, as well as evidence for Target Language gradient wellformedness in the form of a facilitatory effect, there is a conflict in wellformedness between Target and Source Language. Any acquisition trajectory between Target and Source has to resolve this conflict. According to a unified theory of phonotactic wellformedness, the acquisition process has to be unified as well. Acquisition is predicted to proceed towards greater wellformedness. This prediction is testable, as it would amount to the observation of a decrease of the illusion and an increase in facilitatory effects for the most wellformed clusters. This prediction makes the unified account of phonotactic knowledge falsifiable as well, as an observation of an increase in the illusion and in the facilitatory effects should not be possible.

2.1.4 Wellformedness of Dutch /sC/ clusters

Frequency differences between sound combinations serve as a tool to detect possible gradient wellformedness differences. There are different ways to calculate the probability of a sound combination: this chapter does not attempt to test the validity of any method against other possibilities.

As the language of all experiments is Dutch, frequencies of phonemes from the Corpus of Spoken Dutch (CGN; Oostdijk, 2000) were used to calculate wellformedness scores for the /sC/ clusters. Using continuous speech, instead of lexical frequencies, makes it possible to find phonotactic gradient wellformedness patterns that diverge from the lexicon, even though both are still correlated. As lexical neighbourhood density has an effect opposed to phonotactic wellformedness (see above), this is important. As an added benefit, continuous speech is more available to L2 learners: their lexicons might be incomplete or suffer from distortions due to misperceptions. Hence, a difference that can be observed in continuous speech is more likely to be learnable, although it remains an empirical question if the wellformedness differences can be acquired and if lexical neighbourhood density does not cancel the effect.

The calculation used to select and categorise clusters is the observed over expected (O/E) ratio (Pierrehumbert, 2003), calculated over the consonant clusters in the corpus. O/E ratios are calculated by dividing the observed frequency of the combination in the corpus, O, by the estimated frequency, E. The latter is estimated from the frequency of the individual components, the /s/ and the consonant, by multiplying their probabilities with each other and the total number of tokens. In other words, if two phonemes have a high
probability of occurrence, their combination is expected to occur frequently as well. The value for E reflects how often two phonemes would be observed next to each other if there was no preference for any combination over others at all; if there indeed is no preference for a combination, O equals E, yielding an O/E ratio of one, and the frequency of phoneme combinations would be a function of the frequency of the constituent phonemes. Note that the O/E value is corrected for high or low frequencies of combinations that are only due to the high or low frequency of the components.

The O/E ratios for Dutch /sC/ clusters do not all equal one, as can be seen in Table 2.1 for all Dutch /sC/ clusters that occur word-initially in more than a handful of words. If the O/E ratio of a combination of two phonemes is lower than one, this means that there is a tendency not to combine these phonemes. Apart from /sk/, Dutch has more legal /sC/ clusters with a O/E ratio below one, which would fit the purposes of the present study, but there are too few words starting with these clusters.

Two groups of (legal) /sC/ onset clusters were defined; these groups form the two levels of the Phonotactics condition. Good clusters have a higher frequency than expected, while Neutral clusters occur roughly as often as if they were randomly joined.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Type</th>
<th>Observed</th>
<th>Expected</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>[sp]</td>
<td>good</td>
<td>5573</td>
<td>1783</td>
<td>3.12</td>
</tr>
<tr>
<td>[sx]</td>
<td>good</td>
<td>10762</td>
<td>3596</td>
<td>2.99</td>
</tr>
<tr>
<td>[st]</td>
<td>good</td>
<td>32173</td>
<td>15840</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>48508</td>
<td>21219</td>
<td>2.28</td>
</tr>
<tr>
<td>[sm]</td>
<td>neutral</td>
<td>3548</td>
<td>3541</td>
<td>1.00</td>
</tr>
<tr>
<td>[sl]</td>
<td>neutral</td>
<td>3667</td>
<td>3304</td>
<td>1.11</td>
</tr>
<tr>
<td>[sn]</td>
<td>neutral</td>
<td>3742</td>
<td>3136</td>
<td>1.19</td>
</tr>
<tr>
<td>[sk]</td>
<td>neutral</td>
<td>2703</td>
<td>3241</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>13660</td>
<td>13222</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Table 2.1: Ranking of /sC/ clusters on O/E. O/E (Observed over Expected) ratios are calculated using frequencies of the phonetically annotated part of the Corpus of Spoken Dutch (CGN; Oostdijk, 2000). The expected values are based on the frequencies of the phonemes in consonant biphones, without regard to word or syllable boundaries; using the frequency over the whole corpus gives even more pronounced differences.
Note that the use of O/E ratios on continuous speech to detect wellformedness is not trivial; actual evidence for wellformedness has to come from the experiment. However, ranking the clusters on other statistical measures of co-occurrence will not greatly differ. Other ranking methods have the disadvantage that a principled division into good and neutral (and bad, missing here) is not available. For example, transitional probabilities (TP) are unidirectional and influence each other; in the present case, there are so many /st/ tokens that the TP of other clusters has to be low, but that is mainly because /t/ is very frequent after /s/ and does not reflect that /sC/ onsets in general occur quite frequently.

2.1.5 Acquisition and proficiency

By measuring the participants’ proficiency in Dutch, the Target language, the acquisition trajectory could be studied. If the predicted effects of the wellformedness differences between Good and Neutral clusters and the illegality difference of /sC/ clusters between Dutch and the Source languages are indeed observed, the acquisition trajectory has to conform to the unified theory in the way described above; otherwise, the unified theory is falsified.

As the experiment itself employs auditory speech recognition, the proficiency of the participants was assessed in a different way. C-tests provide a score that is a relatively good indicator of general language proficiency (Eckes & Grotjahn, 2006). As the experiment is not longitudinal, it is not possible to map the acquisition trajectory for each participant, which could indicate if phonotactic illusions disappear before facilitation by wellformed clusters appears. However, by calculating the predictive power of proficiency on both factors, it is possible to see if the trajectories have the predicted direction.

2.2 Experiment 1

2.2.1 Method

Participants

The native speakers of Dutch were recruited through the Utrecht Institute of Linguistics OTS subject pool; in total 71 participants completed it. The Spanish group consisted of 36 participants with Spanish as their native language and no other native language that allows /sC/ clusters (other languages were Quechua, Catalan, Galician and Valencian). The Japanese group consisted of 33 participants. The non-native speakers were found in various ways, using
flyers, networks of previous participants and through institutions that teach Dutch as a foreign language. All non-native speakers lived in the Netherlands.

All participants reported not having problems hearing stimuli over headphones, nor reading from a computer screen. Participants were paid 7.50 euros for their participation, which lasted up to one hour.

**Materials**

Items are constructed based on a target word. There were 60 experimental items, of which the target was a word starting with /sC/. For filler items, 60 words not starting with /sC/ and 120 non-words, which followed Dutch phonotactics, were used. Half (60) of the filler non-words started with /sC/ and the other 60 filler non-words did not start with /sC/, preventing a response bias based on the presence of /sC/ in the target.

The words starting with /sC/ clusters were nested in the Phonotactic condition. 30 words started with Good clusters, 30 with Neutral clusters. Words were monosyllabic or bisyllabic. All words were singular nouns or infinitive verbs, selected from the most frequent words starting with sC clusters. These most frequent words were determined using an equally weighed average of the frequency in the CGN Corpus of Spoken Dutch (Oostdijk, 2000) and the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). The weighing corrected for the different sizes of the databases, giving both the same influence. As the most frequent 30 words with Neutral clusters were less frequent on average than those with Good clusters, the words for the Good level of Phonotactics were picked to match the frequency of the most frequent Neutral cluster words one-by-one, as closely as possible. The result of this strategy is that the distribution of the frequencies of the Good words is similar to the distribution of the frequencies of the Neutral words.

For each target, there were three auditory primes, such that trials consisted of prime-target pairs in three Priming levels: Unrelated, Epenthesised or Faithful. In the first case, the prime was a word that was not related to the target. In the last case, the prime was a correct pronunciation of the target word. The Epenthesised condition allows a test of the perceptual illusion of L2 listeners. The epenthesis, which is supposed to correspond to their perceptual illusion, should not have been contrastive, so that the priming strength of Epenthesised and Faithful primes should be similar. If there is a difference, the epenthesis is noticed, meaning that the clusters are contrastive with their epenthesised counterparts.

The epenthesis used was based on the Source Language of the L2 listeners, namely Japanese and Spanish. For Japanese, a cluster starting with /s/ is assumed to be transformed by adding an /u/ after the /s/. The Dutch /u/
differs phonetically from its Japanese counterpart, that can be produced as voiceless or non-rounded. Nevertheless, the Japanese vowel system does not contain a contrast between anything like a Dutch /u/ and a Japanese /u/. Hence, a Dutch /u/ was used.

For Spanish, a short vowel /e/ is added before the cluster. As Dutch does not have a short /e/, but a long /e:/ and a short /e/, the latter was used; it should not be contrastive with Spanish /e/ as Spanish only has the five vowels /a/, /i/, /e/, /o/ and /u/.

The Unrelated and Faithful primes were recorded by a native speaker of Dutch who was unaware of the purpose of the study. Epenthesised primes were constructed using the Faithful recordings of the targets: the /s/ of these recordings was deleted and the remainder spliced after a fragment of /es/ or /su/, carved out of other Dutch words pronounced by the speaker. The /es/ or /su/ were cross-spliced in such a way that no unnatural transition was present. This was done using PRAAT software (Boersma, 2001), combining two fragments at zero-crossings of the amplitude wave, so that amplitude would rise before and after the boundary, or fall before and after the boundary. For the faithful clusters, the same splicing was performed, now pasting a production of /s/ excised from another word, so that these primes as much as possible matched the manipulated primes in possible unnaturalness at the splice point. Three native speakers of Dutch, all linguists, confirmed the items sounded natural to them. The duration of the spliced /e/ was close to 100 ms; the /u/ was close to 80 ms. As the items were spliced at the most natural points, the duration differed slightly between items, but not more than 5 ms.

As the targets were presented in orthographic form, orthotactic influences have to be taken into account. These might be similar to the effects of phonotactics (Bailey & Hahn, 2001). However, the most frequent combinations in speech are expected to also be the most frequent in written language, as Dutch orthography closely follows the phonemic structure for all clusters. The only exception is the Good cluster /sx/, which is written with three graphemes (‘sch’). The letters ‘ch’ are most often pronounced as (close to) [x], but can also occur in loan words, where they are pronounced as /ʃ/. On the other hand, the sound /x/ is usually written as ‘g’, hence ‘sch’ would have a lower Expected value and hence a higher O/E value than /x/ if ‘ch’ is taken as one orthographic unit. The orthotactic frequencies thus differ from the phonotactics for this cluster, but this only enhances the contrast between Good and Neutral. In addition, the baseline of Unrelated primes for every word serves to incorporate word-specific effects, including orthotactic effects.

The filler targets, both non-word and word, were matched to primes in such a way that the phonological (starting with /sC/ or not, epenthesised
or not) and lexical properties (word or not) of the prime did not predict the correct answer.

Procedure

The test was administered in a sound isolated cabin at the Utrecht Institute of Linguistics OTS phonetics laboratory, or in a quiet room at the participants’ home or school, using a computer running under Ubuntu LINUX 6 with Xenomai real-time support. Participants were (individually) seated by the experimenter and received instructions according to a protocol. They were told that they were to hear Dutch words or non-words and also see Dutch words or non-words on the screen. They were instructed to determine, as quickly as possible, whether the target presented on the screen was an existing word of Dutch, while trying not to make too many errors. They responded by pressing the ‘yes’ or the ‘no’ button on a button box. The yes button was placed the side of the dominant hand; buttons were marked with the Dutch words ‘ja’ and ‘nee’, for ‘yes’ and ‘no’, respectively. Participants were told they would occasionally be asked to write down the last word they heard (this happened every 60 trials), in order to avoid them stopping listening to the auditory primes. This also worked to check if they were listening, but virtually no errors were made (apart from spelling oddities, mainly in non-words). Participants could take a break every time they had to write down a word, as the experiment paused there until a key was pressed. No information was given about the nature of the words used in the experiment. Three practice trials were presented, if necessary repeatedly until participants understood the procedure.

Auditory primes were presented over Beyerdynamic DT250/80 headphones at a sound level comfortable to the participant; visual targets on a 15” or 17” computer screen. After the auditory prime was played, there was an 500 ms interval, after which a fixation point (+) was shown for 500 ms in a white background to announce the appearance of the target. The target was shown for 750 ms; participants had 2500 ms to respond from the start of the target presentation. The words were shown in lower case black letters on a white background. After every trial, feedback was presented on screen (‘correct’, ‘wrong’, or ‘too late’ in Dutch).

All targets were presented once, in one of the three Priming levels; the target item were rotated through Priming in a Latin Square design. The order of the trials was randomised separately for each participant. Epenthesised primes were either epenthesised to /esC/ (the Spanish version) or with /suC/ (the Japanese version). In order not to confuse the language of the type of epenthesis with the native language of the participants, the combination of native language and the epenthesis version are referred to as levels of one
covering condition called Group. The Dutch native listeners received either the Spanish or Japanese epenthesised primes; these two groups are called Native-Es (37 participants) and Native-Su (34 participants). Participants with Dutch as a second language were given the epenthesis that is normally found in their own native language. These Groups are therefore simply called Spanish and Japanese.

After the cross-modal priming part of the experiment, participants filled in a questionnaire about their language backgrounds and did a C-test. The C-test consisted of five Dutch texts, ranging from informal to formal. Participants had to complete twenty words in each texts, of which the last half was removed. Their proficiency was expressed as the number of completely correct answers they gave. Only words that were orthographically correct and morphologically and semantically possible in the given context were scored as correct. The C-tests were scored by by one of two experimenters, who discussed answers they had doubts about.

**Analysis**

Accuracy scores of more than 134 good answers out of 240 are less than 5% likely for participants answering randomly, according to the binomial probability distribution. Therefore participants scoring less than 134 correct answers were excluded; they could have responded randomly or else they knew too few Dutch words to be sure that their correct answers were not accidental. For the Japanese group, 30 participants were left, for the Spanish group, 35. No native Dutch speakers were removed. Due to an implementation error, one item from the Good level of Phonotactics had to be discarded.

The dependent variable was the logarithmic transformation of the reaction time in ms. The transformation is applied to remove the usual skewness of reaction times. Only trials in which a word was correctly recognised were used. Responses were defined as outliers and removed when they were more than three times the interquartile range away from the quartiles for each participant. There turned out to be only outliers that were very fast (below 50 ms), not very slow ones, but answers slower than 2500 ms were not possible due to the time-out. Outlier removal removed 0.7% of the observations in the Japanese Group, 1.4% for the Spanish Group, 0.7% for the Native-Es group and 0.6% for the Native-Su group.

The sphericity assumption necessary to use repeated measures ANOVA is not to be expected to be correct in groups differing in L1 and L2. To avoid this incorrect assumption and allow to correct for noise more strictly, the data were analysed in a mixed-effects linear model with crossed random effects for participants and items, as advised by Quené & van den Bergh (2008) and Baayen, Davidson, & Bates (2008).
The contrasts in the Group condition were planned contrasts, between the Native-Es and Native-Su Group, the Spanish and Native-Es group and the Japanese and Native-Su Group, as comparing a different L1 (Japanese and Spanish) and a different type of epenthesis is not informative for the purpose of this study. The C-test score was used as a predictor to assess possible learning effects on illegality and wellformedness. Additional factors added to the model were length of stay in the Netherlands, previous reaction time, correctness of previous answer, trial number and word frequencies from the CGN Corpus of Spoken Dutch and CELEX. These factors were included to see if they improve the model fit; the better the model fits the data, the more reliable the estimates for all conditions will be, barring overfitting.

To find the optimal model of the data, the most complex model was fitted to the data using the ‘lmer’ function of the ‘lme4’ package of the R statistical software (R Development Core Team, 2010; Bates, Maechler, & Bolker, 2011). To simplify the model, terms with small influences were removed. When the model fit was not significantly lower after this removal, the removal was warranted. The model fit differences were evaluated with the ‘anova’ procedure of lme4. This strategy leads to the simplest model that still optimally predicts the dependent variable (Baayen et al., 2008). The only factor that was completely removed was the CELEX frequency of the words; but as the CGN frequency was kept as a factor, frequency information is taken into account. Other factors were not completely removed, but some of the interactions of these factors were. The random slope of item on Priming was also not needed for the model fit.

The factors were checked for collinearity where this was expected to occur; residualisation was used to remove the collinearity. This does not affect other factors in the model, nor the model fit.

There was collinearity between C-test score and Group, which was to be expected as Dutch participants are bound to have a higher proficiency in Dutch than L2 participants. The score on the C-test was residualised to group. This means that Group was used to predict the C-test score and that only the divergence from that predictor was used, allowing to assess the influence of proficiency after the influence of Group has been modelled separately. In practice, for every participant the C-test score was redefined as the difference between the actual score and the group mean. The same correction is applied to the number of years spent in the Netherlands, which should not be collinear with Group either, to avoid dividing the effect of native language over two factors.

The CGN corpus frequency was residualised to Phonotactic condition, allowing the CGN frequency to independently account for any frequency effect without disturbing the estimate for the Phonotactic condition.
As the random effects of the model include complex terms, it was not possible to perform Monte Carlo simulations as recommended by Baayen et al. (2008). Instead significance is calculated using the t-values and a conservative estimate of the degrees of freedom, namely 58, using the lowest possible number among the random effect terms (items) minus 1. Confidence intervals were calculated using the Monte Carlo simulation method from Baayen et al., on a model that was similar, except for the removal of the complex random effect terms.

To make sure no speed/accuracy trade-off explains our results, accuracy data were modelled in a mixed model similar to the model for the latency, but now with the logit of the correctness as the dependent variable (T. F. Jaeger, 2008). A speed/accuracy trade-off is possible when responses are both faster and less accurate, or slower and more accurate, in a certain condition. The accuracy data was not used for hypothesis testing, as the latency data was already used for this purpose.

### 2.2.2 Results

The estimates for the factors that are independent variables are shown in Table 2.2. The accuracy model, shown in Table 2.3, shows no significant effect in the same positive or negative direction as the corresponding latency effect, showing that faster processing is not accompanied by significantly less accurate processing, or vice versa.

Figure 2.6 shows the latencies for Dutch native listeners. Differences between the native and other groups are discussed below, but overall, Epenthesis primes and Faithful primes both prime the targets, as witnessed by significantly faster responses compared to the baseline latency for Unrelated Primes. There is a trend towards a faster response for Good items, but for Epenthesis primes to Good items, there is a contrary effect.

Group differences can be seen in Figure 2.7. L2 learners answered most slowly. Comparing the two Native Groups (Native-Es vs. Native-Su), epenthesis of the /su/C/ kind (anaptyxis) leads to slower responses than the prothesis (/esC/) kind. Note that the participants of these groups were randomly assigned from the same population of Dutch native speakers.

The Spanish group are primed more by Epenthesis primes than Native speakers, ending up at a level of priming approximately as high as the priming by Faithful primes. The Japanese group did not differ from the Dutch group, apart from overall slower performance.

The effects of proficiency on the priming of the Epenthesis primes (indicative of a perceptual illusion) and on of the Phonotactics condition (indicative of the wellformedness facilitatory effects) show in the interactions with
Table 2.2: Experiment 1: latency model. Factor estimates. Interactions between independent variables are not shown when they were not near significance, i.e. they had a p-value above 0.10, with the exception of Epenthesis: Native-Su → Japanese. Nuisance variables are all left out of this table. Significance was assessed using the t-value and assuming 58 d.f. (the number of items -1). Arrows (→) indicate a contrast between two groups in the given direction; intersection signs (∩) mean the intersection of two levels of different conditions, i.e. all the trials that conform to both parts of the intersection. The formula in R used to obtain this model was LogRT ∼ 1 + Phonotactics × Priming × Group × C-test score + years in Netherlands + log(previous RT) + correctness previous trial + trial nr + CGN frequency + (1 + Phonotactics + Priming | part) + (1 | item). The logged likelihood was -763.6. For the 95% Confidence Intervals, the mcmcsamp and HPDinterval functions from the lme4 package were used, as described by Baayen et al. (2008). As these functions are not yet implemented for random slopes, these were dropped, i.e. (1 + Phonotactics + Priming | part) was changed to (1 | part); the estimates were only mildly affected by this change.
Table 2.3: Experiment 1: accuracy model. Factor estimates. Experimental predictors that are not shown were not near significance, i.e. they had a p-value above 0.10. Nuisance variables are all left out of the table, but were left in the model when they improved the model significantly. Arrows (→) mean a contrast between two groups in the given direction; intersection signs ∩ mean the intersection of two levels of different conditions, i.e. all the trials that conform to both parts of the intersection. The formula in R used to obtain this model was Correct ~ 1 + Phonotactics × PrimingU × Group × CgoodRC + yearsInNlC + logprevRT + prevWrong + trialNrC + (1 + Phonotactics + PrimingU | part) + (1 | item). The logged likelihood was -1413.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.639</td>
<td>0.186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>-0.556</td>
<td>0.019</td>
<td>-2.15</td>
<td>0.032 *</td>
</tr>
<tr>
<td>Phonotactics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priming</td>
<td>0.868</td>
<td>0.207</td>
<td>4.18</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Faithful</td>
<td>1.693</td>
<td>0.346</td>
<td>4.90</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native-Es → Spanish</td>
<td>-0.529</td>
<td>0.215</td>
<td>-2.46</td>
<td>0.014 *</td>
</tr>
<tr>
<td>Native-Su → Japanese</td>
<td>0.095</td>
<td>0.271</td>
<td>0.411</td>
<td>0.681 n.s.</td>
</tr>
<tr>
<td>Interaction Group × Priming</td>
<td>-1.243</td>
<td>0.422</td>
<td>-2.95</td>
<td>0.003 **</td>
</tr>
</tbody>
</table>
the C-test score. The higher proficiency is, the faster participants are, but this effect is not significant in general and only a trend for Faithfully primed items. For the Spanish group, there is no evidence that the priming by Epenthesised primes is less when proficiency in Dutch is higher. However, there is one learning effect in the opposite direction: with growing proficiency, Spanish listeners are primed more by the wellformedness of Good but Epenthesised clusters. Native listeners do not show this effect. Figures 2.8 and 2.9 show the correlation between latency and proficiency for the Good \& Epenthesis trials; the interaction factor Good \& Epenthesis: Native-Es→Spanish×C-test in the model gives the difference of the slope of the lines shown in the two graphs. Note that the data is not balanced; some proficiency scores are shared by more than one participant.

2.2.3 Discussion

The results of Experiment 1 are valuable, as they show that the experiment succeeded in finding a perceptual illusion, indicative of illegality, and a facilitatory effect, indicative of gradient wellformedness differences between two sets of legal sound combinations of Dutch.
The comparison between the Japanese and the Native-Su group shows that Japanese listeners do not have perceptual illusions for the /sC/ clusters, at least not illusions that sound identical to the [suC] type of epenthesis. The comparisons between the Spanish and Native-Es groups are highly relevant, as well as the interactions with proficiency, indicative of the acquisition trajectory.

Before drawing general conclusions about the differences between the L2 and native listener groups, it is necessary to assure that the difficulty to distinguish Epenthesised primes from Faithful primes is indeed the result of the transfer of illegality of the /sC/ clusters. It is also possible that second language learners are in general more tolerant of the match between acoustic input and lexical representation. If this had been the case in Experiment 1, the same effect should be found with participants with Dutch as a second language, but a first language that does not have the specific epenthesis of Japanese or Spanish. Experiment 2 serves this purpose. The results from the Native groups of Experiment 1 are used to compare to the results of the L2 listeners in Experiment 2.
Figure 2.8: Experiment 1: latencies for Epenthesised primes, Spanish group, as a function of proficiency as measured by C-test score.
Figure 2.9: Experiment 1: latencies for Epenthesised primes, Native-Es group, as a function of proficiency as measured by C-test score.
2.3 EXPERIMENT 2

To see if any of the effects found for participants with L1 Spanish or Japanese are not just due to the fact that L2 speech processing is less accurate in general, but to the actual L1’s, the experiment was repeated with participants that were also second language learners of Dutch, but had other native languages.¹

2.3.1 Method

Participants

A control group of 31 second language learners of Dutch was tested; 15 were given the version with /tvC/ epenthesis, 16 with /suC/ epenthesis. None of the participants had a native language with a ban on /sC/ or /CC/ clusters similar to Spanish or Japanese. The participants were recruited through flyers and word of mouth. All lived in the Netherlands.

Materials and procedure

Materials and procedure were equal to Experiment 1. Participants were randomly assigned to the Japanese or Spanish version of the experiment; these versions only differ in the type of epenthesis of the Epenthesis primes, as explained above.

Analysis

The data of two participants in each of both versions had to be omitted, because these participants’ accuracy was not significantly above chance level. The same item that was omitted in Experiment 1 was omitted for the analysis of Experiment 2 as well.

The same model as used in Experiment 1 was fitted to the control group of learners of Dutch with another L1, compared to the same Dutch native speakers also included in the analysis of Experiment 1, who serve as a baseline for the comparisons.

2.3.2 Results

The estimated values of the model are shown in Table 2.4.

The non-native listeners in this experiment responded slower than the native listeners overall, similar to the non-native listeners in Experiment 1. There was a trend that non-native listeners responded faster to Good items,

¹ Thanks to Jacques Mehler for pointing out the need for this additional experiment.
Table 2.4: Experiment 2: latency model. Factor estimates. The predictors that were significant in Experiment 1 are shown here, as well as those predictors with a \( p \)-value below 0.10 in Experiment 2. The model specifications were the same as those used for the model of the results of Experiment 1. The logged likelihood was -331.6.
compared to native listeners, in the /su/-version of the experiment. Otherwise, no significant differences between native and non-native listeners were found, unlike in Experiment 1.

The effect of proficiency is similar for Native and Other-L1 participants: for Faithful items, there is a significant decrease in latency with growing proficiency, while for Good items, there is a trend in the same direction.

2.3.3 Discussion

As this experiment is not aimed at rejecting a null hypothesis, the p-values should not be taken as strictly as normal. However, the estimated effects of Group are considerably smaller than in Experiment 1, with the exception of the main effects. This suggests that while second language learners are in general slower when processing speech in the Target language than native listeners of the same language, but that the epenthesis introduced to elicit a perceptual illusion for second language learners with Spanish as a second language is indeed less contrastive to this group than to second language learners in general.

The priming provided by epenthesised primes for non-native listeners in the OtherL1-Es and OtherL1-Su groups do not seem to be different from the Native-Es and Native-Su groups. This failure to reach significance cannot be explained by the slightly higher standard error in Experiment 2, because if the standard error was as small as in Experiment 1, the effect is still not significant ($t = -1.06, p = 0.226$). The difference between Native-Es and OtherL1-Es is almost four times as small as between Native-Es and Spanish.

The finding that the C-test score was related to the latency for Faithful and for Good items shows that this measure of proficiency can predict the increase in the employment of phonological knowledge of Dutch in the processing of Dutch speech. Most likely, an increase in C-test score comes with the acquisition of the phonotactic knowledge that allows to process normal (i.e. identity) priming, especially for wellformed sound combinations.

2.4 General Discussion

The experiments in this chapter were aimed at combining an effect that is commonly attributed to phonotactic illegality to an effect commonly attributed to gradient wellformedness. These two effects have been found. Priming, and therefore word activation, was found for epenthesised versions of words, hence for not completely faithful pronunciations. The effect for Epenthesised primes was smaller than the effect for Faithful primes (note that the significance of this difference was not tested). Epenthesised primes do still contain a large
part of the target words, but they differ crucially at the onset, which makes it still surprising to find a priming effect for these primes at all.

The facilitatory effect from phonotactic wellformedness was not as clear as desired; it reached only near-significance. There is a simple explanation for this lack of overall facilitation: the overall effect of Phonotactics is calculated over the three Prime types, but on the level of Epenthesis, items with Good phonotactics are not recognised much faster than Neutral items. This inhibitory effect of the interaction is significant, while the facilitatory main effect is only near significant. One can conclude that the facilitation provided by the wellformedness of the cluster is lost when the cluster is broken by epenthesis.

The effects of Priming and Phonotactics thus have to be considered in interaction. The results can be considered to be in accordance with the Vitevitch & Luce (1999) proposition that listeners have sublexical representations of wellformed combinations that cause the facilitatory effect on speech processing. If this is the case, there have to be representations of the Good /sC/ clusters, that boost the activation of the word containing these clusters. The adverse lexical neighbourhood is overcome by the strong phonotactic wellformedness, confirming the finding that phonotactics also facilitates on a lexical level (Luce & Large, 2001), but while Luce & Large controlled for lexical neighbourhood, this experiment did not: the Good cluster words come from larger lexical neighbourhoods. This means that phonotactic effects still had to beat lexical effects. Sublexical representations can thus be more powerful than lexical effects. If this is correct, the size of the phonotactic facilitation is larger than the effect size in this experiment suggests. Note that when the prime is Unrelated, targets with Good clusters are still recognised quicker, which might indicate an additional orthotactic effect. Nevertheless, the interaction with epenthesis in the prime, a purely auditory manipulation, means that there is at least a phonological component to the effect.

The trend of an effect of the Phonotactics condition and its interaction with the Priming condition can best be explained as a facilitatory effect of phonotactic wellformedness, as it is positive for Good items, though not significant. Note that the null hypothesis could have been that responses to Good items are slower, as predicted if the denser lexical neighbourhood would cause a strong inhibition through competition between similar words. Such a null hypothesis would have to be rejected, as it is one-tailed, warranting dividing the p-value by 2. However, this study did not start with a prediction on the direction of the effect, so the reported p-value is based on a two-tailed test against the effect being equal to zero.

Following the Vitevitch & Luce (1999) idea of representations of wellformed phoneme combinations, it is possible that the distortion of Good clusters by
epenthesis does not allow the activation of the sublexical representations. According to Vitevitch & Luce, detection of the wellformed combinations in the input activates their sublexical representation, which in turn activates the words containing the combination, facilitating the perception of those words. If epenthesised clusters in the input do not allow the activation of the cluster representations, the facilitatory link is not established and the facilitation does not occur. The conclusion has to be that the clusters are apparently not perceived as clusters if they are epenthesised, so that their representation is left unactivated and the facilitatory effect cancelled. A syllable boundary forced by the epenthetic vowel is a possible reason for the failure to perceive epenthesised clusters as clusters; it is also possible that the cluster is represented with reference to its onset position (/#sC/ ≠ /csC/) and that the epenthesised version thus does not match the representation. In the case of the Neutral clusters, there was probably no facilitation, as these clusters are not expected to be represented separately. For both Good and Neutral Epenthesised primes, there is activation of the word through the phonemes (/s/ and /C/), under the assumption that these phonemes are activated by the input irregardless of epenthesis, as epenthesis does not remove the consonants. Only for the Good primes, there is possible facilitation caused by the wellformedness of the cluster, which can be lost when the cluster is epenthesised.

The Native-Es and Native-Su groups differ in their latency for the Epenthesised primes. As these groups consist of similar participants, but different types of epenthesis, the difference has to be explained by the epenthesis. The Spanish type of epenthesis, with [esC], leaves the cluster intact on the surface, while the latter, with [suC] is more intrusive. It is possible that there is also an effect of other factors, for instance phonetic factors related to the difference in vowels. One suggestion is provided by Fleischhacker (2001), who concludes that intrusive or anaptyctic epenthesis of the [suC] type is more different from the original cluster than the prothetic [esC] epentheses. Fleischhacker based this conclusion on cross-linguistic and experimental data. Anaptyxis is most damaging to sibilant-stop clusters (like /st/, /sp/, /sk/), followed by sibilant-nasal clusters (/sm, sn/, then the sibilant-liquid ones (/sl/), because the perceived difference to the faithful version is larger compared to the same epenthesis in other consonant clusters. This explains why languages might not resolve /sC/ clusters completely with anaptyxis, but instead use prothesis to remain a bit closer to the original cluster, even if that means that consonants are not always prevocalic.

Comparing the results for the Spanish group and the Native-Es group reveals a smaller contrast between Epenthesised prime and Faithful prime. This effect can be attributed to the illegality of the /sC/ clusters in the Source
Language. Most likely, the L1 illegality caused a perceptual illusion in the L2. In Experiment 2, such an effect was not found for non-native listeners with not Spanish but another language as first language.

Unfortunately, the Japanese group did not specifically differ from the Native-Su group in their latency for the epenthised primes: there is no evidence for perceptual illusions in the Japanese group. As primes with the /suC/ epenthesis are also less effective primes for the targets for native listeners, this manipulation can indeed be assumed to be more distinct from the normal pronunciation.

It is possible that the Dutch pronunciation of the vowel /u/ did not sufficiently match the epenthetic vowel that the Japanese listeners perceive, causing a mismatch between the perceptual illusion of the Faithful prime and the epenthised prime. Japanese listeners have specific epenthetic vowels and in word perception other vowels are perceived as different (Dupoux et al., 2001).

The failure to find a difference for the Epenthesised primes between the Japanese and Native-Su group is surprising given the previous findings by Dupoux et al. (1999), who reported that Japanese participant hardly discriminate between consonant clusters and epenthesised versions of those clusters. However, their study focussed on non-lexical processing and their participants were not aware of the relevance of consonant clusters or epenthesis for the task. In the present experiments, the task was to recognise Dutch words, which is a task the participants face in their everyday life in the Netherlands. They also probably received some training in Dutch (or English) and they are therefore far more likely to be aware of their unfaithful perception.

Another relevant earlier finding is one by Davidson et al. (2007). They found that the contrast between a non-native consonant cluster (CC) and an epenthesised version of that cluster of the form CsC is easier to perceive after a minimal pair is given. In comparison to Spanish, the Japanese illegality is more across-the-board: Japanese has no consonant clusters at all, while Spanish has quite a few, just not of the /sC/ type. In addition, the CELEX database for Dutch contains no word pairs that lose their contrast with /vsC/ epenthesis. English has in /wstet/-*/stet/, ‘estate’ – ‘state’. Note that the English pair does not feature the vowel /e/, but /i/ (possibly reduced to a schwa); it is still a relevant pair, because Spanish listeners might map remove that contrast if they map English vowels onto their own system, consisting of only five vowels. For Dutch, no words in CELEX that start with a vowel followed by an sC cluster are also words when the first vowel is removed, with the exception of the uncommon word /asla/ asla ‘ash pan’. Although the stress is placed on the first syllable in standard Dutch, speakers from the North might pronounce it with the stress at the end, which gives the pair /as’la/ – /sla/, asla ‘ash pan’ – sla ‘lettuce’. This pair is so marginal that it is
unlikely to have any influence on the acquisition of Dutch by second language learners.

Japanese listeners are more likely to have encountered minimal pairs, like /ku’lAnt/–/’klAnt/, ‘lenient’–‘customer’; in addition, there are even a handful of contrasts based on /sC/–/suC/, e.g. /su’pera/–/’spera/, ‘to have supper’–‘spears’, but these never involve frequent words. In addition, Fleischhacker (2001) calls the anaptyxis into sibilant-stop clusters in Japanese loan-word adaptation a forced choice, due to the strict, highly ranked constraint that no consonants can occur in non-prevocalic position (except in case of consonant doubling or nasals). As soon as the Japanese listeners realise that consonants can occur in other positions, as in consonant clusters, the /suC/ epenthesis is more notably different from the /sC/ clusters than the prothetic /EsC/ epenthesis.

Apart from the possibility that Japanese listeners had more opportunities to notice their misperceptions or were provided with more input to notice the existence of the non-native contrast, the failure to find a perceptual illusion for the Japanese group might be due to a different type of strategy in the task. They might have focussed more on the visual words, without making use of the auditory primes, as suggested by the accuracy data (see Table 2.3). There is a surprising lack of difference in accuracy between the Japanese group and the Native-Su group, but they show lower accuracy for Faithful priming. This might indicate that they benefitted less from the primes in general and focussed more on the actual task of visual lexical decision. This also explains why they were the group with the highest overall latency.

The failure of finding a perceptual illusion in the Japanese group means that this group cannot be used to test predictions of the unification of categorial illegality and gradient wellformedness. Only the Spanish group can serve this purpose. Due to this loss of data, care must be taken in the interpretation of the remainder of the results.

The results of Experiment 2 indicate that listeners with native languages other than Japanese and Spanish, who should not have transferred illegality of sC-clusters or the type of epenthesis present in Spanish, do distinguish between epenthesised and faithful primes on a near-native level. The perceptual illusion found in the Spanish group cannot easily be explained with the Vitevitch & Luce (1999) account of representations for wellformed clusters; neither illegal nor not-so wellformed (neutral) phoneme combinations would be represented separately, but perceptual illusions only occur for illegal clusters. If the Vitevitch & Luce account is to be used to explain the Phonotactics×Priming interaction, another type of representation has to be responsible for the perceptual illusion. This goes against the unification of phonotactic illegality and wellformedness.
Interestingly, when there are perceptual illusions, i.e. for the Spanish learners, higher proficiency correlates with faster responses to good cluster words with epenthesised primes. This effect follows general effects for neutral cluster words and for the Native-Es group, that were not significant; these effects might be not related to more activation of the words by the primes, but indicate better performance with growing proficiency in general. However, the increase in speed for Spanish listeners is larger for Good clusters than for Neutral clusters, an effect that was not found in the Native-Es group. It thus might be the case that Spanish listeners acquire wellformedness for the wrong clusters, namely epenthesised ones. In other words, they incorrectly acquire that e.g. /est/ is wellformed, instead of /st/.

To draw conclusions based on the acquisition trajectory of the Spanish learners first requires a closer look at possible transfer. Although the existence of a proficiency effect for Good clusters suggest that Spanish L1 learners of Dutch do not transfer knowledge about the gradient wellformedness of /sC/ clusters in Spanish, nor use universal wellformedness, it is important to assess whether the clusters might have similar wellformedness in Spanish. O/E ratios of the /sC/ clusters were calculated for continuous Spanish using the CORLEC corpus (Moreno Sandoval, 2003). Table 2.5 shows Spanish O/E ratios for the different clusters. The differences are far less pronounced in Spanish than in Dutch, providing another indication that Spanish learners indeed are acquiring knowledge on the gradient wellformedness differences in Dutch.

If Spanish listeners would indeed correctly acquire wellformedness for some of the sC-clusters, these cluster should move up on the wellformedness scale and start being legal. However, the illegality of sC-clusters still causes perceptual illusions for epenthesised primes, while wellformedness plays a facilitating role for more proficient learners in exactly these primes. In other words, for more proficient listeners, wellformed epenthesised clusters facilitate word activation more than neutral clusters.

Note that epenthesised clusters still contain the actual clusters for the Spanish and Native-Es groups. In Dutch, the clusters are divided by a syllable boundary, because /e/ is not a legal Dutch syllable. The Dutch syllabification of e.g. [e:stat] is /e:st.at/, making it harder to extract the cluster or the word /stat/ 'city'. It is possible that Spanish L1 learners of Dutch syllabify differently, allowing them to access the cluster representations and ignoring the epenthesised vowel. This option entails growing facilitation for Faithful primes as well, which does not show. It is also unlikely that Spanish allows syllables but not words starting with sC-clusters.

The fact that epenthesised clusters still contain clusters is also problematic for the question how learners acquire the wellformedness of the epenthesised Good clusters, but not of Faithful, unepenthesised, Good clusters. Assuming
Table 2.5: Ranking of sC-clusters on O/E in Spanish

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Type</th>
<th>O</th>
<th>E</th>
<th>Spanish O/E</th>
<th>Dutch O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>[sp]</td>
<td>good</td>
<td>2645</td>
<td>1602</td>
<td>1.65</td>
<td>3.12</td>
</tr>
<tr>
<td>[sx]</td>
<td>good</td>
<td>158</td>
<td>173</td>
<td>0.91</td>
<td>2.99</td>
</tr>
<tr>
<td>[st]</td>
<td>good</td>
<td>8529</td>
<td>4732</td>
<td>1.80</td>
<td>2.03</td>
</tr>
<tr>
<td>Total</td>
<td>good</td>
<td>11332</td>
<td>6507</td>
<td>1.74</td>
<td>2.28</td>
</tr>
<tr>
<td>[sm]</td>
<td>neutral</td>
<td>1616</td>
<td>1095</td>
<td>1.48</td>
<td>1.00</td>
</tr>
<tr>
<td>[sl]</td>
<td>neutral</td>
<td>1197</td>
<td>1703</td>
<td>0.70</td>
<td>1.11</td>
</tr>
<tr>
<td>[sn]</td>
<td>neutral</td>
<td>731</td>
<td>504</td>
<td>1.45</td>
<td>1.19</td>
</tr>
<tr>
<td>[sk]</td>
<td>neutral</td>
<td>4434</td>
<td>2473</td>
<td>1.79</td>
<td>0.83</td>
</tr>
<tr>
<td>Total</td>
<td>neutral</td>
<td>7978</td>
<td>5775</td>
<td>1.38</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Note. O/E (Observed over Expected) values are based on frequencies found in the CORLEC Corpus of conversational Spanish; the conversational part has been used, comprising of 196,917 words. Utterances were separated and otherwise words were joined. The close relation between orthography and phonology in Spanish allowed to obtain a phonemic transcription; however, possible assimilations that were not annotated by the transcribers have been lost. The expected values are again based on the frequencies of the phonemes in consonant biphones; using the frequency over the whole corpus gives even more pronounced differences. For ease of comparison, Dutch O/E values from table 2.1 are repeated here.
that the learning is based on Dutch input, the Spanish L1 learners cannot be assumed to have learned from actual /esC/ sequences in the input, as Dutch does not have many words containing /esC/ sequences. Unepenthesis Good /sC/ clusters encountered by the learners apparently help build a representation of epenthesisised (/esC/) clusters, while the same clusters, together with all other /sC/ clusters, do not build a representation of (faithful) /sC/ clusters. The latter would remove the illegality and thus the perceptual illusion, but that does not happen.

It is of course possible that more evidence is needed to unlearn illegality than to learn a gradient difference in wellformedness. In addition, it is possible that learning takes place on the lexicon, not the input. The representations of L2 listeners do not have to be the correct Target Language form. L2 listeners filter lexical representations through the Source Language filter, as concluded by Pallier, Colomé, & Sebastián-Gallés (2001), who state that “word recognition uses a language-specific phonological representation and that lexical entries are stored in the mental lexicon as abstract forms.” Their conclusion was based on the fact that Spanish dominant Spanish-Catalan bilinguals show priming for Catalan words that differ from the target words in only a vowel contrast that Spanish does not have. Spanish dominant bilinguals were primed by /neta/ ‘granddaughter’, for /nEta/ ‘clean’, while Catalan dominant bilinguals were not. The authors conclude that the ‘a lack of sensitivity to difficult L2 phonemic contrasts (…) extends to the way L2 words are represented in the mental lexicon”. It is possible that a similar thing happened for the Spanish group in this experiment. In this case, the contrast is not between phonemes, but between phonotactic structures, /esC/ and /sC/.

Lexical representations containing the perceptual illusion cannot serve to learn that /sC/ clusters are legal in Dutch and in contrast with epenthesisised versions, while the lexicon with perceptual illusions contains a higher number of epenthesis Good clusters than epenthesis Neutral clusters. However, some researchers state that continuous speech is the source of phonotactic learning. Adriaans & Kager (2010) show computationally that continuous speech without word boundaries contains enough information to induce quite a few phonotactic constraints and Onishi, Chambers, & Fisher (2002) showed that new phonotactic constraints can be learned after brief exposure to syllables exhibiting that constraint. If this is so, the source of phonotactic knowledge is apparently filtered speech and only ‘illusionary’ gradient wellformedness can be learned, while illegal structures stay unattested and their legality will not be acquired.

The absence of a perceptual illusion for the Japanese group is at odds with this view, as it suggests illegality can be unlearned. As this is a null result that could also have its origin in problems with the experimental
set-up, the Japanese group’s performance cannot be taken as a falsification of the special protection for illegality knowledge, as no illusion was found at low levels of proficiency either. It should also be noted that perceptual illusions are not necessarily 100%; the Japanese participants tested by Dupoux et al. (1999, p. 1570) were, compared to French subjects, significantly more likely to perceive an epenthetic vowel /u/ in stimuli like [ebzo]. They were not performing at random (vowel length had a significant influence on the detection of the vowel, $p < 0.005$). If these listeners would start learning a second language, they already start out with a less than full perceptual illusion, as in the present study.

To explain the Spanish group’s results, a very strong Spanish perceptual filter has to be assumed, persevering to the moment in the acquisition trajectory where Dutch gradient wellformedness is being learned, the acquisition of wellformed sC-clusters goes through the perceptual filter and the representations contain the epenthetic vowel. This raises the question why these representations, acquired behind the filter, are not activated by the faithful clusters, that should also be perceived with an epenthetic vowel, be it an illusionary one. In sum, there is a perceptual illusion for the Spanish learners, that removes the contrast of epenthesis vs. non-epenthesis, but this contrast is relevant for wellformedness representations.

This means that the acquisition of wellformedness uses either a filtered lexicon or filtered continuous speech as input and thus indeed contains perceptual illusions. However, the acquired ‘illusionary wellformedness’ facilitates only input completely faithful to it. Hence, the situation in the Spanish group is the reverse from the Native-Es group: in the Spanish group, the representation of wellformedness is specified to contain the epenthetic vowel, in the Native-Es group, it is specified not to contain it.

The theoretically possible unification of categorical phonotactic illegality and gradient phonotactic is thus at odds with the data for a number of reasons. First, the facilitation by wellformed structures (good sC-cluster for native speakers of Dutch, epenthesised good clusters for Spanish L1 learners of L2 Dutch) takes the form of a rapid connection between unfiltered input and representations of lexical entries, while perceptual illusions allows to ignore mismatches between input and lexical representation as problematic. Such knowledge makes sense, as ignoring mismatches that are illegal cannot lead to a loss of contrast when applied in the correct language. Native listeners do not fully activate a word when it is presented with an epenthesised cluster, while Spanish L1 listeners do as the epenthesis is not contrastive in Spanish. Hence, the representation of illegality comes in the form of knowing when certain properties of the input can safely be ignored, providing robustness against noise in communication, while facilitation comes in the form of quickly
recognising certain combinations in the input, allowing an efficient mapping from the input to higher-level representations of lexical entries.

Second, facilitation and filtering cannot be reduced to each other. Facilitation is not similar to ignoring details or a bias to perceive a structure, because it is sensitive to the presence or absence of the epenthetic vowel: The Native-Es group needed the vowel to be absent for facilitation to take place, while the Spanish group acquires the wellformedness of epenthesised clusters and has increasing facilitation for the epenthesised prime only. Filtering can also not be reduced to strong facilitation of legal structures, as facilitation only occurs for structures that pass the filter; the Spanish group does not acquire wellformedness for unepenthesised good clusters; if the epenthesised representations are very easily activated by unepenthesised input, the absence of the vowel should not matter.

Third, the filtering is separate and prior to the representations of well-formed structures for not only acquisition and activation, but quite likely also for adaptation. If L2 learners have representations of wellformed epenthesised clusters and these are like phonemes, these representations would have to be subject to perceptual learning (see for an overview Kraljic, Brennan, & Samuel, 2008; Samuel & Kraljic, 2009); every time the cluster is encountered without the epenthesis and in a lexically unambiguous context, the (second) language learner can use the information from the lexicon to adapt his or her phoneme representation to the speaker. All Dutch sC-clusters are lexically unambiguous, so if Spanish learners speak Dutch with native speakers, they should constantly be recognising words that they represented with epenthesis, while the actual productions do not contain this epenthesis. This should incite perceptual adaptation of phoneme-like sC clusters towards non-epenthesised versions, giving a foothold on stopping the perceptual illusion. As soon as the cluster can be perceived faithfully, the input serves to learn the wellformedness of sC-clusters without epenthesis. However, such an effect is absent from the present results.

Given the above, the enterprise to unify phonotactic effects and represent them on one large wellformedness scale has to be abandoned. Phonotactic illegality knowledge is used to abstract input towards legal structures in the language concerned. The wellformedness that is acquired is that of combinations that occur frequently after the filter. These combinations are easier to recognise or process than others. In a second language, gradient wellformedness is easier to learn than categorical contrasts unknown in the L1, due to the filter protecting itself by distorting the input and/or the lexicon of the L2 learner. Wellformedness knowledge is of a different kind and does not distort the input; it is activated when wellformed combinations are recognised in the input. However, the wellformedness knowledge is distorted
by perceptual illusions. Even if there is one linguistic concept of phonotactic wellformedness, this concept does not correspond to a psychologically real representation that is tapped into by the mechanisms causing facilitation and filtering.
SPEECH SEGMENTATION

ABSTRACT

Phonotactic cues affect speech segmentation early in the lexical recognition process. McQueen (1998) showed that words are harder to spot when misaligned with phonotactic boundaries. These results are fundamentally ambiguous: they are commensurable with segmentation on a either a prelexical or a lexical level. This study present a replication of the McQueen results with a newly developed eye-tracking methodology that detects segmentation attempts in the absence of lexical recognition. Without lexical recognition, listeners still attempt to segment the speech stream. The phonotactic cues that modulate segmentation are partially confounded with the possibility that partially recognised lexical items affect segmentation. However, the results of a second experiment suggest that if segmentations do not differ on lexical likelihood, but only on phonotactic legality, legal segmentations are preferred. An additional experiment compared learners of Dutch with a Slavic native language to native speakers. The cues that drive segmentation show to be both illegality of clusters and their wellformedness. The Slavic listeners, in contrast to the native listeners, do not reject segmentations that are illegal in Dutch. However, they only consider illegal segmentations if more well-formed segmentations do not lead to word recognition. Keywords: Speech perception; Word recognition; Speech segmentation; Lexical access; Eye-tracking; Phonotactics

3.1 INTRODUCTION

Speech segmentation is the process that maps a continuous stream of speech sounds onto a chain of words. Effects of sublexical properties of the speech stream on segmentation have been found at the output of this process, for instance on the latency and accuracy of word recognition (see for overviews Quené, 1989; Cutler, 1994; Saffran et al., 1996; Davis, Marslen-Wilson, & Gaskell, 2002; Mattys et al., 2005). However, there is no consensus on the properties of the segmentation process and on the way sublexical knowledge affects it. Two main ways of conceptualising the process can be identified (Mattys et al., 2005). First, there is a sublexical viewpoint, following the idea that sublexical cues operate independently, generating segmentation hypotheses, e.g. in the form of boundary detection (Quené, 1989; Christiansen, Allen, &
Seidenberg, 1998) or in the form of initiating lexical look-up at strong syllables (Cutler & Norris, 1988). There is also a lexicalist viewpoint, under which sublexical information modulates hypotheses generated by lexical recognition (Norris, 1994; Norris et al., 1995; McQueen, 1998; Norris & McQueen, 2008).

This chapter addresses the influence of phonotactics, one of the types of sublexical cue. Its influence on segmentation is not under debate; the issue is the locus and type of the influence, i.e. the way phonotactic knowledge is employed in segmentation. A landmark study on phonotactic effects in segmentation by McQueen (1998) contrasted the recognition of Dutch words, e.g. /rOk/ rok, ‘skirt’, when they were hidden in different types of nonsense string. In one condition, named the Aligned condition, e.g. in the auditory string [fimrOk], it is easier to spot the target than in the Misaligned condition, e.g. [fidrOk]. The difference between the conditions is the phonotactic status of the cluster containing the word boundary. The cluster */nr/ is not legal and it is thus ‘phonotactically better’ to split it. On the other hand, the cluster */dr/ is legal and it is better not to split it, as the voiced obstruent */d/ is not legal in coda position.

For simplicity’s sake, the phonotactic influence on segmentation can be described as follows: phonotactic properties of sound combinations make it easier to split the speech stream at certain locations than at others. An important question is whether the segmentation process is executed during word recognition or before. One possible answer is that segmentation hypotheses are only generated by lexical knowledge. This type of answer will be referred to as a lexical account. In this account, segmentation starts with the identification of words that are present in the speech input, which are subsequently carved out of the speech stream. If it is phonotactically easy to split the stream at the boundaries of the words that have been identified, the word is recognised with greater ease. The other option is to assume sublexical cues are used at a prelexical level. In this prelexical account, the speech stream is simply split where it is easiest, after which words are identified in the resulting chunks.

The sensitivity of the segmentation process can be the result of phonotactic knowledge, but could also reflect properties of the lexicon. A wellformed sound combination is the start of many words; hence, starting word recognition at the onset of wellformed sound combinations and not at the onset of illegal sound combinations might be the effect of phonotactic knowledge, but would also be predicted if segmentation attempts were the result of the activation of lexical items by imperfect matches. In the [fimrOk] – [fidrOk] example, no words start with */mr/ and */mr/ is phonotactically illegal. Compared with */dr/, which is legal and the onset of many words, it is thus likely that a segmentation at */r/, the onset of the actual embedded target word /rOk/
rok ‘skirt’, is the only option in the Aligned \[\text{fimr}+\text{ok}\], but not in the Misaligned \[\text{fidr}+\text{ok}\], for both phonotactic and lexical reasons.

McQueen (1998) takes a combined lexical and phonotactic stance in his analysis. He concludes that phonotactic knowledge provides information on the likelihood of lexical hypotheses. According to McQueen, the activation of lexical items, such as the target words in the experiments, is modulated by the phonotactic wellformedness of the segmentations corresponding to the lexical boundaries. The presence of the target word in the example leads to the generation of the segmentation hypotheses \[\text{fimr}+\text{ok}\] and \[*\text{fidr}+\text{ok}\], in the Aligned and Misaligned conditions, respectively. As the latter is phonotactically illicit, it is harder to recognise the target word in the Misaligned condition.

The lexical analysis of McQueen is parsimonious, because it employs a theory that had already been proposed to account for other empirical findings. These assumptions include that word recognition is a competitive process using lexical cues and that segmentation is sensitive to the Possible Word Constraint (Norris et al., 1997). To explain the phonotactic effects on speech segmentation, McQueen only has to add the assumption that phonotactic knowledge provides cues. The cues he proposes are of two types: cluster illegality and segmentation illegality. For the Aligned condition (e.g. \[\text{r}+\text{ok}\] in \[\text{fimr}+\text{ok}\]), the fact that \[*\text{mr}+\text{ok}\] is illegal word-internally ensures that the lexical hypothesis is in accordance with phonotactics. In the Misaligned condition, illegality of voiced obstruents in coda position marks the \[d\] as a phonotactically likely onset (or very unlikely to not be an onset). The result is an infelicitous segmentation, e.g. \[\text{fi.d.r}+\text{ok}\], when phonotactic and lexical cues are combined. This type of segmentation is illegal according to the Possible Word Constraint (PWC, Norris et al., 1997), according to which a sequence of speech sounds can only be a word if it contains a vowel. The illegality of the lexically induced boundary inhibits the word that induced the boundary, making its recognition harder in comparison to the other condition.

McQueen does not explicitly take a stance on the type of phonotactic knowledge that provide the cues causing the phonotactic effect on segmentation. Although the analysis he gives is warranted by his experiments, the results themselves are fundamentally ambiguous with regard to two issues. The first issue has already been introduced, namely the question whether segmentation hypotheses are generated by lexical recognition alone, to be later checked by phonotactic knowledge, or the other way around.

The second ambiguity lies in the type of knowledge that explains the effects. There are at least three different ways to formulate a theory that links possible knowledge of the listener to McQueen’s observations of speech segmentation. First, the listener process can employ phonotactic knowledge of cluster illegality: clusters such as \[\text{mr}+\text{ok}\] are illegal within one syllable and hence provide a
cue that there is a boundary that solves this illegality, which constrains the hypothesis space to those segmentations that contain a boundary in the illegal cluster. Second, the higher probability of the legal sound combinations as word onsets, compared to the illegal ones, can cause more competition from alternative segmentation hypotheses in the Misaligned condition than in the Aligned condition, because in the latter there are no words starting with the illegal cluster that can compete with the target word. This latter option is also in line with the experimental evidence and analysis provided by van der Lugt (2001), who found that higher probability of the sound sequence within a target word makes it easier to spot the target. Third, phonotactic knowledge can be used to signal the illegality of certain segmentation hypotheses, importantly those containing voiced obstruents in codas. These segmentations should then be inhibited.

As experiments on segmentation generally compare an illegal and a more wellformed cluster, it is unclear whether knowledge of illegality or knowledge of wellformedness is the driving force for segmentation. In other words, the first and second explanations are confounded; both are supported by the results of McQueen (1998) and similar studies, but one of the two explanations would suffice. If the wellformedness explanation suffices, the phonotactic effect can be further reduced to an effect of lexical knowledge, removing the need for a separate representation of phonotactics in the explanation.

In sum, the origin of boundary hypotheses is unclear, as is the type of knowledge that drives segmentation. The two types of boundary detection, one based on sublexical cues and one on lexical cues, are not mutually exclusive, except in the case of the inhibition of illegal segmentations. The string */fid/, the non-target part of /fidrOk/, is phonotactically illegal in Dutch, and also not a word. If it is compared to /fim/, however, both are not words, but /fim/ is phonotactically legal. Nevertheless, the fact that /rOk/ is easier to detect after /fim/ than after /fid/ does not prove that the phonotactic illegality of */fid/ matters; the information that is employed by the listener could also be the illegality of */mr/ or the wellformedness of /dr/, as well as the larger number of lexical items starting with /dr/.

An argument for the usage of phonotactic illegality as a cue for segmentation, such as the illegality of */mr/, can be found in the results of a study by Weber & Cutler (2006). They used the methodology of McQueen (1998) and observed segmentation behaviour of second language listeners, namely German learners of English, listening to English. These listeners were compared to native listeners of English. The words the participants had to spot started at the second phoneme of clusters, e.g. the words started with /1/ and were embedded in nonsense strings such that the /1/ formed a cluster such as [s1] and [j1]. The crucial clusters were chosen to be either illegal in German, but
3.1 Introduction

legal in English (the [sl] type), or legal in German but illegal in English (the [J] type). The performance is compared to clusters that are either legal or illegal in onset position in both languages (e.g. [kl] and [nl], respectively).

German learners of English used both English and German illegality to segment English speech: if a consonant cluster was illegal in one of the two languages ([sl], [J], [nl]), it was more likely to be split up, leading to a facilitating effect on the spotting of words that start at the second consonant of the cluster. This effect was found by comparing with the overall legal clusters ([kl]) that could legally either be split up or kept intact. English listeners only showed facilitation by clusters that were illegal in their language ([J], [nl]), i.e. by clusters that had to be split up.

Although Weber & Cutler do not draw this conclusion explicitly, their results are an argument for the view that knowledge of phonotactic illegality helps segmenting speech by providing a cue for a boundary in the illegal cluster, because clusters that were illegal in one of the languages were at the same time legal in the other, but were still split. Hence, if legal clusters were hard to split, the L2 learners should have performed less well on both German legal and English legal clusters, as compared to the clusters that were illegal in both languages. If illegal clusters are easy to split, the predicted behaviour matches the results reported by Weber & Cutler, which suggest that knowledge of illegality is indeed used in segmentation.

However, the finding that knowledge of phonotactic illegality from two languages was combined and applied, instead of wellformedness, does not entail that wellformedness by itself is not used in segmentation. To separate the influence of phonotactic wellformedness from the influence of lexical properties, however, the recognition of wellformed words is not an ideal testing ground. Some models of word recognition, including the ART model proposed by Vitevitch & Luce (1999), suggest that frequent sound combinations are represented sublexically and make the activation of words containing frequent combinations easier in certain tasks. Hence, even if a wellformed word is as easy to segment as a not-so-wellformed word, it is possible that the wellformed word is recognised quicker by itself. If this is correct, the effect of phonotactic wellformedness within the target word does not affect the segmentation process directly, but only through lexical recognition.

To address this issue, however, the question whether phonotactic knowledge is applied prelexically has to be answered first. Next, if there turns out to be prelexical segmentation, the question whether non-words that are phonotactically wellformed are also more likely to be segmented from the speech stream can be asked and answered.

In summary, to study the questions stated above, first the null hypothesis that segmentation hypotheses are only made when the presence of a lexical
item is detected has to be tested. This hypothesis is at odds with the idea that the listener can employ phonotactic knowledge independently to segment the speech stream, before lexical recognition takes place. However, if there are segmentation attempts that do not correspond to lexical items, these attempts might still be the result of lexical activation based on a partial match, instead of on phonotactic knowledge; the lexical account of segmentation is only truly falsified if (prelexical) segmentation attempts are inhibited by phonotactic illegality of the segmentation. In this case, the hypothesis that wellformedness only affects word recognition and the hypothesis that wellformedness also affects segmentation can be proposed and answered.

To answer the first question, whether there is a prelexical effect on segmentation, the null hypothesis that there is no attempt to segment speech before the identification of a lexical item will be tested. Note that McQueen did not claim that prelexical segmentation does not exist; his experiment simply gave no reason for the assumption that prelexical segmentation exists as an independent process. The difference between the prelexical and the lexical positions is taken to the extreme here to clarify the possible contribution of phonotactics to speech segmentation. If previous literature is classified as proposing a lexical view, this classification does not imply that its authors deny the possibility that phonotactics operates before lexical knowledge enters the process. As Norris et al. (1997, p. 233) suggest, “it is no trivial matter to incorporate phonotactic constraint information, or language-specific metrical cues, into the initial analysis of the signal which is used as input to the word recognition process.” In later work, Norris & McQueen (2008, p. 363) mention that “[a] more complete model would therefore include modulation of the computation of P(PhonemeString) as a function of transition probabilities.” The present study’s goal is not to refute the Shortlist models of word recognition because they treat segmentation at the lexical level, but to look for a more specific account of the percolation of information provided by phonotactic knowledge towards the process of segmentation.

There are a priori arguments for the existence of a prelexical generation of segmentation hypotheses. Phonotactic knowledge is an abstraction of lexical knowledge, namely of the possible shapes that word forms can take. If this abstraction is more compact than the collection of actual word forms (the form part of the lexicon) and if consulting a more compact (smaller) body of knowledge places less demands on processing, consulting phonotactics for segmentation hypotheses provides an advantage over consulting the lexicon from the onset of the segmentation process, as mentioned by Cutler & Norris (1988): “(…) [A] lexical access attempt could be begun at every segment. This procedure (…) would result in a majority of fruitless access attempts, but these could perhaps be reduced by considering, say, two-segment sequences
and ruling out access attempts when the sequence postulated to begin the word was phonologically illegal (...)”. In other words, if looking up if a sequence is illegal is more efficient than looking up the words starting with it and finding no results, it makes sense to use phonotactics in prelexical speech segmentation in place of full lexical look-up.

The potential of phonotactics to offer useful information for segmentation when words are not yet recognised has been proven by computational modelling as well (among others, by Church, 1987). Nevertheless, as usefulness does not entail usage, these computational findings are only strong arguments, not proof, for the use of phonotactic knowledge on a prelexical level of segmentation. Only empirical evidence would allow to posit the existence of a prelexical stage as the locus of phonotactic effects on segmentation.

Segmentation in the absence of lexical recognition is difficult to observe. The McQueen (1998) and van der Lugt (2001) experiments, as well as the priming studies employed by Mattys et al. (2005), only yield results on the lexical level, i.e. only when a word is recognised. Participants had to press a button when recognising a word in the nonsense string and report the word, which only allows the measurement of the accuracy and latency for segmentations that could have been initiated by lexical cues. Such measurements cannot falsify the lexical null hypothesis. A new type of experiment is needed to directly answer the question whether prelexical segmentation exists and whether it is sensitive to phonotactics. Present day technology makes it possible to take a more detailed look at segmentation as it unfolds. By repeating the McQueen (1998) experiment while tracking the eye movements of the participants (Experiment 3), the phonotactic effect was observed at the potentially prelexical stage, namely in a non-lexical condition, in which the stimulus contains no word at all. The predictions of the prelexical and the lexicalist hypothesis differ here. The eye-tracking methodology used in the present study employs the visual world paradigm, which has been shown to tap into the speech recognition process while it is still unfolding (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Magnuson, Dahan, & Chambers, 2000; McMurray, Tanenhaus, & Aslin, 2002).

However, the classical visual world paradigm is still based on the activation of lexical items. In experiments using the visual world paradigm, a computer screen with a number of objects is shown. Eye fixations of a participant on an object on the screen indicates preferences for this item; over a sufficiently large number of items and participants, fixation probabilities are correlated to the probability that the participant is thinking about the object. The object can be primed by an auditory stimulus containing the its name; this principle allowed for the paradigm to be used in speech recognition research. Even an
incomplete match between an auditory stimulus and the name of an object has a positive effect on the probability of fixation on the object (Allopenna et al., 1998).

It is likely that if the McQueen (1998) experiment were adapted to the visual world paradigm, by showing objects, including one representing the target word, fixations to the target object would be sensitive to the phonotactic alignment of these experiments, but this again only shows the effect at the lexical level, as the objects’ names are lexical items. To study the segmentation hypotheses of the participants, if they exists independently of lexical recognition, named objects will therefore not suffice. Additionally, alternative segmentations, that do not coincide with the boundaries of the target word, might be represented by objects with a name that matches the speech stream from the alternative boundary. E.g., if the auditory stimulus is the Misaligned [fiədrək], with the embedded target word /rək/, and one of the objects presented visually has a name starting with /dr/, e.g. /drəp/ drop ‘licorice’, this object would be fixated if participants entertain the segmentation hypothesis /fi.ədrək/. Such fixations on the phonotactic boundary, misaligned with the target, would also make it less likely to fixate on any object representing the target word. However, a segmentation hypothesis with a boundary before an illegal cluster, such as /fi.əmərək/, cannot be represented with a match of more than one phoneme, as no words start with illegal clusters. Using only objects that match the auditory input for not more than one phoneme would greatly reduce the possibility of attraction of fixations by any non-target object, as the target fully matches the auditory input.

To work around both the problem of the lexical recognition needed for fixations on named objects and the difficulty of representing the illegal segmentation with such an object, orthographic representations have been used. McQueen & Viebahn (2007) showed that printed words instead of depictions can be used as the objects in a visual world experiment as well. If letters correspond closely to phonemes, positions in the speech stream can be represented by letters. Letters would then fulfil the task of representing segmentation hypotheses. As letters are not lexical items and do not mismatch the auditory input if the corresponding phoneme is present at any position, fixations on letters can be assumed to represent recognition of the phoneme/letter in the auditory stream. In the experiments in this chapter, as well as in the next, participants listened to fragments of continuous speech and were asked to spot words, just as in the McQueen (1998) experiment. Every hypothesised boundary is expected to draw the listener’s attention to the phoneme after the boundary. If this phoneme corresponds to a letter on a computer screen, the listener might be more likely to fixate on this letter, anticipating it to be the start of a word. To encourage such fixations, the participants in the experi-
ments presented in this article were asked to respond by looking at the first letter of any word they spotted. This encourages anticipation; if a boundary is detected before a word has actually been recognised, the participant can fixate on the letter in anticipation of it being the first letter of a word.

If the participant detects a boundary that does not correspond to a word, i.e. makes an incorrect, unlexical segmentation hypothesis, the fixations on the ‘wrong’ letter reveal this prelexical segmentation attempt. This allows the detection of distraction by incorrect segmentation directly, not only through the inhibition of word recognition, but also by showing the facilitation of the alternative. To measure the phonotactic role in the generation of erroneous segmentation attempts, the visual presentation (on the computer screen) included letters corresponding to the alternative boundaries that do not correspond to the target word.

If there is a phonotactic influence on the prelexical generation of speech segmentation hypotheses, the phonemes right after phonotactically wellformed boundaries are expected to be likely to become word onsets. Prelexical segmentation does not necessarily predict that participants anticipate the phonologically legal onsets to be the onset of a word and look at the corresponding letter, awaiting the rest of the word to follow, but if such behaviour is observed, it does support the existence of prelexical segmentation. Alternatively, if speech is only segmented as a result of word recognition, there should not be fixations on letters not representing words, as no lexical hypothesis will be formed in such cases. However, if fixations on non-word letters are recorded, they do not entail that listeners hypothesise the segmentations corresponding to these letters. The observation could also be due to noise. Such observations can therefore not be taken to be direct evidence for prelexical chunking. However, if non-target fixations show sensitivity to phonotactic cues, the effects of phonotactics on segmentation can be assumed to be due to distraction by alternative segmentations in case the target is misaligned. If the lexicalist account is correct, phonotactic information would only modulate the likelihood of fixating on letters that are actually word-initial. However, if the prelexical account is correct, a lexical hypothesis is not necessary for listeners to assume a boundary on phonotactic grounds; hence, even if no word is present in a stimulus, phonotactic cues can guide segmentation and hence affect fixations.

As it is not clear if the participant’s task in the experimental method described above is natural enough, an additional requirement for rejection of the lexicalist null hypothesis is that the new method yields results similar to McQueen’s, with regard to latency and accuracy. If this is the case, the new method can be supposed to tap into speech segmentation in the same way as in the experiments of McQueen. If the fixation patterns mirror the latency and
accuracy effects, it can be assumed that fixation proportions can be analysed as indicative of segmentation hypotheses.

If fixation proportions are indicative of segmentation hypotheses, they can replace the latency and accuracy of word recognition as the dependent variable in the experiments. Using fixations on letters as dependent variables also makes it possible to test what kind of phonotactic cues are used in segmentation.

3.2 Experiment 3

3.2.1 Participants

Thirty-two Dutch native speakers, students that volunteered to be in the participant pool of the Utrecht Institute of Linguistics OTS, participated for monetary compensation. More participants were recruited but could not finish the experiment due to problems in the calibration process of the eye-tracker, usually related to a participant’s way of fixating or to a lack of contrast between pupil darkness and darkness of other regions within the tracker’s receptive field, usually eyelashes.

3.2.2 Materials

Items are bisyllabic strings. 120 Dutch words were used to construct the experimental items, that were nested in a Position condition: 60 had words embedded at the beginning of a nonsense string (Initial), 60 at the end (Final). All of McQueen’s 80 words were included in similar strings, but sometimes the strings were changed to avoid that a letter presented visually occurs twice in the same item. All syllables contained unreduced vowels.

All items occurred in four versions, grouped in two conditions (see Table 3.1). The Lexicality condition reflects that every item could actually contain the Word or a derived Non-word that shares the phoneme at the phonotactic boundary with the Word. This Word or Non-word will be referred to as the target; a Non-word target can thus not be spotted as a word, but it does correspond to one of the letters presented visually, as will be discussed below. The strings contained no other words than the target word, hence the Non-word items did not contain any embedded words.

Table 3.1 gives examples of the interaction between Lexicality and Phonotactics and the construction of the auditory material. The Phonotactics condition was a manipulation of the string around the word boundary, such that the target was either Aligned or Misaligned with a phonotactic boundary. Final items all start with one consonant (r in the example /rɔk/). This consonant is
preceded by one consonant of the nonsense syllable. The phonotactic boundary in the Aligned condition coincides with the target, as an illegal cluster is formed by the last consonant of the nonsense string and the first of the target (/m/ and /r/ for /rOk/ in Table 3.1). In the Misaligned version, a boundary before the last consonant of the nonsense syllable is predicted. This boundary is the result of a ban on final voicing for obstruents that made the last consonant of the nonsense syllable (d in the example) illegal as a coda. The cluster formed by the two consonants in the Misaligned versions was always wellformed ( /d/ and /r/ in the example /rOk/ in Table 3.1). In the Initial position, targets were Aligned at their offset when the following biconsonantal onset of the nonsense syllable was legal on its own (/vr/ in the example /wil/), but not together with the targets’ monoconsonantal coda (*/lvr/ in the example). The Initially embedded, Misaligned targets were followed by two consonants that were illegal together, forcing the first consonant of the nonsense syllable to combine with the last consonant of the word; this always formed a legal coda. Table A.1 (p. 190) tabulates all contexts, i.e. all clusters that could occur at the onset or offset of the target words.

Items were rotated in a Latin Square design through the Phonotactics and Lexicality conditions. For tight temporal control, the strings were synthesised with MBROLA (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996); phoneme durations were based on Waals (1999). As duration influences segmentation (Quené, 1989), compression rates of phonemes in illegal clusters were made identical to those in their legal counterparts. This reduces the chance of speaker interpretations of boundaries and thus of other sublexical cues creeping in.

Four letters were displayed in every trial. These letters depended on the item: one corresponds to the target and two letters correspond to distractors, namely the consonant immediately following the phonotactic boundary in the Misaligned condition (Misalignment distractor) and the corresponding position in the Aligned condition (Alignment distractor). The fourth position

<table>
<thead>
<tr>
<th>Position: Final</th>
<th>Position: Initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item /rOk/</td>
<td>Item /wil/</td>
</tr>
<tr>
<td>Phonotactics</td>
<td>Phonotactics</td>
</tr>
<tr>
<td>Lexicality</td>
<td>Lexicality</td>
</tr>
<tr>
<td>Aligned</td>
<td>Aligned</td>
</tr>
<tr>
<td>Misaligned</td>
<td>Misaligned</td>
</tr>
</tbody>
</table>

| Word           | Non-word         |
| [fim.rOk]      | [fim.rypt]       |
| [fi.drOk]      | [fi.drypt]       |
| Non-word       |                  |
| [wilm.ro]      |                  |
| [wilm.rop]     |                  |

Letters shown: r, m, d, f

Table 3.1: Examples of stimuli and conditions. Boldface indicates embedded words, dots indicate phonotactically preferable boundaries.
was used for the initial phoneme in the case of finally aligned targets and for the third consonant in the cluster at the end of the target for initially aligned targets. Each item was always presented with the same four letters. The letters used as non-targets in the experimental items were targets in 120 filler items (50% containing a word).

As the experimental Final items started with only an r or l, fillers had to reduce the probability of an r or l to be the correct answer. The fillers contained words starting with the letters that were used as distractors, so that the probability that letters often used as distractors were correct answers also rose. All letters used in the experiment were correct in roughly 25% of the time, which is chance level given that four letters are presented. In addition, the fillers made sure that the number of consonants did not predict if the item was embedded initially or finally.

3.2.3 Procedure

The experiment was run with an Eyelink I head-mounted eye-tracker, sampling fixation position at 250 Hz. Participants were first seated, after which the helmet was positioned on their heads and the camera’s were placed appropriately. After this, calibration is necessary, which takes a few minutes. Calibration was tested performed with Eyelink software. Participants had to fixate on dots on 9 positions on the screen. If a validation round proved that the calibration was accurate enough, the real experiment started. Participants received instructions only then. They practised on eight trials, which they could repeat when they required or when the experimenter noticed unusual behaviour. The practice trials were easier than the normal trials and participants were expected to succeed in at least the last trials. If this was not the case, the solution was often to recalibrate the tracker. Participants were often unsure if they had understood the instructions, but were quickly reassured when they had practised.

Trials started with the letters on screen in large blue rectangles; the rectangles corresponded to fixation regions. After 3000 ms, a dot appeared in the centre of the screen. Participants were told they could read the letters as long as they wanted and fixated on the dot when ready. The dot was actually placed on the screen by a small helper application that is part of the eye-trackers software. This application takes over from the experiment when a fixation on the dot is detected. The eye-tracker software then changes its own settings, in order to improve the spatial accuracy. This procedure is called drift correction. If it was completed, the dot was removed. Participants do not generally notice the drift correction, as they only see the dot appearing and disappearing.
The auditory stimulus started playing 250 ms after the end of the drift correction. Participants were instructed to indicate if they spotted a word by looking at its first letter, present in one of the four regions. They knew that looking at wrong letters was allowed and did not result in errors and they were encouraged to give a quick response. Participants could only give a correct response in Word trials. The eye-tracker only accepted responses after the sound was played. If a participant fixated on the target letter for 300 ms in Word trials, the square around the letter changed colour to green and the trial ended. If this happened, the trial was recorded as correct and time between target offset and the onset of the fixation was recorded as the reaction time. (McQueen, 1998, also recorded reaction times from target offset.) After 2500 ms, trials timed out; for Word trials, a timeout constituted an incorrect response. In the Non-word condition, all trials timed out and correctness was not recorded. 240 trials were presented randomly in four blocks.

The choice of the intervals between different stages of the trials was based on a pilot run, in which first the author and later six participants participated in the same experiment with different settings. The 300 ms fixation gave participants the idea that the computer actually responded to every intentional fixation they made. It was not easy to ‘walk through’ all four areas within the 2500 ms window before timeout. 200 ms is commonly believed to be the minimum time needed to plan an eye-movement. In theory, it should have been possible to fixate on the four letters one by one. In practice, this turned out not to be the case. Timeouts before 2500 ms made participants feel if they could not keep up with the experiment. A possible explanation might be the distance from the dot in the middle of the screen to the squares; participants had to look at the dot to start the sound. A more likely explanation is that participants enjoyed the experiment. Interviews with the participants in the pilot revealed that they enjoyed the challenge of word spotting and tried to fixate on the target as quickly as possible.

3.2.4 Results

Latency and accuracy

The latencies and accuracies per condition are summarised in Table 3.2. Participants fixated at least 300 ms on the target letter in 63% of the non-filler Word trials (1000 out of 1587, 95% confidence interval (0.60, 0.65)). The average reaction time is 786 ms (SE 13, 95% confidence interval (760, 811)). This is not a large difference from McQueen’s results of 43% and 708 ms for the stimuli that had two strong syllables; however, the fact that both are lower suggests a difference in the speed/accuracy trade-off; participants in the present study took a slightly longer time to respond, but responded more correctly. This can
| Position | Accuracy | | | Latency | | |
|----------|---------|---------|---------|---------|---------|
|          | Phonotactics | Phonotactics |          |          |          |
|          | Aligned | Misaligned | Total | Aligned | Misaligned | Total |
| Final    | 79.7    | 43.7     | 61.4   | 692 (20)| 745 (27) | 711 (16) |
| Initial  | 70.9    | 59.1     | 65.0   | 845 (24)| 913 (39) | 875 (19) |
| Total    | 75.8    | 50.4     | 63.0   | 755 (16)| 830 (21) | 786 (13) |

Table 3.2: Experiment 3, Word trials: percentage of correct answers and average reaction times (SE in parentheses).

probably be explained by the fact that picking the right letter out of the four shown makes it more likely to quickly guess the right answer, as compared to the task in McQueen's experiment, where participants had to correctly repeat target words if they spotted them, without any prior information about the properties of the target word. Errors were thus more likely in McQueen's experiment.

To see if the effects of the experimental conditions as found by McQueen were replicated, all Word trials were analysed in a mixed model with random effects for item and participant (Quené & van den Bergh, 2008; Baayen et al., 2008). The dependent variables were the log-transformed reaction time and the logit of the accuracy. Log-transformed latencies (measured from target offset) were higher for Misaligned words and lower for Initial items. Adding an interaction to the model did not lead to significant improvement and the interaction was not significant. The logit of accuracy (T. F. Jaeger, 2008) shows a similar pattern: Misaligned words were spotted less; Position effects were insignificant, but the interaction Phonotactics × Position reveals that Misalignment matters less in Initial Position, as McQueen also noted. The model estimates are shown in Table 3.3. The results correspond to those reported by McQueen, who found significant effects in the same directions, with one exception: McQueen found a significant interaction between Position and Phonotactics for the latency, as the alignment effects were more marked for Final than for Initial targets, while the present experiment did not even show a trend in this direction.

Fixations

Given that accuracy and latency show phonotactic effects in the Word trials, the hypotheses under discussion can be tested with this new experimental paradigm. The null hypothesis to be tested is that the phonotactic effects on segmentation are not due to distraction by alternative segmentations in the Misaligned but not the Aligned cases, but that they modulate only the
Table 3.3: Experiment 3, Word trials, modelled latency and accuracy. A mixed model with crossed random effects was fitted to the log-transformed reaction times measured from word offset; for accuracy a model was fit on the logit of spotting the word.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Latency (SE)</th>
<th>t-value (31 d.f.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>6.40547 (0.04113)</td>
<td>155.73</td>
<td>&lt;0.0001 ***</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>0.10299 (0.03731)</td>
<td>2.76</td>
<td>0.0118 *</td>
</tr>
<tr>
<td>Position-Initial</td>
<td>0.24174 (0.05220)</td>
<td>-4.83</td>
<td>&lt;0.0001 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Accuracy (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.9040 (0.2642)</td>
<td>7.206</td>
<td>&lt;0.0001 ***</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>-2.3184 (0.2161)</td>
<td>-10.730</td>
<td>&lt;0.0001 ***</td>
</tr>
<tr>
<td>Position-Initial</td>
<td>-0.2175 (0.3923)</td>
<td>-0.554</td>
<td>0.579</td>
</tr>
<tr>
<td>Misaligned × Initial</td>
<td>1.2341 (0.3131)</td>
<td>3.941</td>
<td>&lt;0.0001 ***</td>
</tr>
</tbody>
</table>

activation of the target. As fixations are to initial letters, while phonotactic manipulation affects Initial items at their offset, those items were excluded from the fixation analyses.

The null hypothesis is falsified if (1) target fixations follow the pattern of alignment, hence are sensitive to the Phonotactic condition, (2) regardless of the presence of a Word. If (1) is not true, target fixation probability cannot be used as a dependent variable corresponding to segmentation. Target fixations in the Word trials are the fixations on the letter that corresponds to the embedded word in the Word trials; they are thus predicted to follow the pattern of accuracy and latency for those trials. If the null hypothesis is true, the Lexicality condition should matter for the phonotactic influence on the target, as the target is only lexical at the Word level of this condition.

Another falsification of the null hypothesis can be provided if distractor fixations are responsible for this pattern, i.e. in the Misalignment trials, the distractor is more likely to be fixated on. As participants can only fixate on one letter at a time, every distractor fixation causes a lower proportion of target fixations. The two analyses of target and distractor fixations are thus not necessarily independent (they are dependent if the prelexical account is correct). To avoid capitalisation on chance, the significance threshold was (Bonferroni) corrected to 0.025.

The 95% confidence interval of the reaction times ended at 743 ms. The grand mean of target fixations to targets in a pilot showed a slowly rising pattern, but not in the first 100 ms after stimulus offset. The time windows for the fixation analyses were therefore restricted to 100–740 ms (due to the 4 ms sampling interval, 743 ms was not in the dataset). Growth Curve Analysis (Mirman, Dixon, & Magnuson, 2008; Barr, 2008) on the logit of fixation was
used, without aggregation as there is sufficient data. The target fixation analyses take the logit (log of the odds) that a fixation is on a region as the dependent variable, over all recorded fixations.

The statistical models of the fixations contain third-order orthogonal polynomial time terms, because visual inspection of the pilot data showed two bends in the fixation development. This was confirmed by looking at the significance of the three terms in a model excluding the experimental factors. Modelling started with a base with only the time polynomial factors and random effects for item and participant at each time level; the predictors were added stepwise to each polynomial, allowing likelihood tests to show if the fit improved significantly.

For target fixations, experimental conditions improved the model at all time levels ($\chi^2$ was respectively 3597, 94, 178, 40 (3 d.f.), $p < 0.0001$ in all cases). The best-fitting models are given in the appendix (Tables A.5 and A.6, p. 194). Figures 3.1 and 3.2 show these models imposed over the actual proportions. The modelled lines capture the relative curvature of the fixation probabilities, in as far as these hold for the whole population of both items and participants. Note that models were restricted by the time structure and calculated on the logit space, while in the graphs data and model are converted (back) to probability space.

The fixed effect lines are not necessarily at the minimum distance to the data points, due to use of random effects for items and participants at every time level. In data without time dependencies, the fixed effect is very close to the mean and the random effects have a mean of zero, causing the addition of the fixed effects and the random effects to be very close to the mean. In the data presented here, however, there are four random effects for each item and participant (one at each time level). At the beginning and end of the time window, the random effects at the quadratic and cubic levels have a larger influence on the predicted value. The fixed and random effects are calculated in such a way that for each data point, the difference between prediction and observation is minimal. However, this objective was reached better for the values in the middle of the time window. The random effects were on average zero. However, for participants and items for which the line was lower than the mean trajectory, less data was likely to be available, as these might fixate less in general. Hence, more data comes from participants that were on average better at fixating at the target than from participants that were worse, which means that more data points are corrected upward by the random effects than downward. This upward trend shows most at the beginning and end of the time window, due to the use of second and third powers of the time values. If the random effects are not zero, the fixed effects have to be below the data points. The average of the predicted values for each
data point (fixed effects and random effects added) is closer to the actual data points than to the fixed effects.

As can be seen in the graphs, the null hypothesis that fixations to Non-word targets are insensitive to the Phonotactic condition has to be rejected. Aligned targets were more likely to be fixated on and this effect is even larger for Non-words. The analysis confirms this, first because adding the Phonotactics condition significantly improved the mode and second because of the significance of the terms at the intercept, linear, and cubic level. The significance of the interaction Misaligned × Word shows that the presence of a Word weakens the effect of the Phonotactic condition. This effect is in the opposite direction as the null hypothesis, meaning this hypothesis can be rejected.

The null hypothesis also predicted that fixations to distractors should be unaffected by Phonotactics. This prediction is tested on all cases in which fixations are on one of the two distractors. Note that one of these distractors is present in the auditory stimulus. Distractor fixations are thus assessed on their correspondence to the Phonotactic condition, i.e. on whether the fixation is on the distractor actually in the stimulus or on the other one. The data set was restricted to the fixations on one of the two distractors and the dependent variable was if the fixation was on the distractor corresponding to the Phonotactic condition, or not. If the fixations to distractors were completely random, no condition would matter; nevertheless, the mere auditory presence of the distractor in the nonsense syllable could matter even if the Phonotactics condition does not. The crucial test is the effect for the Misalignment trials, where the corresponding distractor is aligned with the phonotactic boundary (while the target is not). The logit, and hence the proportion of fixations to the distractor, is higher in that condition, indicating that more distraction is caused by phonemes that are at phonotactic boundaries than by phonemes that are not. The effect occurs regardless of Lexicality: the Phonotactic effect is again larger in the Non-word condition.

3.2.5 Discussion

The latency and accuracy results of McQueen (1998) have been replicated. Not only does this provide converging evidence of phonotactic effects on word recognition and speech segmentation, it also means that the new methodology is unlikely to be affected by artefacts, or at least not more so than the original word-spotting paradigm. As Target fixations also show the pattern discovered by McQueen, the methodology gives another, possibly more sensitive measure of the influence of phonotactic alignment on word recognition.
The eye-tracking methodology shows continuous speech processing before its culmination in complete lexical recognition. This means that the experiment has the potential to show the existence of prelexical segmentation. For this reason, latency and accuracy scores were not treated as hypotheses to be tested, but as preconditions to test the hypotheses on the fixations. The lexicalist null hypothesis that phonotactic knowledge only affects the activation of words has to be rejected, as the Target fixations also show that phonotactics has an effect without word recognition. The hypothesis that phonotactic cues are used independently of lexically-driven segmentation is therefore supported. It is most likely that the influence occurs before lexical segmentation takes place: distractor fixations show that phonotactically well-formed segmentations are generated even when they conflict with lexical items. This means that lexical items are not essential for segmentation to start, even though the presence of a lexical item also guides segmentation. It is quite likely that lexical cues overrule phonotactic cues in many instances, as shown by the smaller effect of Phonotactics when a Word is present.

The statistical analysis shows that a model taking time into account can capture the general temporal dependencies between eye-tracking data. A fixation on time $t$ is likely to be followed by a similar fixation on time $t + 1$. 

Figure 3.1: Experiment 3, Target fixation probabilities, by Lexicality × Phonotactics. The fixed effects predicted by the model are superimposed as lines. Points are proportions aggregated over trials and 40 ms bins. The solid line at 0.25 represents chance level.
Figure 3.2: Distractor fixation probabilities by Lexicality × Phonotactics. Over all cases in which fixation was on a distractor, the probability that it is on the distractor corresponding to the Phonotactic condition is given. Points are proportions aggregated over trials. The solid line at 0.5 represents chance level.

but piling all fixations into one collection would amount to using the same measurement many times; aggregating over the whole trial would amount to throwing away the sensitivity of the whole paradigm. In the present case, it would also amount to having to assume the effect of phonotactics to occur equally over the whole segmentation process.

As the fixations on targets and distractors give insight into the unfolding phonotactics-driven segmentation, the methodology is of great use to test more specific hypotheses on the effect of phonotactic knowledge. A priori, one could raise concerns about the validity of the procedure: it is possible that participants go through the four letters until one lights up, but the accuracy data show this is not the case. The same holds for the strategy of looking at every letter that is present in the sound; this would entail that distractor
fixations are not sensitive to the Phonotactic condition, which was not the case.

The results answer the first question asked in the introduction, namely if there are segmentation attempts that do not correspond to the full recognition of a lexical item. However, a lexical item might have been recognised even when it was not actually present in the auditory stimulus. Incomplete lexical recognition could explain the phonotactic effect on the generation of segmentation hypotheses. For instance, in the auditory string [fidrOk], words starting with /dr/ might be activated, causing anticipation by looking at the letter 'd', even if no word starting with /dr/ ends up matching the rest of the string (neither /dra/ nor /drOk/ are Dutch words and the cohort is empty after the /k/).

3.3 EXPERIMENT 4A

Experiment 4a is aimed at addressing the origin of the information causing the generation of different segmentation hypotheses in the two levels of the Phonotactic condition. In a contrast between illegal clusters and legal clusters, phonotactic and lexical cues converge, both predicting that segmentations are not started at an illegal cluster. A segmentation splitting an illegal cluster is favoured in this situation, as compared to a situation in which there is a legal cluster, but this does not entail that the illegal cluster is recognised as such and split for that reason. It is also possible that there is a preference for starting segmentations before the legal cluster, driven by recognition of its wellformedness. Third, it is possible that segmentation that contain phonotactically illegal residues are avoided, which in this case explains why the voiced obstruents of the legal clusters are a problem for detecting a word boundary after these obstruents and before the target.

There are in theory two types of cues that phonotactic knowledge can generate. One type of cue indicates the best sites to insert boundaries in the speech stream, by marking phoneme sequences that are unlikely to occur within words, as in the computational models of Brent & Cartwright (1996) and Cairns, Shillcock, Chater, & Levy (1997). The term ‘splitting’ will be used as a shorthand for the insertion of boundaries.

Phonotactic knowledge could also provide cues that indicate where to chunk speech, i.e. to suppress the placement of boundaries, because phonotactics knowledge includes the wellformedness of often-combined phoneme sequences that have been stored as the result of frequent exposure to these combinations and because “internal representations guide perception”, as discussed by Perruchet & Vinter (1998, p. 249). This is also in accordance with the Cutler & Norris (1988) suggestion that phonotactic illegality can stop
useless lexical look-up. Chunking speech amounts to detecting contiguity, i.e. combinations of phonemes that are likely to co-occur within a word, and only generating segmentations that keep these combinations together.

The difference between chunking and splitting is important, because phonotactics is probably not powerful enough to correctly mark all word boundaries on its own. For instance, a recent model of the use of phonotactic knowledge for boundary detection, the StaGe model by Adriaans & Kager (2010), finds 40–45% of the boundaries in a corpus of spoken Dutch, while the other segmentation models compared in the article only reach higher values (up to 61% for troughs in transitional probabilities) at the cost of a higher false alarm rate (9–13% for StaGe, 22% for segmentation based on troughs in the transitional probability). Even if future models of phonotactics show better performance, it is not likely that human segmentation is driven by phonotactics only, as the influence of other sublexical cues has been shown (see also Mattys et al., 2005, for an overview of more evidence). Other sublexical cues and lexical cues are thus likely to provide the necessary extra information.

If phonotactics were to completely perform the whole segmentation task, it would not matter if it splits or chunks, because then the absence of splitting entails chunking and vice versa. Nevertheless, as this is not likely to be the case, splitting and chunking cannot be supposed to be mirror images, let alone to be mutually exclusive. If phonotactics marks some but not all boundaries, the speech in between the boundaries that are detected cannot be assumed to be only one word, i.e. the absence of boundary detection does not imply the detection of the absence of a boundary. Vice versa, if not all contiguity is detected, it is possible that two consecutive recognised chunks are actually part of the same word; it cannot be assumed that there is a word boundary between chunks. This means that the effects of chunking and splitting are potentially different. However, it is possible that human listeners use both mechanisms. According to computer simulations by Adriaans & Kager (2010), using both knowledge of contiguity and of markedness (illformedness) of sound combinations can improve the performance of a speech segmentation model, because it allows generalisations to be made for the ‘neutral’ phoneme sequences.

Along similar lines, Brent (1999, p.296) classifies segmentation models as either (1) supposing a boundary at sequences that are typical at the boundaries of utterances (as these boundaries are often marked by pauses, prelexical language learners can already notice them, unlike word boundaries) or (2) supposing a boundary at sequences with a low probability, or (3) by hypothesising whole words. This classification is partly based on the source of the phonotactic knowledge; this study investigates a different but related set of hypotheses about the human segmentation process, specifically phonotactically-
driven segmentation: phonotactics might independently guide segmentation by recognising boundaries (Brent’s 1 and 2) or by recognising possible chunks (3). Brent’s third option, positing whole word hypotheses, means that segmentation is a lexical process; however, in a prelexical stage, option (3) has to refer to chunks or incomplete matches to lexical items, instead of words.

Illegal clusters of the form */P_1P_2/* can be directly interpreted as straddling a likely word boundary. By definition, illegal clusters cannot be part of the same word, apart from cases of acoustic reduction. Illegality applies to [fimrOk], where the illegal string */mr/* can point out the segmentation /fim.rOk/, that helps with finding the word /rOk/. In addition, the string /fid/ is also illegal, as a voiced obstruent coda is illegal in Dutch. Dutch does not allow voiced obstruents in coda position, with the exception that there can be regressive voicing assimilation if the following obstruent is voiced (see e.g. Wetzels & Mascaro, 2001; Zonneveld, 2007, for overviews). Note that the exception does not apply to any of the phoneme combinations used in this study. Hence, if a voiced obstruent is encountered, a boundary cannot legally be placed after it, providing a phonotactic cue for chunking the voiced obstruent together with the next one, or for splitting before the voiced obstruent.

On the other hand, chunking wellformed combinations of phonemes of the form P_1P_2 to a single unit (P_1+2) can also explain the avoidance of segmentations in the legal clusters used in Experiment 3. If this has an effect, as suggested by van der Lugt (2001), this effect is confounded by the link between wellformed combinations and word representations, as proposed by Vitevitch & Luce (1999). If a combination of phonemes is no longer seen as a pair of separate phonemes, but as a single unit, as Vitevitch & Luce propose, a boundary within the combination is inhibited. Phonotactic wellformedness then chunks speech and reduces the number of lexical access attempts. In the aforementioned example, the nonsense string [fidrOk] contains wellformed /dr/; this close-knit combination could be kept together, to the detriment of the word-aligned segmentation /fid.rOk/. However, incomplete lexical activation, of words starting with /dr/, is also a possible cue for the chunking of wellformed clusters. Note that it is in principle not illegal to split a wellformed combination of phonemes; the influence of positive knowledge therefore has to be gradient.

Phonotactic wellformedness is not necessarily equivalent to high frequency of occurrence, but highly frequent phoneme sequences are wellformed in the large majority of the cases (e.g. Coleman & Pierrehumbert, 1997). Corpus and computational studies have shown that guidance from wellformedness can be effective for speech segmentation (Brent & Cartwright, 1996; Cairns et al., 1997; Hockema, 2006; Daland, 2009; Adriaans & Kager, 2010); in addition, listeners faced with an unknown (artificial) language use statistical cues for
segmentation (Saffran et al., 1996) and infants already induce phonological patterns from input (Saffran & Thiessen, 2003). Pitt & McQueen (1998) showed that the transitional probability between phonemes affects phoneme perception, suggesting the existence of a prelexical processor guiding the perception process with the use of probabilistic phonotactic knowledge.

In the segmentation models mentioned above (Brent & Cartwright, 1996; Cairns et al., 1997; Hockema, 2006; Daland, 2009) either boundaries are detected and all material between boundaries is regarded as a chunk, or chunks are detected and boundaries are supposed to be between those chunks. An exception is Adriaans & Kager (2010), who separately compute those biphones with likely boundaries and those that are likely to be word-internal, only combining these two types of representations in generation of segmentation hypotheses.

To sum up this overview of the possible causes of phonotactic effects on segmentation, both chunking and splitting behaviour can be caused by negative phonotactic knowledge: boundaries that cause illegal segmentations will be inhibited and illegal clusters are split to remove their illegality. Positive phonotactic knowledge can provide cues that inhibit the placement of boundaries in well-formed sound combinations, leading to chunking. However, as Experiment 3 did not feature a baseline condition, in which non-voiced but legal clusters or voiced but illegal clusters were used, it is not clear if the illegal clusters were split, the legal clusters were chunked, or the illegal segmentations with voiced codas were inhibited. Only the latter effect cannot be the effect of partial lexical activation.

Hence, it is vital to look at the legality of the segmentation as a factor in the generation of segmentations. The next experiment, Experiment 4a, uses the methodology of Experiment 3 to examine whether illegality avoidance is a driving force, by testing whether final voicing avoidance is relevant for fixations to distractors and targets, i.e. whether segmentations containing illegal voiced obstruents are avoided.

If phonotactically illegal segmentations not corresponding to words are inhibited, while segmentations that are not lexical but are phonotactically legal can be made (as shown by Experiment 3), phonotactics has to be directly involved in the segmentation process. The null hypothesis to test is thus that segmentations that differ only in phonotactic illegality, but not the number of partial lexical entries, are preferred.

Experiment 4a serves to investigate whether the voiced obstruent is a key to the difference between the conditions, by manipulating voicing and cluster well-formedness orthogonally. Voicing and cluster legality are now orthogonal cues, that predict splitting of the cluster on different occasions. If the avoidance of voicing at word ends is the driving force, voiced illegal clusters (e.g. */dl/
and /dr/) would never be split, whereas voiceless legal clusters (e.g. /tr/) can be split (if not voiced like /dr/). If voicing is the driving factor, this would provide direct evidence for illegality-driven splitting of the speech stream by phonotactic cues. Experiment 4a achieves the separation of the two factors by adding legal but not voiced clusters (e.g. /tr/) to the equation, as well as illegal but voiced ones (e.g. */dl/). Table 3.4 shows the two dimensions and examples of the clusters. If voicing is decisive and (il)legality irrelevant, voiceless clusters can always be split and /tr/ should behave like */mr/ in Experiment 3, compared to e.g. /dr/. If cluster wellformedness is decisive, words whose onset is hidden in voiceless legal clusters, e.g. /tr/, should be difficult to find, just as words hidden in voiced legal clusters, e.g. /dr/ in Experiment 3. If legality is decisive, words should always be easy to find in e.g. */dl/ and */tl/.

The extension of the set of illegal clusters also makes it possible to study the alternative explanation of phonotactic effects based on incomplete lexical recognition. In the Misalignment condition of McQueen (1998) and of Experiment 3, a cohort of words might be activated by a legal cluster (e.g. /dr/). This cohort of words possibly competes with the target. If this is so, a preference for segmentations at /d/, against one at the target position /r/ could still be driven by lexical cues. This explanation is actually already at odds with McQueen’s finding of an advantage for phonotactically aligned words embedded at the start, e.g. /wil/ (wheel) in the Misaligned [wilrɔp] vs. the Aligned [wilmrɔp]. As both [wilm] and [wilr] have empty cohorts, cohort activation cannot explain the effect directly. Nevertheless, it can be explained as an effect of wordlikeness, i.e. the fact that there are words starting with /vr/, and ending in /lm/, that could be accidentally activated and facilitate the target segmentation in the Aligned condition.

However, the comparison between illegal clusters differing in voicing makes it possible to look for evidence without word explanations; no cluster starts a full cohort, but one of them cannot be split without leaving an illegal segmentation residue.

<table>
<thead>
<tr>
<th>NoVoicedObs</th>
<th>Voiceless</th>
<th>Voiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legality</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>+</td>
<td>dr</td>
<td>tr</td>
</tr>
<tr>
<td>−</td>
<td>dl</td>
<td>tl</td>
</tr>
</tbody>
</table>

Table 3.4: Example of the range of clusters in the experimental set-up of Experiment 4a.
3.3.1 Participants

Thirty-two Dutch native speakers, students from the participant pool of the Utrecht Institute of Linguistics, participated for monetary compensation and completed the experiment. As in Experiment 3, more participants were recruited but could not complete the experiment.

3.3.2 Materials

The difference found between fixations on Misaligned and Aligned targets and distractor the Phonotactic condition in Experiment 3 can be explained by the illegal voiced obstruent coda in the residue of a target-aligned segmentation in the Misaligned trials, as well as by the contrast between cluster legality in the two levels of the Phonotactic condition. These dimensions are made orthogonal in the present experiment, by using different Cluster Pairs. These form a condition. A Cluster Pair differs in the phonological contrast expressed by the Phonotactic condition: voicing, cluster legality, or both. Hence, in a Voicing Cluster pair, the Aligned cluster differs from the Misaligned cluster on voicing. The pairs are named according to the difference between them. The Phonotactic condition thus has a different underlying meaning for the different pairs.

Sixteen clusters of two Dutch consonants were chosen, such that they formed eight pairs in which the second phoneme was held constant; this phoneme is the target phoneme, as a word will be used starting with it. The Voicing-A and Voicing-B pairs differ in voicing only, the Legality pairs differ in cluster wellformedness only and the Both pairs differ in both, with the cues converging in Both-1 and diverging in Both-2, as shown in Table 3.5. Note that within the Voicing-A and Voicing-B pairs that differ on voicing, illegality is varied between Cluster Pairs, not within, and reversely that voicing is different between Legality-1 and Legality-2, but not within them. For the cluster pairs that are legal, the cohort size is given in the table. As can be seen, the voiceless clusters that are compared with voiced clusters have a higher cohort size. If a larger cohort size causes a preference to leave the cluster intact in segmentation, the effect of voicing would be diminished, as the voiceless cluster are expected to be easier to split. A confound is thus avoided, admittedly by coincidence.

48 Dutch words were used to construct items. Six words were used for each Cluster Pair. Items consisted of an auditory nonsense string, at the end containing the target, which is the Dutch word or a close non-word (the close non-word sharing the onset with the word); the syllable before the (non-)word
Table 3.5: Cluster pairs used for comparisons. Asterisks mark illegality in Dutch as a morpheme-internal cluster. Cohort sizes are the number of Dutch phonological lemmas starting with the clusters in the CELEX dictionary.

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Type</th>
<th>Aligned</th>
<th>Misaligned</th>
<th>Contrast</th>
<th>Shared</th>
<th>Cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voicing-A1</td>
<td>tr</td>
<td>dr</td>
<td>Voicing</td>
<td>Legal</td>
<td>516</td>
<td>476</td>
</tr>
<tr>
<td>Voicing-A2</td>
<td>*tl</td>
<td>*dl</td>
<td></td>
<td>Illegal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voicing-B1</td>
<td>pl</td>
<td>bl</td>
<td>Voicing</td>
<td>Legal</td>
<td>436</td>
<td>404</td>
</tr>
<tr>
<td>Voicing-B2</td>
<td>*pw</td>
<td>*bw</td>
<td></td>
<td>Illegal</td>
<td></td>
<td>259</td>
</tr>
<tr>
<td>Both-1</td>
<td>*sw</td>
<td>zw</td>
<td>Teaming up</td>
<td></td>
<td></td>
<td>450</td>
</tr>
<tr>
<td>Both-2</td>
<td>*zl</td>
<td>sl</td>
<td>Conflicting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legality-1</td>
<td>*zr</td>
<td>br</td>
<td>Cluster</td>
<td>Voiced</td>
<td>503</td>
<td></td>
</tr>
<tr>
<td>Legality-2</td>
<td>*sr</td>
<td>pr</td>
<td>Voiceless</td>
<td></td>
<td>946</td>
<td></td>
</tr>
</tbody>
</table>

ended had as a coda the first consonant of the clusters mentioned above, the (non-)word had the second consonant as onset.

Unfortunately, the number of monosyllabic Dutch words with a monoconsonantal onset of /r/, /l/ or /w/ that could be used is low, as these words often turn into other words when embedded in clusters; e.g. /rɔk/, 'skirt', cannot be used in Voicing-A1 as /trɔk/ means 'pulled', nor Legality-1 as /brɔk/ means 'lump'. For this reason, it was not possible to use the same word in four clusters, even if this would have removed much item-related variation. Comparisons are thus only possible between the clusters within the same Cluster Pair. The experiment can therefore not compare the phonotactic cues on their strength, but investigates whether the voicing cue and the cluster cue are used in segmentation, for both levels of the other cue. The cluster cue is tested in fewer comparisons, making the experiment biased with greater sensitivity for the voicing cue than for the wellformedness cue. Analyses are still performed on the dataset as a whole, as it allows better estimates of subject random effects.

All items were constructed in four versions. Items were nested in Cluster Pairs and presented in the two levels of Phonotactics, Aligned or Misaligned, with the Cluster Pair defining the contexts. In pair Both-2, there are two diverging cues; in this case the Phonotactics condition was labelled based on illegality rather than voicing.

The other condition that each item was subject to was Lexicality, either Word or Non-word. This condition was included to follow Experiment 3 in design, as the number of words in target position might lead to learning
Table 3.6: Examples of stimuli over conditions from cluster pair Legality-2. Boldface indicates embedded words, dots indicate predicted phonotactic boundaries.

<table>
<thead>
<tr>
<th>Item /wil/ wiel</th>
<th>Phonotactics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicality</td>
<td></td>
</tr>
<tr>
<td>Aligned</td>
<td></td>
</tr>
<tr>
<td>Misaligned</td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>[zu.s.r]</td>
</tr>
<tr>
<td>Non-word</td>
<td>[zu.s.r]</td>
</tr>
<tr>
<td>Letters shown:</td>
<td>r,s,z,p</td>
</tr>
</tbody>
</table>

as targets are not alternating between words and non-words. However, the Lexicality condition was not part of a hypothesis tested in this experiment. Table 3.6 gives examples of all the versions of each item.

Like the material in Experiment 3, the auditory materials were generated with MBROLA. There were filler items containing words in the Initial position, making the position of the target 50% Initial, 50% Final as in Experiment 3; fillers with Final targets were also constructed in which the letters corresponding to the distracters in the experimental items were targets and vice versa, avoiding a bias to the target letters (l, r and w).

The items were checked for possible embedded words other than the target. For this check, the CELEX database was used. Due to an error, the word [trEm] tram ‘tramway’ was present, but not detected, in one of the auditory materials (the Aligned version for the word /rEm/ rem ‘brake’). This item was therefore omitted from all further analyses.

3.3.3 Procedure

The procedure was the same as that of Experiment 3, but the four blocks contained 48 instead of 60 trials each.

3.3.4 Results

Latency and accuracy

This experiment had a hit rate of 41%, which is lower than the 61% of Experiment 3 (Final items). There were no significant effects of experimental factors on latency, likely due to the reduced number of hits and the lower number of items. Table 3.7 shows the latencies.

Mean latency for the experimental trials was 899 ms. The accuracy (error rate) does show influence of phonotactic alignment. Nevertheless, these values
Table 3.7: Experiment 4a: Latencies. Means (ms) per combination of experimental factors, SE in parenthesis. Note that these means are not comparable, due to the use of different items in each Cluster Pair.

<table>
<thead>
<tr>
<th></th>
<th>Legality</th>
<th>Illegal</th>
<th>Legal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voicing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td>833 (31)</td>
<td>882 (45)</td>
<td>854 (26)</td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>867 (35)</td>
<td>891 (43)</td>
<td>877 (27)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>850 (24)</td>
<td>886 (31)</td>
<td>865 (19)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8: Experiment 4a: Hit rates. Probability per combination of experimental factors. Note that these numbers are not comparable, due to the use of different items in each Cluster Pair.

<table>
<thead>
<tr>
<th></th>
<th>Legality</th>
<th>Illegal</th>
<th>Legal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voicing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td>0.49</td>
<td>0.37</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>0.47</td>
<td>0.33</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.48</td>
<td>0.35</td>
<td>0.41</td>
<td></td>
</tr>
</tbody>
</table>

are not used to test any hypotheses and are only reported here to allow the reader to appreciate the data. The values shown in Table 3.9 are corrected for differences between Cluster Pairs, which are not of interest. The alignment effects by Cluster Pair were grouped together, as there were no significant effects for Phonotactics per Cluster Pair; the Pairs in which the difference was cluster wellformedness (Legal vs. Illegal) showed a significant effect of Phonotactics, but when the difference was based on voicing (Voiced vs. Voiceless), there was no significant difference, as shown in Table 3.9. Although the effect of voicing does not reach significance, it is interesting that this effect is only present for Illegal clusters: the base level of the model is Voiceless and Illegal. Therefore the Voicing effect in Table 3.9 is not a main effect but the effect for Illegal clusters; the interaction shows that this effect completely evaporates when the Voicing occurs in Legal clusters. So, if clusters are illegal, they are easy to split and then Voicing might matter, but if they are legal, they are not split at all, which makes Voicing irrelevant. This suggest that Cluster legality is the primary phonotactic cue for segmentation, although this could not be tested in a controlled way with this experiment.

For illustrative purposes, Figure 3.3 shows the (non-significant) differences in accuracy per Cluster Pair; the pairs Voicing-A1 (based on Voicing) and Both-1 (with converging cues) do not seem to follow the predictions. Note that these have the lowest respectively highest overall accuracy, which could mean
that effects of Phonotactics were lost in ceiling and floor effects. Voicing-A\textsubscript{1} and Both\textsubscript{-1} also showed very small differences, just as Voicing-B\textsubscript{1} and Voicing-B\textsubscript{2}, that were also based on voicing but do follow the predicted pattern. The Legality pairs show large differences irrespective of voicing (Legality\textsubscript{-1} is voiced, Legality\textsubscript{-2} voiceless). Note again that these differences are not tested as hypotheses.

The considerable accuracy differences suggest a speed/accuracy trade-off, as the context of illegal clusters, supposedly more difficult to split, words are more difficult to find. They might be so hard to find that only the most salient words are found; therefore the latency differences, which are only collected for correct responses, are subject to both the increase caused by the added difficulty and the decrease due to the elimination of the most difficult items. Just as in Experiment 3, fixations can be expected to show a much sharper image as they are also gathered in the Non-word condition and when no word is recognised.

**Fixations**

The hypothesis under scrutiny is that fixations to targets and distractors follow the Phonotactic alignment differences within Cluster Pairs. This amounts to the interaction of Phonotactics $\times$ Cluster Pair. A fixed effect for Cluster Pair was added to correct the estimates for the difference between the sets of words used in the different pairs, but this effect is not of interest to the hypotheses under scrutiny. As Experiment 3 showed that Lexicality of the target influences the fixation behaviour, a main effect of Lexicality was also included.

As in Experiment 3, two analyses were performed, on target and on distractor fixations. Both were logistic regressions on the probability of fixation. The analyses can be assumed to be dependent, because the prelexical account turned out to be correct; in order to be conservative the significance threshold was again (Bonferroni) corrected to 0.025. Another correction was needed, as

<table>
<thead>
<tr>
<th>Factor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Cluster</td>
<td>-1.082 (0.297)</td>
<td>-3.639</td>
<td>0.000274 ***</td>
</tr>
<tr>
<td>Voiced Obstruent</td>
<td>-0.401 (0.258)</td>
<td>-1.551</td>
<td>0.120827</td>
</tr>
<tr>
<td>Legal $\times$ Voiced</td>
<td>0.441 (0.368)</td>
<td>1.201</td>
<td>0.229911</td>
</tr>
</tbody>
</table>

Table 3.9: Accuracy for Native Dutch listeners. A mixed model with crossed random effects for participant and item was fitted to the logit of spotting the word, also taking cluster pair and word frequency into account.
unlike in Experiment 3, the analysis was not just to show that the data are modelled better when a phonotactic and lexical term were introduced at the temporal terms. Instead, the best-fitting model was calculated from models containing Cluster pair and Phonotactics interaction terms, varying only in the number of time levels considered. The best model had three time levels and the effects of the Cluster Pair and Phonotactics might have surfaced at each of them; there was no a priori prediction of the exact shape of the effect. The significance threshold was thus lowered again, by dividing it by the number of time levels. This correction can be assumed to be overly conservative, as the effects at the different time levels are neither completely independent, nor completely dependent: an effect might occur at only one level, but is not likely to, as a higher overall probability (intercept) can reasonably be argued to spur a quicker rise of the probability. However, it is most conservative to use the correction proposed here, at the cost of severely reducing power. As only positive test results are considered, this is not detrimental to the analysis.

Fixations to targets in a pilot showed a rising pattern starting at 100 ms after word offset. After 904 ms more than 95% of trials had ended in this experiment and in Experiment 4b. For reasons of comparability the same time window was used in the two experiments. The analysis was therefore restricted to the window 100–904 ms.

Adding the three levels of time (without experimental factors) significantly improved the target fixation model and adding the Cluster Pair and
Phonotactics terms to each time level also lead to significant improvements ($\chi^2$ values were 2075, 655, 844, 254, each with 16 d.f. and $p < 0.0001$). The significance threshold was therefore set to 0.025/4 = 0.00625 ($p < 0.00625$: *, $p < 0.00125$: **, $p < 0.000125$: ***).

Table A.7 in the appendix shows the model for target fixations. Effects of Phonotactics were found for all Clusters Pairs, with the exception of the Voicing-B1 pair (pl-bl). Note that the effect on the linear time component for Voicing-B1 is only not found because of the high corrected threshold and might thus be a false negative. For the other Voicing-B-type Cluster Pair, Voicing-B2, with illegal clusters, as well as the Voicing-A-types, there were effects of Phonotactics, which for these pairs is instantiated by voicing differences. The wellformedness differences of the Legality-pairs and the converging and diverging cues of voicing and wellformedness in the Both-type pairs also show effects; the conflict between the two cues in Both-2 turns out to be decided in favour of cluster wellformedness. The graphs in Figures 3.4, 3.5, 3.6 and 3.7 give a more insightful summary of the data. Triangles refer to the supposedly easier (Aligned) situations for finding the word, and black symbols refer to the Word level of the Lexicality condition, while circles refer to the Misaligned situation and open symbols refer to Non-words. In most graphs, black triangles, i.e. Aligned Word targets, are fixated on quickly with a higher probability. Note again that the modelled lines are conservative due to corrections for shrinkage and thus closer to the mean; however, they also start low (close to zero) due to the restrictions of the time structure. This explains the apparent failure of the model to extend to the peaks.

Effects on distractor fixations were found for all Cluster Pairs and correspond inversely to the patterns found in target fixations. There are more fixations to distractors that correspond to the start of a legal chunk (Misaligned in Legality and in Both-2, where it is a stronger force than the opposed voicing cue; note that distractors in the Misaligned case are actually aligned with phonotactic boundaries). In the Cluster Pairs Voicing-A and Voicing-B that differ in voicing, the voiced clusters are more likely to be distracting from the target segmentation than their voiceless counterparts. The graphs in Figure 3.8 and 3.9 show the fixation pattern to the distractors.

3.3.5 Discussion

The aim of Experiment 4a was to pin down the phonotactic segmentation effects found in Experiment 3. If those effects had been due to voicing differences, this could only be explained by illegality avoidance, showing quite directly that negative phonotactic knowledge affects segmentation. The presence of voicing effects show that illegal voiced codas are indeed avoided,
Figure 3.4. Experiment 4a, Target fixation probabilities for Cluster Pairs Voicing-A. The plotted points are proportions (aggregated over trials and 40ms bins). The solid line at 0.25 represents chance level.
Figure 3.5: Experiment 4a, Target fixation probabilities for Cluster Pair Voicing-B. The plotted points are proportions (aggregated over trials and 40 ms bins). The solid line at 0.25 represents chance level.
Figure 3.6: Experiment 4a, Target fixation probabilities for Cluster Pair Both. The plotted points are proportions (aggregated over trials and 40 ms bins). The solid line at 0.25 represents chance level.
Figure 3.7: Experiment 4a, Target fixation probabilities for Cluster Pair Legality. The plotted points are proportions (aggregated over trials and 40 ms bins). The solid line at 0.25 represents chance level.
Figure 3.8: Experiment 4a, Distractor fixation probabilities for Cluster Pair Voicing-A–Voicing-B. The plotted points are proportions (aggregated over trials and 40 ms bins). The solid line at 0.5 represents chance level. Circles represent distractors hypothesised to cause misalignment for the target, triangles represent those hypothesised to cause alignment. Open symbols refer to the Non-word level of Lexicality, closed ones to the Word level.
Figure 3.9: Experiment 4a, Distractor fixation probabilities for Cluster Pair Both–Cluster. The plotted points are proportions (aggregated over trials). The solid line at 0.5 represents chance level.
making segmentations that contain such illegal residues less distracting for the segmentations aligned with the targets. This effect of voicing, found with the Cluster Pairs Voicing-A and B (except for Voicing-B1 target fixations), must be an effect at the level of the segmentation hypothesis needed to find the word, as the clusters that are compared are either both wellformed (Voicing-A1, Voicing-B1) or both illformed (Voicing-A2, Voicing-B2). Cohort sizes hardly differ and only in the opposite direction, as mentioned above: the clusters /tr/ and /pl/ have fuller cohorts than the clusters /dr/ and /br/ and the latter should thus have been easier to split if cohorts cause activation of alternative segmentations.

Only when segmentation at the target onset is performed does illegality arise; it is not present in the speech stream before segmentation. In the Legality pairs this is not the case; the illegal clusters are compared to legal clusters and the legal clusters attract more attention, while the target attracts more attention when it is embedded in an illegal cluster; the illegality is already present in the speech stream. Wellformedness differences were nevertheless also convincingly influential on the segmentation.

Two origins of the phonotactic effect on segmentation can be identified. First, segmentations that are phonotactically illegal are not preferred. Second, the more wellformed a cluster, the less likely it is to be considered to have occurred over a word boundary, or, reversely, the more illformed a cluster, the more likely it is to be split. The effect of wellformedness can be explained at the level of the (unsegmented) speech stream, in contrast to the illegality effect of voiced obstruent codas. However, the illegality and/or wellformedness in the speech stream is also present at the segmentation level; it is thus most parsimonious to assume that segmentations, not clusters, are assessed on wellformedness.

Distractor fixations are in line with target fixations: more distraction is caused by wellformed clusters, but also by clusters starting with voiced obstruents when compared to voiceless obstruents. This convergence of the voicing and cluster wellformedness effects again suggests that segmentation wellformedness is assessed on a global level, in a comparison of different possibilities. In Cluster Pairs Voicing-A1 and Voicing-B1, both clusters are legal and in Voicing-A2 and B2 both are illegal, but voicing matters for the likelihood of the distractor to be considered a word onset, which only makes sense if the rest of the segmentation is taken into account, as the residue contains the legal or illegal residue with the first consonant of the cluster as a coda.

In summary, Experiment 4a has shown that one of the driving forces for prelexical segmentation is the avoidance of illegal segmentations. In addition, cluster legality and/or wellformedness plays a role. The Phonotactics effect
found for the C1 Cluster pair, where both cues team up, can be explained by both voicing and cluster wellformedness. However, when the two cue types conflict (in pair C2), the cluster legality difference is dominant.

The effect of phonotactics on segmentation has to be attributed to actual phonotactic knowledge. Lexical information cannot completely explain the segmentation effect. As Experiment 3 suggests that segmentation hypotheses are made for non-lexical boundaries, segmentation seems to be a process that runs regardless of the presence of lexical cues. Of course, this does not mean that lexical information is not also used to segment speech.

3.4 Experiment 4B

The results of Experiment 4a suggest that knowledge of phonotactic illegality drives segmentation. The possibility that wellformedness also plays a role is still open. If wellformedness differences are the origin of the effect found in the Both-2 and Legality Cluster pairs, an account of the usage of phonotactic knowledge in the generation of segmentation hypotheses has to be based on the assumption that targets (both words and non-words) that start in the middle of a legal cluster are harder to segment, because listeners tend to hypothesise the whole cluster to be an onset, while they do not consider an illegal cluster to be one.

Alternatively, a reverse explanation based on avoiding illegal segmentations can also explain the difference between the illegal and legal clusters. The wellformed clusters do not contain a likely boundary, hence only the recognition of the target word, a lexical cue, points to the segmentation that is in accordance with the target word. The target word is thus not aided at all by prelexical information. The illegal clusters, on the other hand, can directly be considered as straddling a word boundary and thus a word onset in an illegal cluster is already found prelexically.

The two explanations, illegality-driven splitting and wellformedness-driven chunking of clusters, do not differ in their predictions for the current experiment. However, if the same result is obtained in a situation where the illegal clusters are not illegal, while the legal clusters are still wellformed, there would be evidence for the use of knowledge of wellformedness for speech segmentation. The illegal clusters are not illegal in all languages. Hence, the experiment can be repeated with listeners with phonotactic knowledge of another language.

Given that the results of Experiment 4a constitute a ‘normal’, i.e. native-like level of fixation to a set of legal and illegal clusters, these can be used to contrast native listeners with L2 listeners. The L2 listeners should have a native language that allows them to transfer cluster wellformedness, but
not cluster illegality knowledge. Such listeners should, if wellformedness knowledge is a separate source for segmentation, prefer to chunk the legal clusters. However, if only illegality knowledge plays a role, the L2 listeners are predicted to fail to disregard illegal clusters as word onsets and benefit less from the Phonotactic Alignment.

The next null hypothesis is that wellformedness knowledge is not used as a cue for word segmentation. Hence, the difference between legal and illegal clusters is caused only by the illegality of the illegal clusters. Listeners that transfer phonotactic knowledge from their native language to a second language can thus only transfer illegality knowledge; the null hypothesis predicts no transfer of wellformedness knowledge. Hence, if there is no illegality knowledge to transfer from first to second language, second language listeners cannot use this knowledge unless they acquire it in the second language. If second language listeners are less sensitive to the phonotactic alignment due to illegality, they have not acquired illegality yet. However, if they consider the cluster onsets to be more likely onsets than native listeners, this is chunking behaviour, which can only be explained as the usage of wellformedness knowledge, not illegality knowledge.

Experiment 4b is set up to test the null hypothesis that there is no transfer of wellformedness knowledge, as a mirror image to the results of Weber & Cutler (2006), that show transfer of illegality knowledge. Wellformedness-driven segmentation predicts that legal clusters of the form C1C2 could attract more attention to C1 as an onset than illegal clusters would, to the detriment of attention to C2; these segmentation attempts can be detected by eye-tracking. The L2 participants had a native language with a set of permissible onsets that is a superset of the L1. They would therefore never transfer cluster illegality. If they show better performance at words that are aligned according to the Target Language’s phonotactics, they have either learnt L2 phonotactics, or applied transferred L1 gradient phonotactic knowledge of wellformedness; some C1C2 clusters might be more wellformed and attract more attention to the C1 than others.

If wellformedness knowledge has no effect, there is no difference in attention to clusters that are not illegal; the L2 participants should therefore not pay more attention to Target Language illegal, but Source Language legal, clusters. However, if the acquisition problem is that L2 learners have to acquire the illegality of some clusters, they might not distinguish illegal from legal clusters and not benefit from the illegality of clusters when looking for words that start within those clusters.
3.4.1 Participants

Twenty-three non-native speakers of Dutch participated for monetary compensation. They were recruited through word-of-mouth, web communities and flyers at cultural centers and shops targeted at Slavic immigrants. The first languages are Russian (16), Polish (2), Bulgarian (4) and Czech (1). Three of the Russian language participants also had Ukrainian as a mother language. All of these Slavic languages allowed most clusters in this experiment at word onsets, but featured final obstruent devoicing like Dutch. Nevertheless, these languages have no /w/ phoneme and therefore the cluster pairs Voicing-B2 and Both-1, ending in /w/, are not allowed. Therefore, these clusters were not included in the analysis.

As most of the recruited participants spoke Russian, the word-initial biconsonantal clusters of the Uppsala corpus of Russian (Lönngren, 1993) were assessed to gather some insight into the gradient wellformedness of the clusters in this language. The probabilities of each consonant were calculated for first and second position, after which the expected frequency of each sequence was calculated. This expected frequency was then used to calculate the observed/expected ratio, with observed referring to the frequency of each cluster type in the same set of word-initial onsets. If clusters were the result of random combination of phonemes, the O/E ratio would always be 1. A lower value indicates underrepresentation, a higher value overrepresentation (Adriaans & Kager, 2010).

The corpus reveals that Dutch illegal clusters are indeed legal in Russian, though they tend to be less wellformed. Table 3.10 (p. 136) shows the frequencies of the clusters. Note that the Russian wellformedness of /zl/ and /dl/ is quite close to 1, the neutral value.

3.4.2 Materials

The same materials were used as in Experiment 4a. However, as mentioned above, less of the material was used as experimental item; the remainder was used as fillers. The materials were now not compared in cluster pairs; instead, the clusters were ranked on Legal vs. Illegal. These levels differ in more than just the legality, because the sets of words are different. Nevertheless, the fixations are compared to a base level, set by the Dutch participants in Experiment 4a, which means that the comparison is made in interaction with Native Language and that analyses are performed on the combined data of Experiment 4a and Experiment 4b. Therefore, no effects that are not interactions with Native Language are interpreted, as such effects are not properly corrected for the different items in the conditions.
### Table 3.10: Cluster pairs’ O/E ratios in the Uppsala corpus. Asterisks mark illegality in Dutch as a word-internal cluster. In Russian, the Dutch legal clusters are more wellformed than the Dutch illegal clusters as well, but the Dutch illegal ones are not illegal in Russian. As Russian has no /w/ phoneme, pairs Voicing-B2 and Both-1 were removed.

#### 3.4.3 Procedure

The procedure of Experiment 4b was the same as for Experiment 4a.

#### 3.4.4 Results

**Latency and accuracy**

The latency and accuracy were again only analysed to allow more insight into the data. Tables 3.11 and 3.12 show the behaviour of the L2 Dutch participants with a Slavic L1 and Table 3.13 shows a model of the data, including significance values. There seems to be another speed/accuracy trade-off present in these results: in the context of legal clusters, supposedly harder to split, words are less likely to be spotted, but if they are, they are spotted faster. Just like in Experiment 4a, the listeners might not need a lot of need phonotactic help to find the really salient words, and on the other hand fail to find the misaligned words at all, thereby causing the latency data to incorrectly show an advantage for misaligned words as the reaction time was only recorded for very salient words, whereas the aligned words were also spotted when they were less salient.

**Fixations**

Target fixations were analysed in the same way as in Experiment 4a. Again intercept, linear, quadratic and cubic terms for the experimental factors im-
### Table 3.11: Experiment 4b: Latencies. Means (ms) per combination of experimental factors, SE in parenthesis. Note that these means are not comparable, due to the use of different items in each Cluster Pair.

<table>
<thead>
<tr>
<th>Voicing</th>
<th>Legality</th>
<th>Illegal</th>
<th>Legal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiced</td>
<td>949 (61)</td>
<td>1031 (92)</td>
<td>983 (52)</td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>1039 (85)</td>
<td>870 (78)</td>
<td>953 (58)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>986 (50)</td>
<td>951 (61)</td>
<td>970 (39)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.12: Experiment 4b: Hit rates. Probability per combination of experimental factors. Note that these numbers are not comparable, due to the use of different items in each Cluster Pair.

<table>
<thead>
<tr>
<th>Voicing</th>
<th>Legality</th>
<th>Illegal</th>
<th>Legal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiced</td>
<td>0.42</td>
<td>0.23</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>0.29</td>
<td>0.22</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.35</td>
<td>0.22</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.13: Modelled latency and accuracy. A mixed model with crossed random effects was fitted to the log-transformed reaction times measured from word offset; for accuracy a model was fit on the logit of spotting the word.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Latency (SE)</th>
<th>t-value (19 d.f.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Cluster</td>
<td>-0.3369 (0.1771)</td>
<td>-1.902</td>
<td>0.0689</td>
</tr>
<tr>
<td>Voiced Obstruent</td>
<td>-0.0859 (0.1648)</td>
<td>-0.521</td>
<td>0.3417</td>
</tr>
<tr>
<td>Legal × Voiced</td>
<td>0.2716 (0.2360)</td>
<td>1.151</td>
<td>0.2007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Accuracy (SD)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Cluster</td>
<td>-0.8537 (0.4055)</td>
<td>-2.105</td>
<td>0.0352 *</td>
</tr>
<tr>
<td>Voiced Obstruent</td>
<td>-0.2046 (0.3970)</td>
<td>-0.515</td>
<td>0.6064</td>
</tr>
<tr>
<td>Legal × Voiced</td>
<td>-0.0576 (0.5305)</td>
<td>0.109</td>
<td>0.9135</td>
</tr>
</tbody>
</table>
prove the model significantly. Therefore, the significance threshold is again corrected to $0.025/4 = 0.00625$ ($p < 0.00625$: *, $p < 0.00125$: **, $p < 0.000125$: ***).

For fixations to targets, the difference between L1 and L2 Dutch listeners is significant in the linear term for Non-words; the difference between legal and illegal clusters is smaller for L1 Slavic L2 Dutch listeners, compared to L1 Dutch listeners. For Words, the difference is attenuated for the linear term, but present in the quadratic and cubic terms. The graphs in Figure 3.10 show this in the most accessible way; Table A.9 shows the estimates and their significance.

For fixations to distractors, the difference between L1 Slavic and L1 Dutch listeners is not as easy to interpret straightforwardly. L1 Slavic L2 Dutch listeners are less susceptible to cluster wellformedness of distractors, but not if a Word is present. Note in Table A.10 that the constant effect of L1-Slavic $\times$ Cluster-Legal (for the Non-word base level of Lexicality) is curiously similar in size as the same effect at the Word level, but in the opposite direction (0.4814 vs. -0.4749); this neatness of the result should of course not lead to overly confident interpretation. The two graphs in Figure 3.11 show how the other time levels also influence the lines; the data points do follow the general pattern. Illegal clusters are more distracting to L2 listeners in the Non-word trials, in comparison to the native listeners; this corresponds to the finding that L2 listeners are less likely to fixate at Aligned targets when these targets are not words.

3.4.5 Discussion

L2 listeners’ knowledge of Dutch phonotactics can originate either from acquisition or from transfer from the L1, as pointed out by Weber & Cutler (2006). Importantly, the L1s in the present experiment could not have provided these listeners with illegality knowledge. The result that L2 listeners of Experiment 4b showed more fixations on the Dutch illegal clusters in the Non-word condition than native Dutch listeners makes it unlikely that the L2 listeners acquired categorical Dutch illegality knowledge. However, in the Word condition, listeners with Dutch as an L1 and with Dutch as an L2 all benefit from the phonotactic alignment provided by cluster illformedness; they spot the word with greater ease.

Hence, gradient wellformedness differences seem to make the L2 listeners correctly decide on the word-aligned segmentation as most wellformed in the Aligned Word condition. As a word is actually present, lexical cues guide them in the same direction, making other segmentation hypotheses
Figure 3.10: Experiment 4b: target fixation probabilities for Words and Non-words. The plotted points are proportions (aggregated over trials and over 40 ms bins) and the lines represent the modelled trajectories. The solid line at 0.25 represents chance level. Circles represent illegal clusters, triangles and dotted lines represent legal ones. The open symbols and the dashed and dot-dashed lines represent native listeners.
Figure 3.11: Experiment 4b: distractor fixation probabilities for Words and Non-words. The proportion of fixations on the distractor present in the auditory stimulus is given, over all fixations on one of the two distractors. The plotted points are proportions (aggregated over trials and over 40 ms bins) and the lines represent the modelled trajectories. The solid line at 0.5 represents chance level. Circles represent illegal clusters, triangles and dotted lines represent legal ones. The open symbols and the dashed and dot-dashed lines represent native listeners.
quite irrelevant, even if these alternatives are incorrectly assumed to be ‘not wellformed’, instead of illegal.

However, when in the Non-word condition no word exists at the target position to provide lexical cues, segmentation hypotheses that are less well-formed are also considered; this explains why the L2 listeners in the Aligned Non-word condition fixate more on the illegal cluster than the native Dutch listeners: the L2 listeners consider the alternatives wellformed, while the Dutch participants seem to consider them illegal. However, the low fixation probability of Dutch participants for the distractor in the Aligned Non-word condition does not entail complete absence of segmentation hypotheses that keep the illegal cluster intact; it is also possible that Dutch listeners consider these hypotheses even more illformed than the L2 listeners do, but not illegal. The methodology does not make it possible to examine if there is such a difference.

Still, the result that L2 listeners benefit from alignment caused by cluster wellformedness in the Word condition, like the Dutch listeners do, indicates gradient wellformedness knowledge that is similar to Dutch categorical legality. Evidence that Slavic L2 learners of Dutch acquire wellformedness was found by Trapman & Kager (2009). In addition, the corpus data reported above indicates that Dutch illegality/legality contrasts are often correlated with Russian wellformedness. This means that gradient wellformedness could both be transferred from the first (Slavic) languages and that it also could have been acquired. The latter is more likely, as Trapman & Kager (2009) reported that more advanced learners have more knowledge of Dutch phonotactic wellformedness and as Weber & Cutler (2006) showed that phonotactic knowledge can be acquired for a second language, but it does not matter for the present discussion which one of the two is correct.

The important result of Experiment 4b is that L2 listeners have not acquired illegality. It is possible that the native listeners do not have a separate type of representation for cluster illegality, but just a much lower wellformedness value for the ‘illegal’ clusters than the L2 listeners. This suggests that gradient phonotactic wellformedness knowledge is used in segmentation.

If a cluster is legal, segmentations are more likely to chunk the whole cluster. However, for both the native and L2 listeners, fixations also go to the target; possibly because no word starts at the cluster and hence lexical cues start to favour the target as a boundary. Lexically-driven segmentation is likely to be slower and therefore a wellformed cluster, i.e. the presence of phonotactic cues against the word onset, slows down recognition of words that start within wellformed clusters. Illegality seems to help the native listeners to find the target quickly, whether or not lexical cues are present. This is not the case for the Slavic L1 listeners of Experiment 4b, who consider the illegal
cluster as a possible onset, i.e. keep entertaining segmentations with these clusters intact.

3.5 GENERAL DISCUSSION

The eye-tracking methodology introduced for the present study and explained in this article is a sensitive tool to assess segmentation as it unfolds, instead of at its end stage of lexical recognition. After conservative statistical analysis, it has been shown to yield a large power to test hypotheses. This means that this methodology can be used with smaller datasets, that are often needed due to the sparsity of (lexical) items fulfilling a large number of requirements. In addition, Experiment 4a and Experiment 4b showed that data gathered on fixations show significant differences that are lost in latency and/or accuracy data, most likely due to speed/accuracy trade-offs.

The results indicates that speech can be segmented independently of the lexical correctness of the segmentations: even if no words are present in the speech stream, segmentation hypotheses are entertained. In addition, preferences for segmentation hypotheses can be based on phonotactic cues only. It is likely that other sublexical cues operate in a similar way. Lexical cues, provided by words that are recognised in the speech stream, also have an important effect on segmentation, but are not necessary for segmentation to occur. This was shown in Experiment 3 by the effect of the Phonotactics condition in the Non-word trials and by the lower accuracy of word spotting in the Misaligned Word trials, but more accurately in Experiment 4a, when participants avoided segmentations that left voiced obstruents in codas.

In light of the results, the conclusion that there is already speech segmentation before words are identified can be made. The most important characteristic of prelexical segmentation is that it is not generated by words and then assessed on sublexical properties, but that instead, sublexical properties generate or evaluate segmentation hypotheses. The term ‘prelexical’ is not so much meant to denote a process completely before lexical segmentation, but as a process that is not only driven by cues derived from lexical knowledge. Even though the time course shown in the graphs of fixation proportions suggests that the advantage of phonotactically aligned segmentations starts before the target is recognised, this can only be concluded tentatively: the fact that the target word has not yet been spotted does not necessarily imply that the word is not already influencing the segmentation. Lexical cues might already be active even if words have not reached the threshold of recognition.

The data do not directly justify the conclusion that phonotactics operates before lexical cues in segmentation. Nevertheless, this view is theoretically the most reasonable one, as phonotactics can help find words, the very purpose
of word segmentation. The reverse, recognised words asking for phonotactic approval, makes less sense: if words are recognised, this normally implies that a phonotactically sound segmentation is made, as words are by definition phonotactically legal.

Under the assumption that speech segmentation is directed towards word recognition, it is not surprising that the low-level effects of phonotactics appear most clearly either before words are recognised or when no words can be recognised. Nevertheless, as soon as words are found, their boundaries are also used to segment the speech stream. The findings can thus be interpreted along the lines of Mattys et al. (2005), who compared the importance of different segmentation cues by pitting them against each other. They concluded that there is a hierarchy of segmentation constraints, with lexical cues being most important when available and phonotactics occupying the next highest level of importance. Nevertheless, Mattys et al. also looked at lexical recognition latency and accuracy, that are measured at the end of the segmentation process, which emphasises lexical cues. It seems that phonotactic cues that are not backed up by lexical viability are dropped in order to use less wellformed segmentations that do lead to word recognition; when measuring at the lexical end of the process, the phonotactics-only segmentation attempts will not be detected, but only the delay they cause to the lexical attempts by competing with them.

If a target is a word, as well as phonotactically aligned, all information converges on it to be recognised; if no word is present while the target is aligned, phonotactics signals the target just as well but on its own. In fact, if no word is present, there are only phonotactic cues to word boundaries, hence the phonotactic effect is most visible in such a case; because in the misaligned non-word case, only the distractor is supported, namely by phonotactics, while segmenting at the target is not supported by any information. Only when phonotactics conflicts with words, which should be rare in natural speech, do the two segmentation hypotheses have to struggle; the word target has to share its attention with the distracting segmentation(s). This explains the phonotactic effect on word recognition, but it also explains how word presence is not essential for phonotactic cues to have an effect.

This explanation diverges from that given by McQueen (1998, p. 40), who states that “any candidate word in the competition process has its activation reduced when the stretch of speech between the edge of that word and a likely word boundary location is not a possible word”, thus locating the phonotactic effects at the level of words. A string like /fi.draɪk/ has a likely word boundary between /d/ and /r/, according to McQueen. Segmenting the word then leads to /fi.draɪk/, leaving /d/ which violates the Possible Word Constraint (PWC; Norris et al., 1997) as it contains no vowel. However, this lexicalist
account needs to compute the likely word boundary before recognition of words. For illegal clusters, this can be done locally, but phonotactics does not signal a word boundary before voiced obstruents like /d/ as (e.g.) /id/ is legal within words, e.g. /ide/, ‘idea’. Only the fact that */fid/* is not legal forces a boundary before /d/, but as /fi/ is not a word, the preference for the segmentation /fi.drOk/ can only be computed by performing the complete segmentation, partially using lexical cues, partially using phonotactic cues.

Even if lexical cues are not directly necessary for segmentations and might be only taken into account after sublexical knowledge yields segmentation hypotheses, it is possible that phonotactics does not just mark boundaries or chunks. Experiment 4a suggested that segmentations are primarily influenced by (gradient) cluster wellformedness and that final devoicing is not always taken into account. Interestingly, the voicing effect found in Experiment 4a is present in illegal clusters, which suggests that it is not the wellformedness of the cluster (low in both cases in a comparison) that makes it hard to split voiced clusters, but the illegal properties of a segmentation hypothesis that is generated by the cluster illegality and possibly also by lexical cues.

These results suggest that both (local) cluster wellformedness and (global) segmentation illegality are important factors for phonotactics-driven segmentation and that they can conflict at the level of the whole segmentation. It would otherwise have been theoretically parsimonious to assume that cluster wellformedness leads to a local decision to split or chunk a cluster. It seems that a more global evaluation of segmentations takes place.

The knowledge used in such a global evaluation could be of any kind, both constraints against certain illegal features or constraints favouring certain wellformed combinations. The effects observed in the present study are mostly related to cluster wellformedness, which could be based on legality, illegality, or even both. Experiment 4a could not determine the primacy of either illegality or legality, but Experiment 4b did show the use of gradient wellformedness, which implies at least a positive representation of the wellformedness of all clusters. The aforementioned experiments by Weber & Cutler (2006) suggested that illegality knowledge is also used by second language (L2) listeners, as they showed transfer of illegality knowledge.

Still, one has to be careful not to jump to conclusions on the basis of the experiment of Weber & Cutler (2006) alone. It is important to note that the experiment was not designed to test the difference between legality and illegality (it was designed to test the acquisition of phonotactic knowledge). It can thus be taken as exploratory for the present purpose, but not confirmatory of the existence of illegality, only of the existence of transfer of phonotactic knowledge and of acquisition of second language phonotactics.
There is an alternative explanation for the fact that illegality knowledge seems to be transferred. It could also be that L2 learners have flawed lexical representations and therefore no representations of words in the Target Language starting with clusters illegal in their native language. E.g., Germans might not represent words like /slip/ 'sleep' faithfully, as their native language constraints might cause them to filter out illegal clusters like /sl/ and perceive these as the nearest legal ones, probably [ʃ]; for this reason they might represent ‘sleep’ as /ʃlip/. This might make them more likely to consider the alternative /lip/, as they do not recognise /sl/ as the onset of any English words. For instance, in the actual experimental item [daʃrlid³an], containing the word /lid³an/, ‘legion’, an English native listener might start entertaining the word onset /sl i . . ./ as many words (e.g. /slip/ ‘sleep’) start with these three phonemes, while a German listener might not ‘suffer’ from this interference because she has represented those words with a /ʃli/ onset (e.g. /ʃlip/).

This kind of lexical influence might affect the outcome at the lexical end-stage of the segmentation process, which is where Weber & Cutler, following the McQueen (1998) paradigm, gathered their data. There is ample evidence for flawed lexical representations in L2 learners. Pallier et al. (2001) showed that Spanish-Catalan bilinguals that are highly proficient in Catalan, but dominant in Spanish, ignore the Catalan vowel contrast /E/-/e/. Broersma & Cutler (2008) showed that Dutch listeners do not use the English contrast of voicing on word-final obstruents. Therefore, the transfer of phonotactic illegality as found by Weber & Cutler (2006) could have been an artefact of L2 lexical representations, instead of L2 segmentation.

As the paradigm introduced in this article allowed a direct look at prelexical segmentation, the results of Experiment 4b can be attributed to phonotactics without confounds with lexical factors. In addition, the L2 learners all had an L1 in which all clusters used in the experiment are legal. This makes it unlikely that misperceptions or flawed lexical representations play a role. In addition, where illegality is the most likely form of transferred knowledge in the Weber & Cutler (2006) study, the L2 listeners in the present study cannot transfer any illegality knowledge to Dutch from their native language. Therefore, Experiment 4b suggests that there is both use of knowledge of illegality and of knowledge from wellformedness, but in different ways. Nevertheless, such a conclusion can only be made tentatively. The native listeners still made some fixations on the illegal clusters and no principled grounds are available to decide that these fixations are only due to noise. Nevertheless, this study suggests that both representations of wellformedness and categorical illegality of phonotactic knowledge are active in the prelexical segmentation stage.
The independent status of phonotactic knowledge in word segmentation calls for a model of its machinery. Such a model would not only serve to capture the observations made here, but also partially provide a computational solution to the tasks facing listeners listening to speech. Listeners have to look up many words and if no sublexical information was available to guide them, they would have to make numerous lexical look-up attempts. Using sublexical cues, such as phonotactics, saves them from the unnecessary ones.

A model for segmentation cannot be proposed based on only the data from the present experiments and previous findings, as many details cannot be filled in. However, the model should fulfill a number of requirements. It has to take different representations of phonotactic knowledge into account in an evaluation that precedes lexical look-up. Phonotactic knowledge is used to form bottom-up segmentation hypotheses containing likely candidate words, inciting lexical look-up at the adequate places. The process crucially does not hinge on the lexical status of the chunks, but on the suspicion that a chunk is likely to be a word. Lexical knowledge also has an independent role, because it can provide confirmation for segmentation hypotheses. Incorporating phonotactic chunking in lexicalist segmentation models is possible and for the Shortlist B model, this was already envisaged, although not implemented (Norris & McQueen, 2008, p. 363, 377). However, the suggested role of phonotactics in Shortlist B is the role of raising the probability to recognise a word when it occurs in a well-formed segmentation path.

The results of the present experiments suggest that phonotactics influences the probability of a segmentation hypothesis to be considered, instead of the probability of words. That is not to say that word competition is not important in segmentation, but phonotactic knowledge is relatively independent from it; words get activated easier if they are already found by prelexically likely segmentations, but still have to compete with other words in case there is overlapping lexical material.

Hence, a parsimonious model of prelexical segmentation has to assume a general measure of segmentation wellformedness, taking into account at least the factors for which evidence was found in the present study and most likely also other sublexical cues. The most wellformed (or least illegal) segmentations are likely to be used first, but the fact that participants are not completely unable to find misaligned words suggests that less optimal segmentations are also taken into account. However, this is less likely to happen, probably because it occurs only after the more wellformed segmentations have failed to yield a recognised word.

A general evaluation mechanism for phonotactics-driven segmentation has been proposed by Kager (2010). This mechanism takes the entire string and all phonotactic knowledge into account and calculates the wellformedness
of segmentations. The string [CVdlVC] might be segmented at either /d.l/ or at /.dl/, but one segmentation might be more wellformed than another, even if both violate a phonotactic constraint. In other cases, e.g. for the string CVtrVC, the segmentations CV.trVC and CV.trVC are both legal, but one might again be more wellformed than the other and both are more wellformed than the segmentations of [lll].

The Kager (2010) account formalises the evaluation of segmentations using one of the most common frameworks in phonology to address conflicting information: Optimality Theory (OT; Prince & Smolensky, 1993/2002). In OT, candidates are evaluated based on their violations of constraints; the constraints are ranked and violations of the highest ranked constraints are most important; they eliminate candidates. In the case of segmentation, input would be an unsegmented string and output candidates would be the (logically) possible segmentations. For instance, for the input [fisrOk], candidates are /fi.srOk/, /fis.rOk/, /fisr.Ok/, etc. The candidates that contain the illegal string */sr/ violate the constraint */sr/. If this constraint is high-ranked, these candidates are eliminated. After that, the next constraint is evaluated, until only one candidate remains. This candidate is optimal, because it violates less important constraints than any other option.

The standard assumption in OT is that the optimal candidate is the only legal candidate. Therefore, the wellformedness of optimal candidates is not calculated, it is only determined that they have the highest wellformedness possible. With phonotactic segmentation, this will not suffice; many segmentations can be legal, especially in the case of legal clusters.

Nevertheless, there are alternative versions of OT that can capture gradient wellformedness differences (Coetzee, 2008, 2009). In general, one can assume that phonotactic constraints yield a legal candidate, which can be passed on to lexical look-up to evaluate if it contains words. If not, other candidates can be tried, in order of wellformedness. This might seem like an extension of OT grammars, but in fact OT is a special case of so-called Harmonic Grammars, in which all constraints have a weight and the weighted constraint violations add up to yield a harmony score (Legendre et al., 1990). OT is a form of Harmonic Grammar that does not allow the violation of any constraint to be less severe than any combination of lower-ranked constraint violations. Note that it is also possible that only one candidate is considered, but that the selection of this candidate is stochastic; the more wellformed a candidate, the more likely it is to be picked. This option is in line with Stochastic OT (Boersma, 1997; Boersma & Hayes, 2001).

The data gathered in this article are not precise enough to take a stance on the exact implementation of the non-categoricalness of the segmentation process; it suffices to say that segmentation candidates are ranked by well-
formedness. Whether these candidates are then evaluated one by one, or all together but with more attention for the more wellformed candidates, cannot be answered by the present data.

The data do show that because the best segmentation candidate might not yield a word, the segmentation hypothesis selection process is not guaranteed to terminate at the level of phonotactic knowledge. The process will have to backtrack through other segmentation hypotheses, or keep multiple segmentation hypotheses active. If the process is sequential, the last eliminated hypotheses becomes relevant again when the optimal candidate does not concur with any words. Alternatively, if a set of alternative hypotheses are concurrently forwarded to lexical access, but with higher levels of attention for the wellformed ones, the less wellformed ones can still ‘win’ the race by yielding recognised lexical items, only they will be less likely or less quick to do so.

A special case occurs for some candidates that need not even be considered lexically before being decided: illegal clusters have zero wellformedness and are thus not considered as possible word onsets. Segmentation hypotheses containing illegal clusters can be assigned a very low or even zero probability and are thus hardly considered. The difference between wellformedness and illegality is more a difference between gradience and categorical illformedness. Nevertheless, it is unclear whether categorically illegal options are within the ranked segmentation hypothesis set at an extremely low level, or whether they are completely pruned out as soon as it is clear they are illegal. The latter would mean a return to classic OT for the lower half of the candidate set; these are irrelevant again, while the higher half (the legal candidates) are ranked.

Whichever model is preferred for speech segmentation, an algorithmic one similar to the Kager (2010) proposal or one based on activation levels to calculate probabilities of lexical recognition, like the Shortlist B architecture, matters more for the implementation of illegality knowledge. If illegality is more than just low wellformedness, a mechanism has to be assumed to be responsible for the related pruning-down of the segmentation hypothesis space. If illegality is just low wellformedness, both types of model can do with a simple assignment of either rankings or probabilities. In fact, OT and especially Harmonic Grammar are very closely related to connectionist models and Shortlist A is a connectionist model as well, while Shortlist B is Bayesian. It is left to other researchers to decide between the options.

This study provided evidence that phonotactics influences speech segmentation independently of lexical cues. This does not mean that segmentation is not influenced by lexical cues; it is, but also by phonotactic cues and phonotactics gives its opinion whether lexical recognition asks for it or not. The phonotactic influence is not categorical; it does not decide on the way to go,
but guides lexical access, that is therefore more likely to be needed sparingly before a correct segmentation is encountered.
Phonotactic cues affect speech segmentation. Results from a newly developed methodology using eye-tracking made it possible to detect segmentation attempts in the absence of lexical items or contrary to lexical segmentations. These attempts were modulated by phonotactic cues. Second language learners, with English as Target and Dutch as Source language, use the cues of the correct language in contexts in which the languages disagree. This suggests that phonotactic knowledge is truly linguistic knowledge, tied to the language for or in which it is acquired. Phonotactic effects on speech segmentation can therefore not be explained as the result of pure probabilistic learning.

Keywords: Word recognition; Speech segmentation; Eye-tracking; Phonotactics; Second Language Acquisition; Language mode

4.1 Introduction

Phonotactic knowledge is one of the types of non-lexical knowledge that affects speech segmentation, the process that maps the continuous speech stream onto a chain of words (see for overviews Cutler, 1994; Mattys et al., 2005). The phonotactic knowledge and distributional cues that have been found by numerous researchers to influence segmentation (Saffran et al., 1996; McQueen, 1998; Mattys et al., 2005) are largely language-specific, even if phonotactic restrictions or allowances on word forms are (partially) grounded in phonetic markedness. If phonotactics were language-independent, languages would universally allow the same word forms. This is not the case, as witnessed by e.g. the adaptations that some loanwords have to undergo to match the borrowing language’s phonotactics.

The differences between languages’ phonotactics, however, need not imply that the latter are differentiated in cognition. If statistical learning is the source of phonotactic knowledge, a language learner who is exposed to two languages will possibly simply learn a mix of the phonotactics of both languages. In other
words, one language might have different phonotactics from another, which is a typological observation, but a language user that knows both languages might not have two sets of phonotactic knowledge; the knowledge might be stored in a single compound set. A language user who knows only one language, e.g. English, might not have represented her phonotactic knowledge as ‘the phonotactics of English’, but just as ‘phonotactics’ and, upon starting to learn another language, have added the phonotactics from another language to this set. The difference, quite inconsequential for a monolingual, is thus quite important for a bilingual.

Weber (2001) showed that phonotactics of a second language can be acquired, but the results they reported do not make it possible to answer the question whether these second-language phonotactics are added to the first language, or stored as part of the knowledge of the second language; the experiments contrasted phoneme sequences that are illegal in either German or English and found that German L1 English L2 listeners apply the German phonotactics when listening to English. It is possible that they also applied their English knowledge to German.

However, if the listeners represent phonotactic knowledge as belonging to the language they acquired it on, they would possibly show interference of the knowledge of one language in the other, but still a difference between the two. This chapter examines whether second language listeners can use phonotactic knowledge of two languages independently.

It is informative to look at the situation for other aspects of linguistic knowledge. For words, interference occurs, but lexical items can be intuitively assumed to be represented in the mind with a label that assigns them to one language. An English learner of German might, when thinking of a table, think of the word “Tisch”, or even say that word, but she will still be aware that the word is German and not English and she will normally also be able to choose the language she is speaking. This is not an easy task and interference needs to be constantly suppressed, as witnessed by the neuroimaging study by van Heuven, Schriefers, Dijkstra, & Hagoort (2008).

For phonology, there are ample research results on language interference in phoneme perception and production. Flege (1987) (see also Flege, Frieda, & Nozawa, 1997) proposes that if phonemes in two languages overlap in their acoustic dimensions, they are hard to distinguish for learners of the one language with the other as native language. One the other hand, phonemes that occupy a part of phonetic space in one language that is not used in another are much easier to learn. Escudero & Boersma (2004) argue that second language learners start learning by transferring their first language categories. They are sensitive to probability distributions of the sounds they perceive in the second language; when these show categories (in the form
of a multimodal distribution) on a dimension that was previously not used for categorisation, new categories are easily formed. Furthermore, the new grammar, consisting of a copy of the constraints of the native language, including its constraint hierarchy, is adapted by reranking the constraints. The influence of the first language is thus that it defines the starting point for learning.

Phonotactics that conflict on one and the same phoneme sequence have not yet been studied between languages. It is important to notice that conflicting knowledge might be hard to acquire, as native phonotactic knowledge is often persistently used to filter perception. Such phonotactic filters do not only apply to loanwords, but more generally to the perception of phoneme combinations that are illegal in one’s first language; these combinations are prone to misperceptions (Halle et al., 1998; Dupoux et al., 1999, 2001; Berent et al., 2007). This filtering might deprive a second language learner of input, either from speech or from lexical items, that could be used to learn phonotactics.

On the other hand, highly frequent structures have been found to facilitate tasks that involve perception, such as non-word repetition and lexical decision (Vitevitch et al., 1997; Vitevitch & Luce, 1998, 1999, 2005; Vitevitch, Luce, Fisoni, & Auer, 1999) and phoneme recognition (McQueen & Pitt, 1996). Although high frequency does not necessarily entail high phonotactic well-formedness, there seems to be an asymmetrical difference between illegality and wellformedness: wellformed structures facilitate perception, whereas illegal structures are filtered out.

It is likely that negative phonotactic knowledge, i.e. the knowledge that certain combinations are illegal, has effects that only really facilitate speech perception in the language that it belongs to, not in a second language with different phonotactic properties. In the Dupoux et al. (1999) experiments, Japanese listeners have trouble distinguishing between [ebzo] and [ebuzo], likely because */ebzo/* is illegal in Japanese and the listeners perceived both inputs as */ebuzo/*. This is good for Japanese listeners listening to Japanese, because a [bz] sequence in their input is never part of the intended production; there might have been noise in the reception of the input, or perhaps the speaker was careless, but repairing the illegal structure in perception can never hinder lexical recognition as there are no lexical items containing */bz*/.

However, filtering out consonant sequences in a second language by means of first language knowledge is usually not beneficial to the listener. For example, */bz/* is not illegal in all languages; applying the Japanese knowledge to French leads to misperceptions, as French allows many more consonant clusters. The same holds in reverse for the French participants, who had more trouble distinguishing the vowel length contrast in [ebuzo]–[ebu:zo], that
Japanese listeners could perceive. The L1 French listeners thus misperceive Japanese because they apply French phonotactics to it, and vice versa.

Applying knowledge of gradient wellformedness of a language to facilitate the recognition of words containing wellformed sound combinations does not have the same possibly categorical effects and therefore can be tentatively assumed not to be as dangerous when incorrectly applied to Target Language perception. When a combination that is wellformed in the Source Language, but not in the Target Language, is given preference by a listener listening to Target Language input, this might give the wrong kind of facilitation, but it does not delete the right percept.

In comparison to the effect of wellformedness, Source Language phonotactic illegality has the potential to do greater damage to Target Language perception, because the set of possible percepts might be pruned down too narrowly and no longer contain the correct percept. This is not the case for speech segmentation, where illegality does not entail a restriction, but rather opens possibilities. While illegal structures in the input might be due to noise (mispronunciations or misperceptions) and thus have to be filtered out, illegal combinations in continuous speech can occur, but only across word boundaries; hence they effectively signal such boundaries (McQueen, 1998) (see Chapter 3) and solve a part of the computationally challenging task of recognising words in continuous speech. If Source Language phonotactic illegality is applied to Target Language segmentation, it might lead to oversegmentation of the speech stream, i.e. word boundaries might be ‘detected’ in the Target Language speech stream at positions were they are not actually likely to be.

The oversegmentation of a second language is more or less what Weber & Cutler (2006) found. They showed, first and foremost, that phonotactic knowledge of illegality in a second language can be acquired and put to use in segmentation, in addition to first language knowledge. The participants of the Weber & Cutler (2006) study were highly proficient German learners of English, who proved to use both English and German phonotactic cues combined. This indicates that these learners had transferred German phonotactic knowledge to their representation of English phonotactics and later added English knowledge due to exposure. The results of Weber & Cutler (2006) coincidentally also show that illegality knowledge from two languages can be combined, allowing both languages’ phonotactics to be simultaneously active in the detection of boundaries.

The phonotactic knowledge that signals boundaries in the study by Weber & Cutler (2006) is knowledge of the illegality of phoneme combinations. German phonotactics forbids e.g. word-initial /sl/, which is legal in English, while English forbids e.g. word-initial /fl/, which is legal in German. The
participants in the experiment had to spot words in nonsense strings and illegal clusters provided a benefit to word recognition, because they cue a boundary aligned with the word. Hence, in the case of the nonsense strings [darslidzn] and [darSlidzn], containing the English word /lidzn/, ‘legion’, the L1 German listeners used both English and German illegality knowledge when listening to English. In the example, the German illegality of /sl/ cues the segmentation /dars.lidzn/, which allows the word to be found with greater ease. English L1 listeners do not consider /sl/ illegal, as it is not illegal in English; therefore they do not benefit from any cue; they only benefit in the case of [darSlidzn], as /Sl/ is illegal. The German L1 listeners, proficient in English, also benefitted from these kind of boundaries, induced by English illegality.

Interestingly, if the German learners had acquired the wellformedness of /sl/ in English, instead of the English illegality of /fl/, they would have lost the advantage of the phonotactic boundary, instead of gaining an advantage. This suggests that illegality, as opposed to legality, is most active in segmentation and that illegality knowledge from Source Language and Target Language is combined.

Note that in the Weber & Cutler experiment, the second language learner is at an advantage, but normally, if the words did not start in the middle of the cluster, but had the whole cluster as their onset, would be at a disadvantage, as a word like /slip/ ‘sleep’ would then have been hard to spot in [darslip] for them. L1 German listeners would prefer the segmentation /dars.lip/ due to their German constraint against /sl/ and miss the contiguity of the cluster, while L1 English listeners do not have this problem due to the absence of an English constraint against /sl/. The English listeners can therefore be expected to detect the many English words starting with /sl/ with greater ease than the L1 German listeners will be able to do.

However, there are no results on potentially adverse conditions caused by Source Language interference in Target Language segmentation. This makes the results of Weber (2001) in principle compatible with several interpretations. These interpretations can be viewed as lying between two extremes. One extreme is that illegality knowledge from both languages is lumped together and used by one and the same segmentation process. This implies that knowledge of the source language is no longer attached to the phonotactic constraints. It is tantamount to the supposition that phonotactically guided segmentation is not language-specific and that if segmentation is language-specific, this is only because it uses phonotactic knowledge that happens to be language-specific, for instance because it has been acquired by exposure to the native language (Adriaans & Kager, 2010; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993).
The other extreme on the spectrum is the assumption that segmentation of the speech stream is performed separately by the phonotactic systems of the languages the listener knows, with solutions combined at the end of the process in case of interference. Figure 4.1 shows the two extremes.

\[
\text{Segmentation}_{L1}(\text{phonotactics}_{L1}, \text{speech}) \rightarrow \{\text{Boundaries}_{L1}\} \rightarrow \{\text{Boundaries}_{L1}\} \\
\text{Segmentation}_{L2}(\text{phonotactics}_{L2}, \text{speech}) \rightarrow \{\text{Boundaries}_{L2}\} \rightarrow \{\text{Boundaries}_{L2}\}
\]

Segmentation \(\text{phonotactics}_{L1}, \text{phonotactics}_{L2}, \text{speech}\) \(\rightarrow \{\text{Boundaries}\}\)

Figure 4.1: Two options for segmentation interference: a disjoint segmentation process generating the boundaries of both languages (top) or a joint segmentation process using the phonotactics of both languages. Segmentation is a function using phonotactics and the speech stream, yielding segmentation points.

If phonotactically driven segmentation is based on detecting boundaries with the use of illegality knowledge and the two languages are not in conflict, the union of the boundaries marked by either L1 or L2 is the outcome of both the combination of two disjoint segmentation processes and of one joint segmentation process using the phonotactics from two languages. However, if the cues conflict, only one of the two can be preferred per segmentation process. The first option, disjoint segmentation processes, would then cause both boundaries to be detected, while a joint segmentation process has to resolve the conflict between cues of two languages in a similar way as it resolves the conflicts between different cues of one language.

The case of conflicting cues in segmentation therefore provides a valuable test case for hypotheses on the representation of phonotactic knowledge. If phonotactic constraints are used in a joint segmentation process, conflict situations would not surface in the segmentation hypotheses entertained; one segmentation would be preferred. If two segmentation processes are active, two segmentation processes that are not unifiable can both affect speech segmentation. Figure 4.2 shows the application of the two models to speech input that elicits conflicting constraints in two languages.

A problem for testing the two models given above is that it is possible that when segmenting a speech stream known to belong to a particular language, information relevant to this language might be given more importance. If that is the case, the disjoint segmentation option might look like the joint segmentation option, because the outcome of the appropriate segmentation can then become the actual outcome, pushing out the other segmentation.
4.1 INTRODUCTION

However, it is not likely that Weber (2001) would have found an effect of Source Language (L1) phonotactics in Target Language (L2) segmentation in such a case: if segmentation is disjoint, German L1 listeners should have shifted to the English segmentation and not have benefitted (as much) from German cues when listening to English speech. Note that if segmentation is a joint process, it is possible that attention is focussed on one of the two constraints. This is what Grosjean (1999) calls a language mode effect. However, it would in the present case be tantamount to having a different segmentation process that operates differently on the same set of knowledge and hence should not be described as a joint segmentation process. Note that the two options given are only the end points of a spectrum.

The present chapter tries to eliminate one end point of the spectrum, the idea that segmentation is performed in one joint process that uses all acquired phonotactic knowledge indiscriminately. If a shift in segmentation behaviour due to language mode is observed, this would provide evidence for two separate segmentation processes, though the separation might lie in different evaluations of the importance of constraints in a joint set. The simplest model, the joint segmentation model, can then be rejected, as that cannot describe the segmentation process. Note that this does not mean that the other extreme has to be accepted; the two segmentations still do not have to be completely independent.

This chapter presents evidence from segmentation of Dutch and English. In Dutch, final devoicing makes it illegal to split clusters starting with a voiced obstruent, because such a segmentation would contain a voiced obstruent in coda position. English, on the other hand, allows voiced obstruents. In
addition, some clusters starting with voiced obstruents are illegal as onsets, notably /vr/ and /vl/. These are expected to be split by English listeners, while they are expected to be kept together by Dutch listeners. Similarly, but in the opposite direction, the cluster /kn/ is a legal onset in Dutch, but not in English. While it is legal to split it in Dutch, it can be assumed that this is not the first strategy used by Dutch listeners, as onsets are preferred over codas (MAXIMIZEONSET (see Church, 1987)). However, if illegality is the only type of phonotactic knowledge used for segmentation, /kn/, /vl/ and /vr/ should behave differently in English, where they are illegal, compared to Dutch, where they are legal.

The experiments reported in this chapter feature participants that are proficient in English, but have Dutch as a first language. In comparison to the experiment reported by Weber (2001), these listeners are less proficient in their second language; they are not interpreters but university students. However, due to the advanced level of English taught at the secondary schools that prepare for access to universities and the omnipresence of English in Dutch cinema’s and on Dutch television (films and TV programmes are subtitled, not dubbed), these listeners have a reasonably advanced level of English and also to have received substantial exposure to native English speech.

Just as in Weber (2001) and McQueen (1998), a word-spotting experiment was used to assess what the effect of conflicting cues is. In previous experiments (Experiment 3 in Chapter 3) the interdependence of speed and accuracy was found to affect responses to words; latency is only recorded when a word is actually spotted, which might in the more difficult conditions be only when the word is very salient. When phonotactic cues conflict, the non-lexical (prelexical) effect of phonotactics might become even less relevant; participants might mostly respond to words they recognise without prelexical segmentation based on sublexical cues.

To avoid null results on lexical response variables, the current experiment uses the eye-tracking paradigm of the previous chapter. This methodology allows to detect segmentation behaviour when no word is recognised, as well as when a word is recognised. The paradigm can reveal preferences to segment in the middle of the cluster or at its onset.

Segmentations at the onset of cluster are expected for legal clusters, while segmentations within a cluster is expected for illegal clusters. This means that if speech segmentation is performed in English, clusters like /vl/ will not be segmented as a one unit, i.e. with a boundary before the /v/, but split at the /l/. In Dutch, segmentation at the middle is not even allowed for /vl/, as this would leave a residue with an illegal voiced obstruent in coda position.

If advanced Dutch L1 learners of English have acquired the English illegality of the clusters /vl/, /vr/ and /kn/, then segmentations before the cluster
should be made less, as segmentation in the middle becomes more likely. The present chapter assesses whether this prediction is correct and if so, if the knowledge acquired about English is applied to English speech only, or also to Dutch speech. Comparing the two segmentation locations among Dutch listeners in both English and Dutch can indicate if listeners distinguishing sets of phonotactic knowledge from different languages.

The crucial comparison is thus between fixations to (distracting) clusters and targets; this difference might be modulated by the language of the experiment and the illegality in that language. As the same clusters can be used in are the same in the two versions, experimental control can be tighter and conflicting phonotactic knowledge can more easily be teased out. The clusters that are illegal in English only are compared to two other types. The clusters /kn/, /vl/ and /vr/ are illegal in English, not Dutch. The clusters /sn/, /l/ and /fr/ are legal in English and Dutch, while the clusters /mn/, /ml/, /sr/, are not legal in either of the languages. For these groups of clusters, different predictions can be made. Table 4.1 shows the predictions on how hard it is to segment a cluster.

<table>
<thead>
<tr>
<th>Phonotactics</th>
<th>Context</th>
<th>English</th>
<th>Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligned</td>
<td>mn, ml, sr</td>
<td>easy</td>
<td>easy</td>
</tr>
<tr>
<td>English-aligned</td>
<td>kn, vl, vr</td>
<td>easy</td>
<td>hard</td>
</tr>
<tr>
<td>Misaligned</td>
<td>sn, fl, fr</td>
<td>hard</td>
<td>hard</td>
</tr>
</tbody>
</table>

Table 4.1: Experiment 5a: Phonotactic conditions and predictions for the ease of splitting them in the two languages.

As word recognition also has a strong effect on segmentation, the trials in which the second phoneme of the cluster is the start of a word are not used. Instead, trials in which no word is present are crucial. In these trials, only phonotactic cues can be used. It is quite likely that Dutch learners of English can recognise monosyllabic English words. In addition, English words very often have Dutch cognates that start with the same onset; lexical interference will thus not be problematic and any word, either Dutch or English, will be activated if present in the auditory stimulus. As phonotactic effects are likely to be obscured in the results for word trials, these trials were omitted from the analysis.
4.2 EXPERIMENT 5A

4.2.1 Procedure

This experiment was run with an Eyelink I head-mounted eye-tracker, sampling fixation position at 250 Hz. After calibration, participants received instructions and practised on nine trials, repeated if needed.

Trials started with the visual presentation of four letters in large blue rectangles on a white background on a computer screen in front of the participants; the rectangles corresponded to fixation regions. After 3000 ms, a dot appeared in the centre of the screen centre. Participants could take as much time as desired to read the letters; they were told to fixate on the dot when they were ready for the auditory stimulus. When the eye-tracker detected fixation on the dot, it performed drift correction and removed the dot. 250 ms later, a sound file was played containing a nonsense string of two syllables. Participants were instructed to indicate if they spotted a word in the nonsense string by looking at the first letter of the word. This letter was present in one of the four regions. They were informed of the language version of the experiment they would get: either Dutch or English. Instructions on screen were in the appropriate language. Participants were told to respond only to words in the language they were tested in. They knew that looking at wrong letters was allowed and did not result in errors, but that it made it harder to find the right letter in time. When participants fixated for 300 ms on the target in the Word condition, the target changed to green and the trial ended. This constitutes a correct response. Reaction times were aligned with the ends of the words. Trials timed out after 2500 ms. In the Non-word condition, all trials timed out. 120 trials were presented randomly, divided in four blocks with pauses in between.

4.2.2 Participants

Forty-nine native speakers of Dutch from the Utrecht Institute of Linguistics OTS participant pool participated for monetary compensation. More participants were recruited but could not finish the experiment due to problems in the calibration process of the eye-tracker, usually related to a participant’s way of fixating or to a lack of contrast between pupil darkness and darkness of other regions within the tracker’s receptive field, usually eyelashes. Assignment of participants to the Dutch or English Language versions was by one-by-one alternation. 25 participants completed the English version and 24 the Dutch version.
### 4.2 Experiment 5a

#### Materials

The experiment was implemented in two versions, English and Dutch. The versions are treated as levels of the condition Language and participants were nested under this condition. The difference between the versions is that they consisted of different sets of items, but the items were generated using the same principles.

A set of 30 Dutch or 30 English words was used to construct the items. For each word, a non-word was created that matched the onset of the word as far as possible. The Lexicality condition differs on whether the target is the word (Word) or the Non-word (Non-word level). The two targets of every item were embedded in three nonsense strings, that differed only in the clusters at the boundary, as described above. This yields six versions of each item. All syllables contained unreduced vowels.

Experimental items all contained the targets embedded at the end of the nonsense string. Three clusters were used for each item and the first letters of these cluster, plus the first latter of the target, were the visual display part of every item. Table 4.2 shows how words and non-words are embedded in three different contexts.

Items were rotated in a Latin Square design through the Phonotactics and Lexicality conditions. For tight control over duration factors and to avoid production effects of boundaries assumed by speakers, the strings were synthesised with MBROLA (Dutoit et al., 1996); phoneme durations were

<table>
<thead>
<tr>
<th>Lexicality</th>
<th>Phonotactics</th>
<th>Dutch</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>[sym.ləŋ]</td>
<td>[sy.vɨləŋ]</td>
<td>[sy.fləŋ]</td>
</tr>
<tr>
<td>Non-word</td>
<td>[sym.ləj]</td>
<td>[sy.vɨləj]</td>
<td>[sy.fləj]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexicality</th>
<th>Phonotactics</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>[nɛm.ˈlʌk]</td>
<td>[nɛv.ˈlʌk]</td>
</tr>
<tr>
<td>Non-word</td>
<td>[nɛm.ˈlɐ]</td>
<td>[nɛv.ˈlɐ]</td>
</tr>
</tbody>
</table>

Letters shown: l, m, v, f

Table 4.2: Examples of stimuli and conditions. Boldface indicates embedded words, dots indicate phonotactically preferable boundaries.
based on Waals (1999) for Dutch. For English, consonant durations were taken from Umeda (1977). The lengths given by Umeda are on average 67% of those given by Waals (average of all monoconsonantal consonants that are in both studies, namely /p, b, f, v, m, t, d, s, n, l, k/). Therefore, the Umeda values were divided by 0.67 to make the durations longer. The voices that were used were the Dutch 'nl2' and the English 'en1', provided with MBROLA.

As duration influences segmentation (Quené, 1989), compression rates of phonemes in illegal clusters were made identical to those in their legal counterparts. This reduces the chance of speaker interpretations of boundaries and thus of other sublexical cues creeping in. Note that the compression thus preserves durational differences between voiced and voiceless consonants. This was also roughly the case for pairs of legal voiced and voiceless clusters such as /pr/-/br/.

Every item was assigned four letters for visual display: one corresponding to the target and three distractors corresponding to the boundaries: one to the phonotactic boundary in the Dutch-Misaligned condition, one to the Aligned condition and one to the Misaligned condition. Note that the in the Misaligned condition, the target is misaligned but the distractor is actually aligned with a phonotactic boundary.

Ninety filler items were constructed to avoid participants noticing that the three target letters (r,l,n) are word onsets in 50% of the trials and the distractors never are word onsets. In the fillers, consonants that were distractors in the experimental trials were targets. Thirty of the fillers contained the word embedded at the end, sixty at the beginning of the nonsense string; words (and in general targets) had thus a 50% chance of being in either position over all items. The Latin Square rotation included the fillers with regard to the Lexicality conditions. Half of the trials in the expirement contained an actual word.

4.2.4 Results

Latency and accuracy

Overall accuracy on the experimental items was 51% (369 out of 720) and the average reaction time is 821 ms (SE 23). This is not a large deviance from McQueen’s results of 43% and 708 ms for the stimuli that had two strong syllables.

Latency and accuracy effects are not used to test any hypotheses, as fixations give more accurate data and multiple comparisons would reduce power due to the necessary corrections. However, to give insight into the data, all Word trials were analysed in a mixed model with random effects for item and participant (Quené & van den Bergh, 2008; Baayen et al., 2008).
Table 4.3: Modelled latency and accuracy. A mixed model with crossed random effects was fitted to the log-transformed reaction times measured from word offset; for accuracy a model was fit on the logit of spotting the word.

The models, fit to probability of correctness (transformed to their logits) and the log-transform of the reaction time, are in line with the prediction that when Phonotactics is English-Aligned, accuracy goes down and latency up, but less so when the experiment is presented in English. However, the significance values are not convincing enough to make it possible to draw conclusions. Table 4.3 shows the estimates from the model.

Fixations

The hypothesis under discussion can be tested with eye fixation data, especially when gathered in the Non-word trials; these are least sensitive to word recognition. An analysis of the fixation data was performed on the ratio between target and distracting cluster. To obtain this ratio, those fixations that were on the distractor that was present in the sound file or on the target were used. The dependent variable is the probability that these fixations are on the target. The more distracting the alternative word boundary, the less likely this is. A disadvantage of this analysis would be that certain intrinsic characteristics of letters could make them more attractive to look at, irrespective of properties of experimental factors. However, as the hypothesis is related to a possible shift in behaviour between the Language levels, these effects are not a confound for the hypothesis test; the comparisons are between sets of trials that show the same letters.

The 95% confidence interval of all reaction time, including all participants, ended at 865 ms. Fixations to targets in a pilot had a slowly rising pattern from word offset; very few fixations are recorded before this point. Analyses
were therefore restricted to 0–865 ms after word offset. Growth Curve Analysis (Mirman et al., 2008; Barr, 2008) on the logit was used, without aggregation, as there is sufficient data. Models contain third-order orthogonal polynomial time terms, because visual inspection of the pilot data showed two bends in the fixation development. This was confirmed by looking at the grand mean significance of the three terms.

The factors influencing fixation are Language, Phonotactics and Lexicality. Main effects of all these factors were expected, but interactions of Language and Phonotactics are relevant for falsification of the null hypothesis that there is no language effect on the use of phonotactic knowledge. Therefore the main effects of the three factors and the interactions between Phonotactics and Lexicality were used in one model. This model was then compared to a model that also has interactions with Language. This last model is significantly better in capturing the data ($\chi^2 = 1179.5$ (20 d.f.), $p < 0.00001 \ast \ast \ast$).

The best-fitting model is fully given in the Table B.1, p. 202. Figure 4.3 shows the modelled trajectory imposed over the actual proportions for the ratio between target and distractors in the English-aligned trials.

The null hypothesis that fixations to English-Aligned targets in the Non-word condition are as likely in the English Language version of the experiment as in the Dutch version has to be rejected; targets are more likely to be fixated on (and distractors less) in the English-Aligned trials when they are presented in the English Language version of the experiment. The effect is less prominent for the Word trials and in the English version, Word trials are less different from Non-word trials; targets corresponding to English words are thus in general not found as easily as Dutch word targets.

4.2.5 Discussion

The results show that Dutch listeners are capable of alternating between Dutch and English phonotactics in speech segmentation. This means that phonotactic constraints are not used indiscriminately in one segmentation process. The participants did not only have knowledge about the phonotactics of both Dutch and English, but also know to what language that knowledge belongs; e.g., the constraint /*vl/ is labelled as a constraint on English and applied to English speech, but not to Dutch speech. One could assume that constraints are switched ‘on’ and ‘off’ depending on the language mode of the user; however, this would be at odds with the findings of Weber (2001).

The language effects partially confirm the predicted effects of cluster legality and illegality. For English, this suggests that the Dutch learners behave like native listeners. However, before drawing more elaborate conclusions, the behaviour of native listeners of English has to be assessed. The next experiment
Figure 4.3: Experiment 5a: ratio of fixation on target to fixation on distractor present in auditory stimulus, for trials containing the English-Aligned cluster. Fitted model predictions are superimposed as lines. Points are proportions aggregated over trials. The solid line at 0.5 represents chance level (one target / one distractor).

establishes whether the pattern in the results of the Dutch listeners for the English version is similar to the pattern of native English listeners.

4.3 EXPERIMENT 5B

4.3.1 Procedure

Experiment 5a was run again with a small group of native English listeners, to see if the Dutch participants in the English version had behaved like native English listeners would. The procedure was the same as in Experiment 5a, but only the English Language version was used.
4.3.2 Participants

Seven native speakers of English participated for monetary compensation. They spoke no Dutch or only a limited amount; none spoke Dutch on a daily basis or more than English. However, as they all lived in the Netherlands, they were probably more or less systematically exposed to a substantial amount of Dutch.

4.3.3 Materials

Materials were the same as in Experiment 5a. Only the English Language version was used.

4.3.4 Results

Latency and accuracy

As the latency and accuracy for Experiment 5a did not reveal any effects, they were not analysed for Experiment 5b as there is no point in comparing insignificant effects. The reaction time was on average 877 (95% confidence interval 779–975); the accuracy was 64%. The higher accuracy in the control group is thus paired with a lower speed, possibly because native listeners spotted some words that were hard to spot.

Fixations

The same model was fitted to the fixation data of the control group and the Dutch group performing the English experiment in Experiment 5a. The Language condition was now replaced with a L1 condition, with the Dutch group having Dutch as an L1 and the native group English.

Apart from baseline differences, significant differences between Dutch native and English native listeners, on the base level irrespective of time, were only found for the factors reported below. For the Misaligned clusters on the Non-word baseline, the probability of target fixations compared to distractor fixations was higher for the Dutch group (0.565 (se 0.120), \(z = 4.717, p < 0.00001\) ***), possibly because they did not reach the conclusion that no word was present as fast as native speakers; the Dutch speakers had a lower preference for the targets in the Word condition (-1.398 (se 0.120), \(z = -11.625, p < 0.00001\) ***). The Dutch speakers fixated more on the target for the English-aligned, Word trials (1.835 (se 0.196), \(z = 9.376, p < 0.000001\) ***) but also for the overall Misaligned Word trials (0.805 (se 0.169), \(z = 4.779, p < 0.00001\) ***), confirming that they are more sensitive to lexical cues.
than phonotactic cues. Hence, the Non-word baseline was indeed the best place to test the null hypothesis.

Because the terms on other time levels are harder to interpret and skewed due to the different reaction time and latency, the interpretation of these differences is left to a visual inspection of the data as given in Figure 4.4. The graph shows that the Dutch participants seem to be less sensitive to phonotactic alignment: whereas the English participants look at Aligned targets from the start, the Dutch participants are more confused. Later on, both the Misaligned and Dutch-Misaligned condition show greater attention to distractors for English native listeners, but this is likely to be due to the fact that these participants detected more words, making the data of the last part of the time window based on the trials in which the word was absent. In contrast, in the beginning the data are also based on many trials in which the word was still to be found.

4.3.5 Discussion

The Dutch participants in the English task do seem to behave more English-like, as they do not differ on the crucial variable of Experiment 5a, the preference for the English-aligned target over the distractor, but the non-native participants did show poorer word-spotting skills and less sensitivity to the phonotactics of the speech stream.

4.4 GENERAL DISCUSSION

It is most likely that the segmentation process is sensitive to language mode. In case two languages have conflicting segmentation constraints, the system responsible for generating segmentation cues using phonotactic knowledge has to make decisions. However, one might wonder if the Weber (2001) experiment should not also have incurred conflicts between wellformedness in one language and illegality in another. E.g., /sl/ is illegal in German, but wellformed in English, and the German illegality ‘won’, i.e., provided a benefit. On the other hand, /Sl/ is legal in German, but illegal in English, and the English illegality won. Note however that the results gathered in the Weber (2001) experiment are lexical; they only include trials that included a word.

In the present study, the Word trials do not show the largest difference in segmentation behaviour. It is in fact in the Non-word case that English-aligned trials show a Language-dependent effect on the target/distractor fixation ratio. In case a Word is present, this effect is attenuated; the phonotactic properties of the target alignment matter less for Word trials. It is thus likely that when either of the two languages signals a boundary, a segmentation hypothesis
containing this boundary is entertained. The segmentation process considers the best hypothesis first, but other hypotheses are also considered. If the preferred hypothesis does not yield a word, the second best is assessed on its merits; if the second best yields a word, it will be taken into account.

As an example, an auditory stimulus containing \([\text{vt}]\) is segmented as \(/.\text{vt}/\) using Dutch phonotactics (*/v*/r/) and as \(/\text{v}.\text{t}/\) using English phonotactics (*/v/\text{t}/\)). Depending on the language mode of the listener, one of them is preferred. However, if, in Dutch, \(/.\text{vt}/\) does not yield a word (because the target starts at the \(/\text{t}/\)), the second option \(/\text{v}.\text{t}/\) is considered, which is generated when English phonotactics are applied. This segmentation supports recognition of the word in the Word trials, providing the target with an advantage over the Misaligned trials, in which neither of the languages' phonotactics provides a cue for the target boundary.

Thus, phonotactic knowledge from a second language can spawn segmentation hypotheses, but these segmentations are considered according to the language mode of the listener. The phonotactic knowledge that explains the effect cannot be seen as purely a collection of statistical regularities learned on the input, as the representation is labelled as belonging to one of the two languages. Segmenting at phonotactically likely boundaries is usually correct and thus successful and such behaviour might simply be reinforced by the statistics of the task. However, the learning that is the result of these statistics is performed in the native language most of the time; if phonotactic effects on segmentation were a matter of optimisation, Dutch listeners would not use Dutch phonotactics in Dutch only, but they would use one set of phonotactics they learn over all the input that they have had to segment, which is heavily correlated to Dutch phonotactics as the input is mostly Dutch. Although such a statistical view would be technically compatible with the results reported by Weber (2001), it is not compatible with the findings presented here.
Figure 4.4: Target fixation ratio to distractor present in auditory stimulus, for all trials, by level of the Phonotactic condition. Points are proportions aggregated over trials. The solid line at 0.5 represents chance level (one target / one distractor).
DISCUSSION AND CONCLUSIONS

ABSTRACT

In a broad sense, the topic of this thesis is phonotactic knowledge used in speech perception. Phonotactics defines the wellformedness and/or legality of sound combinations. This wellformedness and legality might be derived from a set of constraints or a representation of the probability of (abstractions of) sound combinations. There are separate effects of both types of phonotactic knowledge. Effects attributed to phonotactic markedness and phonotactic probability are relatively independent, against the predictions of a unified account of phonotactic wellformedness. The formal properties of the phonotactic knowledge to which the effects reported in the previous chapters have to be described hinge on the existence of positive and negative knowledge, that interact in one grammar that can yield both categorical and gradient effects. The evidence from the eye-tracking experiments on segmentation is an argument to consider speech segmentation as a process that is independent of word recognition. The phonotactic grammar therefore has to be included separately in models of word recognition. Chapter 4 showed that phonotactic effects are language dependent; hence, multiple phonotactic grammars can be assumed to be represented in the mind of multilingual listeners.

Keywords: Optimality Theory, Harmonic Grammar, speech perception, word recognition models

5.1 EFFECTS OF PHONOTACTIC KNOWLEDGE

5.1.1 Facilitation and inhibition

The lexical decision experiment of Chapter 2 showed that categorical phonotactic illegality inhibits perception independently of facilitation of high-probability sound combinations. The filter against the combination of an /s/ and a consonant was strongly in place despite its mismatch with probability input, while on the other hand the ‘probability’ of the facilitated epenthesised clusters is
low in the input. A facilitatory effect, presumably caused by knowledge of probabilities, was acquired for the epenthesised forms of frequent clusters.

This facilitatory effect cannot be attributed to exposure to high frequencies of the most wellformed clusters, as the facilitation would then also have to occur for clusters that were not epenthesised, which occur more frequently than the epenthesised versions anyway. The probabilistic knowledge that causes the facilitatory effects is thus not learnt from the acoustic input. This dissociation indicates that not all phonotactic effects on speech perception can be reduced to probabilities: restrictive effects at least are not purely probabilistic. Probabilistic phonotactics also does not completely explain the facilitative effect on epenthesised clusters. Hence, the separate existence of facilitating knowledge and filtering knowledge has to be assumed.

An attempt to attribute the facilitatory effect to markedness constraints fails to explain how these constraints can change their ranking in L2 learners, whereas the constraints causing epentheses stay put. An interpretation in which the filtering effect is attributed to positive representations is not successful, either. For positive representations, such as the sublexical units in the ART model as it has been proposed by Vitevitch & Luce (1999), the problem is that even though many of the findings in their article and in the study by Luce & Large (2001) were replicated in Chapter 2, the model crucially does not account for the inhibition of the perception of illegal clusters. The fact that facilitatory effects were found makes it appropriate to conclude that there are sublexical representations of wellformed /sC/ clusters. However, the level at which these effects were observed does not correspond to the input, but to the output of the phonotactic filtering.

The different effects of illegality and probabilistic wellformedness on speech segmentation are also informative. Segmentation can a priori be expected to benefit from both knowledge of which sound combinations are wellformed and which combinations are illegal. Both kinds of information can potentially have effects in the form of chunking and splitting: negative knowledge can facilitate word recognition by inducing splitting of illegal clusters and rejecting illegal residues, while positive knowledge can inhibit boundaries within wellformed clusters. There was evidence for the effect of illegality at the level of segmentation candidates, as target words that require segmentations containing chunks with voiced obstruent codas were harder to find, when the stimuli were controlled for cluster legality. This kind of rejection of illegal segmentation residues is similar to the effects of impossible words, as expressed by the Possible Word Constraint (Norris et al., 1997). Nevertheless, the PWC was only proposed to account for the inhibitory effect against segmentations that leave chunks that do not fulfil the most basic requirements for ‘wordhood’, most notably the requirement that each word
must contain a vowel. Experiment 4a suggests that even less serious violations of ‘wordhood’, such as the illegal and marked phenomenon of final voicing in Dutch, have an inhibitory effect on segmentations that cause these violations.

At the level of clusters, illegal and legal clusters were compared. Differences found with such comparisons could be due to the illegality of the illegal clusters, or the wellformedness of the legal ones. The results of Experiment 3 are in accordance with the idea that legal clusters are not split as easily as illegal ones, but the effect can also be explained in reverse, namely that legal clusters are more likely to be chunked or that illegal segmentations are avoided. In Experiment 4a, these factors were partially disentangled, by manipulating cluster legality separately from segmentation legality. As influences of both were found, the remaining uncertainty is whether cluster illegality or cluster wellformedness, or both, provide segmentation cues.

Earlier experiments on segmentation in a second language by Weber & Cutler (2006) suggest that the knowledge used in segmentation involves cluster illegality, but this still does not exclude the involvement of probabilistic phonotactic knowledge in segmentation. Experiment 4b indeed suggests that Slavic learners, in whose native language the clusters that are illegal in Dutch are legal, do not consider these clusters impossible as word onsets. Still, they consider illegal clusters as less wellformed and prefer to split them.

This finding takes us to a second important distinction. Illegality is normally categorical; if a cluster is illegal, it has to straddle a word boundary, while a wellformed cluster only possibly does not contain a boundary. The effects of wellformedness are thus necessarily gradient, but the effects of illegality can also be. If illegality is gradient, it should rather be seen as markedness. The next section describes the differences between gradience and categoricalness.

5.1.2 Gradience and categoricalness

Facilitatory effects of phonotactics are in general gradient. This might follow from their relation with probabilistic knowledge, which is by definition gradient as probabilities are not binary. Nevertheless, as discussed in the introduction, markedness constraints can also explain gradient effects.

The facilitation found in Chapter 2 has to be related to probabilistic knowledge of highly frequent /sC/ clusters, as it does not match the effect of markedness effect against all /sC/ clusters. Tentatively, the facilitatory effect is likely to be gradient, since it is seems to have a relatively linear acquisition trajectory instead of a discontinuous jump. However, there are also categorical effects that have to be attributed to the knowledge of markedness. The contrast between epenthesised and faithful versions of /sC/ clusters was absent for
the Spanish group. This effect has to be attributed to the markedness of this type of cluster in Spanish. The markedness constraint is apparently so highly ranked that it causes the /sC/ clusters to be illegal overall.

The experiments on speech segmentation provide additional information on the categoricalness of phonotactic cues. The effects of the different clusters were not of similar sizes. The effects might all be attributed to the markedness of clusters or segmentations, but not to their categorical illegality. Rather, the effects of markedness on segmentation are gradient.

Experiment 4b, a repetition of Experiment 4a with Slavic listeners, proves the possibility of gradient effects of phonotactics on segmentation. The two experiments together (Experiment 4a and Experiment 4b) show the same effect, but with a gradient difference, in the Non-word Misaligned condition. Dutch participants avoid chunks containing the illegal, hence marked, clusters. The effect looks categorical in nature, but due to noise that conclusion is tentative. However, the Slavic participants show a more gradient difference, preferring chunks without the illegal clusters. This means that the difference in wellformedness between clusters that are legal versus illegal in Dutch has to be gradient for the Slavic listeners. In turn, this proves that there are gradient phonotactic effects on segmentation. The difference observed for Slavic listeners cannot have been a categorical effect obscured by noise, as the amount of noise would not be expected to differ so much from the Dutch listeners.

The Slavic listeners are likely to have either transferred the gradient difference from their native languages, or else they learnt the markedness from Dutch input but have not yet fully ranked it high enough for it to constitute categorical illegality. In other words, the Slavic listeners have not yet reached the more categorical difference of native listeners. The categoricalness of the effect for the Dutch listeners can be debated, as mentioned before; it can be an extreme case of a gradient difference.

In sum, the phonotactic knowledge used in segmentation can be either categorical or gradient, but the effects are mostly gradient. Whether the underlying knowledge is categorical or gradient can therefore not be decided. Nevertheless, the Dutch ban on voiced obstruents in codas is, in principle, categorical and knowledge of this ban is applied to speech segmentation. Similarly, the illegal clusters are categorically illegal. The same holds for the phonotactic knowledge that causes the perceptual illusion for the Spanish group in the cross-modal priming experiment of Chapter 2. This illusion was at least good enough to reduce the lexical contrast to an indistinguishable level. There are thus theoretical reasons to assume that phonotactic knowledge is categorical and no empirical reasons to reject this idea. On the other hand, the phonotactic knowledge is capable of causing gradient effects.
The theoretical interpretation of the body of experimental results is that phonotactic knowledge is partially based on markedness and partially on probabilistic wellformedness. In addition, the effects derived from phonotactic knowledge, especially markedness, can be categorical as well as gradient. This section briefly discusses the incorporation of these findings in models of speech recognition.

An explanation based on markedness constraints can explain the categorical filters on illegal sound combinations, while it can also operate gradiently, thus explaining the gradience of markedness effects on segmentation.

If the architecture of the ART model as proposed by its designers is stretched (only tentatively), it is possible to imagine that with the right vocabulary and sublexical units, illegal clusters activate the nearest legal cluster, hence explaining filtering effects. Still, such a model lacks an explanation for the acquisition pattern shown by the Spanish group. They acquire probabilistic knowledge after the markedness filter: the probability of epenthesised clusters can only be high after the markedness filter, as epenthesised clusters are not frequent in normal input. Nevertheless, the Spanish group also did not acquire that some unepenthesised clusters are wellformedness. Hence, the representations learnt from probabilities in the input do not actually match the input. To incorporate this finding in the ART model, the listeners in Spanish must be assumed not to notice the probability of the clusters that are actually present in the acoustic input, while building representations for phoneme combinations that are actually not present in the input. These assumptions are quite absurd and violate the architectural principles of the ART model.

It is thus difficult to capture the effects of markedness reported in Chapter 2 in an ART model. As ART does not directly address speech segmentation, the effects found there are even more complicated to derive from the ART model. The same holds for the Shortlist B model, but as suggested, Shortlist B can accommodate an influx of external information and allow it to modulate the probability of the recognition of words. This same strategy could be applied to ARTWORD. Nevertheless, this strategy does not match the finding that speech segmentation is not a side-effect of successful lexical recognition. Segmentation attempts are made before words are recognised, as shown by Experiment 3. The same applies to filtering effects. As far as both effects are caused by markedness, they seem to shape the input to a word recognition model. In other words, markedness knowledge can be seen as information used to preparse the speech input, while there is as yet no reason to assume that probabilistic knowledge does not take the place between phonemes and words, as proposed for both ARTWORD and Shortlist B.
In the next section, this proposed preparsing process will be described in more detail. Afterwards, a formal account of the role of phonotactics in the process will be discussed.

5.2.1 A preparser for speech input

Given the evidence presented in this thesis, as well as previous theories and empirical results, it can be proposed that phonotactic cues for speech segmentation have to be generated independently of other information generated in or by the speech perception process.

In Experiment 3, the effect of phonotactics on speech segmentation was shown to exist independently of lexical recognition. It would have been possible that phoneme combinations that occur at the start of a large cohort of words attract ‘attention’ purely through the activation of this cohort, even if none of those words were present; this is suggested by Norris (1994, p. 224). However, this does not explain the phonotactic effects found by McQueen (1998) for the alignment at the offset of words; it can only account for effects at onsets. It also fails to explain how there can be a difference in the splitting of illegal clusters, as found in Experiment 4a.

The finding that phonotactics is used before or at least independently of lexical access indicates the possibility that phonotactic knowledge is activated before lexical access takes place. This entails that the preparser does not operate on lexical input, but on input at a lower level.

Hence, the preparser is proposed to perform a check on the phonotactic grammaticality of the input. In case the input is too marked, it is adapted. Adaptations include perceptual illusions, but importantly, also the use of word boundaries. Only segmentations that do not contain marked structures will be allowed as output. All proposed changes to the input can be argued to aid subsequent phoneme recognition and word recognition in later stages, at least in normal situations, as marked structures are unlikely to occur in the correct perception. Even stronger, illegal structures are not only unlikely to occur, but in fact should not occur in the correct perception.

The preparser does not have to be restricted to a system that only evaluates phonotactic grammaticality. It is possibly useful to include other phonological effects in the preparser as well. Knowledge of phonologically predictable phenomena such as assimilations cannot be represented directly in an ART model, although it can usually perform correctly if the lexical representations contain the right canonical forms. If assimilation, such as ‘gardem bench’ for ‘garden bench’ occurs, the word ‘garden’ is activated, as there is no real competition from ‘gardem’ (or “guard ’em”), as there is no word ‘gardem’ and ‘guard him’ does not fit the semantic and syntactic context. However, in
the sentence ‘gun production’, ‘gun’ can be pronounced as ‘gum’. Without knowledge of the phonological regularity, the word ‘gun’ pronounced as ‘gum’, it would be recognised as ‘gum’ by the ART model, as that is by far the best match. In the right phonological context, ‘gum’ should be expected for ‘gun’. Reversely, in a sentence without phonological reasons to assimilate, e.g. ‘this is a gum’, the word ‘gun’ is not a good match (see for an overview Mitterer, 2011).

As the word recognition models discussed above are not frivolously designed, but based on careful research and literature review, the research presented here is not enough to demand a change to a model or to propose the most adequate adaptations. Improvements of existing word recognition models are therefore only suggested tentatively, until more evidence or better solutions become available. Especially the ART model is substantially changed if a phonological preparser is added, as this deforms the architecture and hence the nature of the ART model, because the input stops being the actual acoustic input if a phonotactic preparser is allowed. In addition, the proposed phonological preparser does not offer the same detail and neurological plausibility as the theory of adaptive resonance can provide. Still, it should be mentioned that a constraint-based grammar is an approximation of a neural network (Prince & Smolensky, 1997). If such a grammar is proposed, it should in theory be possible to connect the ARTWORD model and the preparser in a more unified model of speech perception.

Interestingly, the same holds for the introduction of a phonological preparser in a word recognition model of the form of Shortlist B. As Shortlist B assigns probabilities to perception, the preparser would have to yield probabilities of sound combinations, to account for facilitation, as well as probabilities for segmentation paths.

The requirements for the preparser are thus that it generates probabilities, or wellformedness judgements, for all possible interpretations of the input. Hence, the preparser has to generate felicitous input-output mappings. The input is the speech signal, possibly recognised as a string of phonemes. The output contains an adapted string of phonemes that includes boundaries and contains no illegal sound combinations. The preparser captures only a small part of speech recognition and can thus not discard marked but legal outputs. This means the output should contain at least all legal interpretations of the input. However, the preparser can indicate that such outputs are marked, by providing a gradient wellformedness value of some form for each output.

The results of the application of the preparser can in some respects still be categorical, as this allows it to capture the illusions in Chapter 2 by virtue of the principle that the output of perception should not contain illegal sound combinations. However, the preparser needs to allow gradient effects as well,
at least to model the gradient preference for splitting illegal clusters shown by the Slavic listeners in the Non-word condition of Experiment 4b. The grammar likely also needs to model categorical filtering.

Other demands on the preparser are that it should capture the use of both markedness and probabilistic knowledge. It might be possible to displace an account of probability effects by coupling the preparser to an ART model with the sublexical representations proposed by Vitevitch & Luce (1999), but the preparser should ideally also capture the facilitation for wellformed sound structures that are not faithful to the input (as found for the Spanish group in Chapter 2).

5.2.2 Formalizing the phonotactic grammar

As discussed in the introduction, the constraint-based framework of phonological theories can capture both gradient and categorical effects of markedness. Hence, there is no need to supply a new formalism. There are a number of theoretical solutions for generating gradience from a collection of categorical knowledge in the form of constraints (Coetzee, 2008, 2009; Anttila, 2008; Boersma, 1997; Hayes, 1997; Hayes & Wilson, 2008). These solutions have the advantage that the categorical constraints parsimoniously capture linguistic phenomena, while they can still yield gradient wellformedness. As they are based on OT grammars, they are not neurologically implausible. In addition, they can be used to generate probabilities.

The preparser proposed above can be formalised as an OT grammar for perception, as already proposed by Boersma (1998, 1999). He proposes that a perception grammar maps phonetic forms to phonological structures and that these structures are mapped to words by a recognition grammar. The perception grammar’s input is a set of phonetic cues and its output is a string of phonemes. It is thus not directly ready to model phonotactic information. However, when whole strings of input are taken into account, it is possible to add phonotactic markedness to the model. This also makes it possible to model how phonotactic considerations sometimes overrule acoustic cues for phoneme recognition. If the acoustic input is not close to any category, the perception grammar will still recognise it as the nearest category, but phonotactic cues can modulate this process towards the nearest legal category. The possible conflict between ‘near’ and ‘legal’ can be captured with standard OT constraint interaction.

Phonotactic cues have to be taken into account together with acoustic cue constraints. The phonotactic cues should be positioned between cue constraints that protect perfect matches and those that provide shoddy matches. In this light, it is informative to look at the stimuli of Massaro & Cohen (1983).
They used stimuli such as /ple/-/pre/, both legal, but also /tle/-/tre/, where the /l/ is not legal, or /sle/-/sre/, where the /r/ is not legal. Disregarding the first phoneme for a moment, perception of the second phoneme depended on the F3 value at the onset of the glide. It was manipulated to be 2397, 2263, 2136, 2016, 1903, 1796 or 1695 Hz. The first is most like /l/, the last most like /r/. To put this in the Boersma (1999) format of cue constraints, one can suppose the constraint *Warp(2136), 2397), militating against perceiving a stimulus with an F3 value of 2136 as belonging to the category of 2397 Hz. The latter category corresponds to the phoneme /l/; for ease of exposition the cue-to-category constraints will be expressed with reference to phonemes, hence *Warp(2136), /l/) expresses that the F3 value 2136 should not be categorised as /l/. There is also a constraint *Warp(2136), /r/) that forbids mapping input with a 2136 F3 value to /r/ and a general constraint against not categorising the input, *Categ.

Hence, the perceptual input-output mapping of a stimulus with an F3 of 2137 Hz will proceed as shown in the tableau of Table 5.1 for neutral contexts, such as /p#e/.

<table>
<thead>
<tr>
<th>p[2136]e</th>
<th>*Categ</th>
<th>*Warp(2136), /l/)</th>
<th>*Warp(2136), /r/)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pre/</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>/p2136e/</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>/ple/</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 5.1: Perception without phonotactics: ambiguous input

For ambiguous input, the ranking of the two constraints directly implies the best output, namely /r/; the constraint against categorising the 2136 Hz F3 as /l/ is evaluated first, as it is ranked higher (shown in the left-to-right order of the constraints). The exclamation mark at the asterisk marking violation of *Warp(2136), /l/) by the output /l/ expresses that this is a fatal violation; there are better candidates. Hence, /l/ is out. In the end, only /r/ is allowed.

For perfect input, such as input with an F3 of 2397, i.e. a perfect /l/, the constraint against perceiving the input correctly, *Warp(2397), /l/), must be ranked very low, while the constraint against perceiving the input incorrectly, *Warp(2397), /r/) must be high. The tableau then changes to the one shown in Table 5.2.

<table>
<thead>
<tr>
<th>p[2397]e</th>
<th>*Categ</th>
<th>*Warp(2397), /l/)</th>
<th>*Warp(2136), /l/)</th>
<th>*Warp(2136), /r/)</th>
<th>*Warp(2397), /r/)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pre/</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p2397e/</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ple/</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2: Perception without phonotactics: perfect input. *Warp has been abbreviated to *W.
However, Massaro & Cohen (1983) found that the ambiguous stimuli were more likely to be perceived as the phoneme that was phonotactically legal, given the context. Hence, an ambiguous sound in the context s#e was more likely to be perceived as /l/ than as /r/. However, if s[2136]e is the input to the tableau in Table 5.1, the outcome would still be /sre/. If the example of a sound with an F3 of 2136 is assumed to be subject to a phonotactic influence, a constraint against /sr/ is enough to model this effect, as shown in Table 5.3.

<table>
<thead>
<tr>
<th>t[2136]e</th>
<th>*W([2397], /r/)</th>
<th>*/sr/</th>
<th>*W([2136], /l/)</th>
<th>*W([2136], /r/)</th>
<th>*W([2397], /l/)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sre/</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>/sle/</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Perception with phonotactics: ambiguous input. The option of not categorising and the constraint against it have been suppressed.

The phonotactic effect cannot prevent illegal perceptions if the input is a perfect match with an illegal combination, as shown in Table 5.4, now for the illegal combination /tl/, against which a constraint is also supposed to exist.

<table>
<thead>
<tr>
<th>t[2397]e</th>
<th>*W([2397], /r/)</th>
<th>*/tl/</th>
<th>*W([2136], /l/)</th>
<th>*W([2136], /r/)</th>
<th>*W([2397], /l/)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sre/</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>/tle/</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Perception with phonotactics: perfect input.

The example shows that phonotactic information can be ranked above the cues provided by ambiguous input, but not necessarily above perfect input. This was necessary to model the results of Massaro & Cohen (1983); with only cue constraints, the perception of ambiguous input would always be the same and the ambiguity would in fact not matter. Note that the perception grammar as envisaged by Boersma (1999) does not predict that ambiguity does not exist; it is meant to model the interaction of different types of cues, whereas the example only includes one. If the F3 cue is ambiguous, another phonetic cue might decide what the best perceptual output is. Phonotactic information thus acts like cue constraints.

This adaptation of perception grammar with phonotactic constraints can also explain categorical effects. If the constraint */tl/ was ranked at the top of the grammar in Table 5.4, the percept would never have been the illegal /tl/, however well the F3 matches the /l/.

Hence, an adaptation of perception grammar can capture gradient and categorical effects of phonotactic information on speech perception, when it comes to perceptual illusions. Nevertheless, the story for positive information, hence the bias to recognise wellformed combinations, is more complicated to capture in this type of grammar. Positive information can be added to the
model in the shape of contiguity constraints, in a fashion similar to Adriaans & Kager (2010). They propose that the high probability of some phoneme sequences should be taken into account in speech segmentation as well; otherwise, only marked structures would be split, but generalisations of markedness might become overly active in splitting if they are not curbed by opposing forces. Hence, contiguity constraints represent positive wellformedness, not markedness. However, they can in fact present a ban on splitting wellformed structures and function similarly to markedness constraints, in that they are assessed by violations, not by satisfactions.

Violations of positive constraints in speech perception would be odd. If a combination, e.g. /kæ/, is very wellformed, a constraint against not recognising it would be violated by many legal words. If such constraints were to force their biases on perception, they might change the input to the most wellformed word, which would pose a serious problem to the general faithfulness of perception to the input; e.g., an input of the form [fif] might then by more likely to be perceived as ‘cat’ than as ‘fish’, because the latter violates the constraint in favour of recognising /kæ/. Positive constraints would thus apparently be more useful as faithfulness constraints, that are only there to protect the very wellformed combinations in the input against being changed in the perceptual output. This would prevent such abrupt changes. However, this kind of faithfulness constraint for the mapping of features of the input to the output has to be supposed to exist in general. If there is markedness in the input, the perceptual output will be different unless a change in perception is banned by a high-ranked faithfulness constraint. So, faithfulness constraints protecting the phonemes /f/, /s/ and /f/ already stop the recognition of [fif] as ‘cat’. Similar constraints also assure the recognition of the phonemes of [kæt]. Hence, preferring the wellformed biphones in this word does not make this mapping more optimal, as it is already optimal, just as the recognition of /kæ/. Nevertheless, Vitevitch & Luce (1999) suggest that frequent biphone words such as /kæt/ should be easier to recognise, hence be more optimal.

There are two ways to make wellformed legal candidates more legal than marginally legal candidates. First, the optimal candidate that is the mapping [kæt]–/kæt/ can be compared against the optimal candidate [fif]–/fif/. As these candidates do not share input, one is not more optimal than another. However, as Coetzee (2009) suggests, both can be assessed on the same set of markedness constraints. Hence, [fif]–/fif/ must violate more important markedness constraints than [kæt]–/kæt/. This proposal as is thus entails that the apparently positive bias in favour of highly probable combinations such as

---

1 Boersma (1999) actually proposes a recognition grammar that links the perception /kæt/ to the word [kæt]. This step is left out for now, but note that this level models the influence of the probability of the word, by containing constraints against recognising less probable words.
is actually caused by a negative bias against improbable combinations such as /fi/. However, the positive constraint can now be reintroduced. A faithfulness constraint arguing in favour of recognising wellformed combinations might be unnecessary to make these combinations the optimal perceptual candidate, as no relevant markedness constraints are supposed to be violated by this candidate. However, if the wellformedness of the optimal candidate is not assessed just on the markedness constraints it violates, but also on the faithfulness constraints it fulfils, the positive effects of high probability are captured as well. Neither option is extremely attractive, as both do not explain the positive effect but merely posit it.

The role of probabilistic knowledge is quite different in speech segmentation. Here, phonotactic knowledge is surely not supposed to identify only one optimal candidate. The role of phonotactics is rather to guide perception in the direction that is most likely to yield a felicitous perception, or to steer perception away from fruitless attempts at lexical look-up. Other information, most notably phoneme recognition and lexical knowledge, should normally be enough to finish the perception process. The OT grammar that models the phonotactic preparser for speech segmentation should thus not eliminate all sub-optimal candidates, but just gradiently prefer the optimal one. The other candidates should also be assigned a wellformedness (or rather, harmony) value, in order to capture differences between the wellformedness of non-optimal candidates.

When a phonotactic grammar is used to rank perception candidates, instead of selecting only one candidate, positive information becomes as influential as negative information, if the proposed contiguity constraints of Adriaans & Kager (2010) are adopted. These constraints can be violated by segmentation candidates that put a boundary within a wellformed sound combinations. However, if there is a segmentation candidates that does not violate contiguity constraints, it is not guaranteed to be correct; a word can start in the middle of a wellformed cluster. Wellformed combinations have to be considered as word-internal in first instance, but not exclusively. Positive knowledge thus has to affect speech perception gradiently, not categorically.

A ranked constraint set is still an appropriate model of the phonotactic grammar containing positive information on wellformedness in the form of violable constraints, as well as markedness constraints. However, the gradience of such a grammar has to be addressed. Firstly, the grammar should rank its output candidates, i.e. the segmentation candidates. The optimal candidate cannot be seen as the only legal output.

Nevertheless, standard OT already makes it possible to rank candidates, when ‘optimality as legality’ is not taken into account. As described in the introduction, two output candidates for the same input are ranked relative to
each other if one is eliminated before the other, i.e. if one violates a constraint that the other does not and there is no higher ranked constraint for which this is the case. This can be visualised by assuming a large number of candidates, that are eliminated in small groups. The elimination is based on constraint violation; the candidates that violate the highest ranked constraint are eliminated first. If this elimination is not seen as a destruction, but a setting aside of candidates on a stack, the bottom of the stack contains the candidate that was eliminated first. This is the worst candidate, as it violates the most important constraint. After the evaluation of all the candidates against all constraints, the candidates that have not been eliminated can be placed on top of the stack.

The imaginary stack is now a ranking of the candidates. The best candidate is on top. Classic OT (Prince & Smolensky, 1993/2002) only considers this best candidate, as optimality entails legality, but non-optimal also entails illegality. However, if the optimal candidate does not turn out to be felicitous after lexical access, the evaluation can retrace its steps and yield a new optimal candidate, that only lost to the previously optimal candidate. This was proposed by Kager (2010) especially for speech segmentation, an optimal segmentation candidate might simply not correspond to a string of words.

The only thing not captured by a ranking generated with a classic OT grammar are the gradient differences between different candidates under different grammars. The Dutch participants in Experiment 4a considered the optimal segmentation for e.g. [sr] to be /s.r/, as did the Slavic participants in Experiment 4b. However, the candidate /sr/ was not considered as much by the Dutch as by the Slavic listeners in case the optimal segmentation did not yield a word (in the Non-word condition). If the /s.r/ solution is rejected, both languages would probably come up with /sr/ as the next option; at least, there are no obvious alternatives that could be better. Nevertheless, /sr/ is much worse under the Dutch phonotactic grammar than under the Slavic phonotactic grammar. The tableau in Table 5.5 shows the nature of the problem. There might be a contiguity constraint against the markedness of clusters in general, but such a constraint can never be ranked higher than a constraint against */sr/, otherwise the listeners would not prefer to split it. This means that the Dutch and Slavic grammars, tuned to their preference for segmentations aligned with the constraint */sr/, cannot explain the gradient difference between their assessments of /fi.srOk/.

This observation is similar to that of words containing highly-probable phoneme combinations, that are recognised more easily than words containing less probable phoneme combinations. However, this observation referred to different inputs to the same grammar. The difference between Dutch and Slavic just described is an instance of different grammars evaluating the same
Table 5.5: Segmentation with phonotactics: relative rankings. The optimal candidate in the top tableau is the best. Without this candidate, all the other candidates are optimal, hence their ranking at place 2.

<table>
<thead>
<tr>
<th>Input</th>
<th>*/sr/</th>
<th>Contiguity (CC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[fis,rok]</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>/\l,rok/</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>*/,rok/</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*/fis,rok/</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.5 shows the relative rankings of different input candidates with respect to the */sr/ constraint and contiguity (CC) constraints. The optimal candidate in the top tableau is the best. Without this candidate, all the other candidates are optimal, hence their ranking at place 2.

Input. It is desirable to not only assess whether one input-output mapping is better than another within the same input and grammar, but also to assign an absolute well-formedness value, which allows to compare input-output mappings for different inputs or different grammars.

In stochastic OT (Boersma, 1997), an output candidate has a probability of being optimal in a series of cases of the same input, but this does not directly translate to a well-formedness in an individual case, unless the evaluation process is assumed to be repeated many times. The latter can be assumed to have happened in the segmentation experiments given above and in the word recognition examples, as the results reported on these experiments are averaged over participants and items. This entails that in the Dutch grammar, the */sr/ constraint is ranked higher than in the Slavic grammar. The Slavic grammar has the */sr/ still close to competing contiguity constraints and in some evaluations, random noise on the ranking values can flip the order of the constraints and hence change the perceptual outcome.

Another option is to assign a harmony value to every candidate based on the harmony of the constraints they violate, as in Harmonic Grammar (Legendre et al., 1990). This entails that every constraint be assigned a value, with the most important ones having the highest value. Candidates are given a harmony value that is decreased with the value of each constraint it violates. The lower the number of violated constraints, the higher the harmony of the candidate. Harmonic Grammar is different from OT in exactly this aspect, which has the implication that Harmonic Grammar allows a large number of violations of a low-ranked constraint to add up to be more important than one violation of a high-ranked constraint, against OT’s principle of strict domination, where a candidate loses from another if it violates the highest-ranked constraint on which the two candidates differ, regardless of other differences in violation. With Harmonic Grammar, it is in principle possible to assume that contiguity constraints are violated by word boundaries, just as in
5.2 Phonotactic knowledge in speech perception

Table 5.6: Stochastic OT in segmentation. For Slavic, the two constraints are ranked close, at 100 and 102 in this fictional example. Random noise values are added that can cause the ranking to skip. In the Dutch grammar, the ranking values are assumed to be further apart, virtually excluding the possibility of inverted rankings.

<table>
<thead>
<tr>
<th>[fisr.ok]</th>
<th>*/sr/</th>
<th>Contiguity → 100 - 1 = 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>/l.srOk/</td>
<td>*/sr/</td>
<td>Contiguity(CC)</td>
</tr>
<tr>
<td>*/fisr.ok/</td>
<td>*/sr/</td>
<td>Contiguity(CC)</td>
</tr>
</tbody>
</table>

Table 5.7: Harmonic Grammar in segmentation. For Slavic, the two constraints are ranked close, at 2 and 1 in this fictional example. The difference between the best and second-best candidates is only a factor 2. If Dutch grammar has a higher value for the first constraint, the difference between the first and the second candidates will be larger.

<table>
<thead>
<tr>
<th>[fisr.ok]</th>
<th>*/sr/</th>
<th>Contiguity(CC)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>*/l.srOk/</td>
<td>*/sr/</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>/l.srOk/</td>
<td>*/sr/</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>*/fisr.ok/</td>
<td>*/sr/</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>*/fisr.ok/</td>
<td>*/sr/</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>
that input with wellformed sound combinations be more robustly recognised than input with less wellformed ones does not follow directly from any of the systems proposed above, even if it can in principle be added. Given the above, it seems that the effect of high probabilities should be captured at a higher level, namely at the level of actual word recognition. Nevertheless, this does not explain the finding that listeners in the Spanish group in Chapter 2 have representations of epenthesised clusters. These representations could be based on the frequency of epenthesised clusters after the filter, but this does not explain why unepenthesised frequent clusters do not cause more facilitation. This result remains unexplained. However, it should be noted that it depends on a null result: it is possible that there was more facilitation for unepenthesised clusters for more proficient listeners, but that it was not found.

5.3 SECOND LANGUAGE LISTENING

Phonotactic knowledge turns out to be language-dependent. In fact, language-dependency was found in all experiments, with the exception of Experiment 3 as its participants were only native listeners. However, language-mode effects were also found. The effects thus cannot have an alternative phonetic explanation.

The acquisition trajectory of phonotactic knowledge might be different for probabilistic knowledge and markedness knowledge. If markedness is categorical, i.e. in the case of illegality, this shapes perception of a second language. Even if illegal sound combinations are not illegal in the second language, listeners will filter these combinations out and therefore have trouble finding evidence against their markedness itself. This showed in the difficulty of Spanish native listeners that had acquired Dutch to lose the perceptual illusion of epenthesis, but it interestingly did not show in the Japanese native listeners group.

However, most of the results in this thesis indicate that it is normally possible to acquire the phonotactics of a second language and apply it to L2 perception. This corroborates findings for very proficient second language learners (interpreters), as found by Weber & Cutler (2006). The Dutch listeners tested in Experiment 5a were students that were proficient in English but not as proficient as the German interpreters of Weber & Cutler. Nevertheless, they applied English phonotactics to English speech and Dutch phonotactics to Dutch speech.

On the other hand, it might be difficult to reach native categorical differences. Categorical application of markedness constraints was not found in Experiment 4b, only gradient application of phonotactic knowledge that came
close to the native behaviour. Learning in OT from a set with wellformedness constraints that were present in L1 is expected to prefer promoting relevant constraints for L2, and not to add new constraints directly. This is in accordance with the second language learning model proposed by Escudero & Boersma (2004) (also Escudero, 2005). These authors propose that learning of a second language starts with a copy of the grammar of the first language. Given input, the new grammar is slowly adapted by reranking its constraints, until it matches the second language. It is possible that the L1 Slavic learners of Dutch that participated in Experiment 4b were in between Slavic legality and Dutch illegality. If so, they had been ranking their markedness constraints higher and higher while they were exposed to Dutch, but not yet high enough for them to categorically break up illegal clusters.

The Dutch learners of English in Experiment 5a showed that in second language learning, the two language’s phonotactics are indeed separate systems, at least as far as segmentation tasks are concerned. In addition, the results of Experiment 5a, as well as the results of Weber & Cutler (2006), show that illegality in a second language can be acquired even if it does not correspond to illegality in the first language at all. The illegality is nevertheless likely to be gradient during the acquisition trajectory.

In sum, the acquisition of second language phonotactics is predictable regarding markedness; although there is no direct test of the correctness of the Escudero & Boersma (2004) account of learning, all data reported in this thesis follow the general principle that a second language phonotactic grammar starts as a copy of the native language phonotactic grammar and is subsequently adapted to the second language. On the other hand, the acquisition of probabilistic phonotactics is less clear. Wellformedness is acquired for input that has already been changed by markedness constraints, but still seems to be specific enough to ignore the markedness constraint. More research is needed on the interaction between phonotactic illegality and the learning of phonotactic probabilities.

5.4 CONCLUSION

To model possibly conflicting wellformedness and illegality knowledge, the different functionality of the two was taken into account. Illegality provides cues about output that has to be avoided, while wellformedness improves efficiency by providing biases that are likely to speed up word recognition in continuous speech. A set of contiguity constraints, that argue against the breaking up of wellformed sound combinations that are present in the input, as well as a set of markedness constraints, that argue against marked sound combinations in the output, can be used to evaluate candidates for percep-
These candidates are then assigned an absolute wellformedness value, depending on the ranking in the grammar. This value allows the perceptual candidates to be ranked compared to other candidates that correspond to the same input, but also to input-output candidates generated with other grammars.

A constraint-based grammar performs the evaluation of output candidates against the set of constraints constituting phonotactic knowledge. The cues provided by this evaluation are effective in speech segmentation, but can also explain categorical perceptual illusions and phonotactic influences on the perception of sequences of ambiguous sound. In general, the perceptual output is a ranked set of candidates on a level before lexical access. Word recognition, as described by the ART and Shortlist models, can employ the candidate evaluation results as the best candidates contain valuable cues such as corrected errors, resolved ambiguities and probable word boundaries. The highest-ranked candidate is not optimal in the classic sense, it is only most probably the correct solution. As it is not guaranteed to be the only solution, less wellformed candidates are also submitted to lexical look-up. In this process, probabilistic knowledge of frequent sound combinations might be added by an a priori probability (Shortlist B), sublexical representations that aid resonance of the right words (ARTWORD), or by higher harmony values for wellformed combinations generated by the positive evaluation of faithfulness constraints by the phonotactic parser.

In sum, a listener’s knowledge of the sound combinations of his language defines how he perceives it. Phonotactic knowledge aids the speech recognition process at the prelexical level and also explains perception of non-lexical input, that cannot be resolved as a string of words. A phonotactic parser always maps raw acoustic input to a string of pseudowords. The pseudowords are normalised by other (phonetic) constraints in the perceptual grammar. The product of the phonotactic parser restricts the hypothesis space for word recognition and thus allows lexical material to connect more efficiently to the input, bypassing unnecessary computations such as the lexical look-up of a pseudoword with an illegal sound combination. The knowledge provided by phonotactics can be successfully modelled as an interaction of markedness and contiguity constraints in a perceptual grammar, that enhances the acoustic input with information that guides word recognition in the right direction.
ADDITINAL MATERIAL FOR CHAPTER 3

A.1 STIMULI

A.2 FULL STATISTICAL MODELS
<table>
<thead>
<tr>
<th>Aligned</th>
<th>Misaligned</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>f.kl</td>
<td>ft.l</td>
<td>8</td>
</tr>
<tr>
<td>f.pl</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>l.br</td>
<td>lm.r</td>
<td>5</td>
</tr>
<tr>
<td>l.dr</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>l.vr</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>n.kl</td>
<td>nt.l</td>
<td>5</td>
</tr>
<tr>
<td>n.pl</td>
<td>nt.l</td>
<td>8</td>
</tr>
<tr>
<td>r.bl</td>
<td>rm.l</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>m.l</td>
<td>6</td>
</tr>
<tr>
<td>r.vl</td>
<td>rm.l</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>m.l</td>
<td>4</td>
</tr>
<tr>
<td>m.l</td>
<td>.bl</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>.vl</td>
<td>7</td>
</tr>
<tr>
<td>m.r</td>
<td>.br</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>.dr</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>.vr</td>
<td>5</td>
</tr>
<tr>
<td>n.l</td>
<td>.bl</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>.vl</td>
<td>8</td>
</tr>
<tr>
<td>n.r</td>
<td>.br</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>.dr</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>.vr</td>
<td>4</td>
</tr>
<tr>
<td>t.l</td>
<td>.bl</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>.vl</td>
<td>2</td>
</tr>
</tbody>
</table>

Table A.1: Experiment 3: clusters in which targets have been embedded. Bold-face indicates the target end/beginning. The last column states the number of times an item of this form was used in the experiment.
## A.2 Full statistical models

Table A.2: Experiment 3: experimental items that were embedded in the Final position. The two contexts show the phonemes that preceded the Word or Non-word. The frequency count is derived from CELEX. Type: M for items copied from McQueen (1998), MA for items copied with adaptations, ME for items, based on the same specifications but newly constructed.
<table>
<thead>
<tr>
<th>Word</th>
<th>Orthography</th>
<th>IPA</th>
<th>Freq</th>
<th>Aligned Context Cluster</th>
<th>Misaligned Context Cluster</th>
<th>Non-word Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>duif</td>
<td>dœyf</td>
<td>823</td>
<td>klœym</td>
<td>fkl tloeym f1</td>
<td>tloeym f1 dref</td>
<td>M</td>
</tr>
<tr>
<td>graaf</td>
<td>xraf</td>
<td>1013</td>
<td>klœyf</td>
<td>fkl tlyf f1</td>
<td>tlyf f1 xuf</td>
<td>M</td>
</tr>
<tr>
<td>schijf</td>
<td>sxrif</td>
<td>541</td>
<td>klœym</td>
<td>fkl tlym f1</td>
<td>tlym f1 xef</td>
<td>M</td>
</tr>
<tr>
<td>druif</td>
<td>dreyf</td>
<td>1009</td>
<td>klœym</td>
<td>fkl tlam f1</td>
<td>tlam f1 dwhif</td>
<td>M</td>
</tr>
<tr>
<td>hof</td>
<td>haf</td>
<td>1315</td>
<td>klœif</td>
<td>fkl tlœif f1</td>
<td>tlœif f1 hyif</td>
<td>M</td>
</tr>
<tr>
<td>boef</td>
<td>buf</td>
<td>139</td>
<td>klœym</td>
<td>fkl tlœym f1</td>
<td>tlœym f1 bref</td>
<td>M</td>
</tr>
<tr>
<td>vijf</td>
<td>vef</td>
<td>787</td>
<td>klœeyf</td>
<td>fkl tlin f1</td>
<td>tlin f1 vryf</td>
<td>M</td>
</tr>
<tr>
<td>suf</td>
<td>svf</td>
<td>182</td>
<td>klœym</td>
<td>fkl tlim f1</td>
<td>tlim f1 snuf</td>
<td>M</td>
</tr>
<tr>
<td>scherf</td>
<td>sxref</td>
<td>436</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 xuf</td>
<td>M</td>
</tr>
<tr>
<td>gleuf</td>
<td>xlof</td>
<td>199</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 snuf</td>
<td>M</td>
</tr>
<tr>
<td>stal</td>
<td>staf</td>
<td>1236</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 xef</td>
<td>M</td>
</tr>
<tr>
<td>kuijf</td>
<td>krif</td>
<td>66</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 kruif</td>
<td>MA</td>
</tr>
<tr>
<td>korf</td>
<td>karf</td>
<td>118</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 kuif</td>
<td>MA</td>
</tr>
<tr>
<td>juf</td>
<td>jyf</td>
<td>207</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 jif</td>
<td>M</td>
</tr>
<tr>
<td>diuf</td>
<td>dif</td>
<td>572</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 duf</td>
<td>M</td>
</tr>
<tr>
<td>knol</td>
<td>knol</td>
<td>190</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 kod</td>
<td>M</td>
</tr>
<tr>
<td>vel</td>
<td>vel</td>
<td>1386</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 vroel</td>
<td>M</td>
</tr>
<tr>
<td>poel</td>
<td>pul</td>
<td>157</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 plul</td>
<td>M</td>
</tr>
<tr>
<td>fel</td>
<td>fel</td>
<td>2985</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 frul</td>
<td>M</td>
</tr>
<tr>
<td>haal</td>
<td>hal</td>
<td>191</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 hyul</td>
<td>M</td>
</tr>
<tr>
<td>spul</td>
<td>spyl</td>
<td>1391</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 stpl</td>
<td>M</td>
</tr>
<tr>
<td>kuil</td>
<td>kœyl</td>
<td>724</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 kruel</td>
<td>M</td>
</tr>
<tr>
<td>bol</td>
<td>bol</td>
<td>1208</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 bel</td>
<td>M</td>
</tr>
<tr>
<td>zaal</td>
<td>zœyl</td>
<td>667</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 zyl</td>
<td>M</td>
</tr>
<tr>
<td>nul</td>
<td>nvl</td>
<td>625</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 nul</td>
<td>M</td>
</tr>
<tr>
<td>bijl</td>
<td>bœyl</td>
<td>445</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 biu</td>
<td>M</td>
</tr>
<tr>
<td>wiel</td>
<td>wœil</td>
<td>891</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 wroel</td>
<td>M</td>
</tr>
<tr>
<td>puel</td>
<td>pœyl</td>
<td>1137</td>
<td>plœuif</td>
<td>fpl tlœuif f1</td>
<td>tlœuif f1 puel</td>
<td>M</td>
</tr>
<tr>
<td>pul</td>
<td>pul</td>
<td>156</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 prol</td>
<td>M</td>
</tr>
<tr>
<td>kool</td>
<td>kol</td>
<td>449</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 korl</td>
<td>M</td>
</tr>
<tr>
<td>dal</td>
<td>dœyl</td>
<td>1513</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 dol</td>
<td>M</td>
</tr>
<tr>
<td>boel</td>
<td>bul</td>
<td>907</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 boel</td>
<td>M</td>
</tr>
<tr>
<td>ziel</td>
<td>zœil</td>
<td>3324</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 zwol</td>
<td>M</td>
</tr>
<tr>
<td>taf</td>
<td>tal</td>
<td>1157</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 tral</td>
<td>M</td>
</tr>
<tr>
<td>pum</td>
<td>pœyn</td>
<td>413</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 pram</td>
<td>M</td>
</tr>
<tr>
<td>plein</td>
<td>pleyn</td>
<td>1402</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 pram</td>
<td>M</td>
</tr>
<tr>
<td>ren</td>
<td>ren</td>
<td>3707</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 ren</td>
<td>M</td>
</tr>
<tr>
<td>baan</td>
<td>han</td>
<td>717</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 hin</td>
<td>M</td>
</tr>
<tr>
<td>rein</td>
<td>riin</td>
<td>735</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 ran</td>
<td>M</td>
</tr>
<tr>
<td>schoen</td>
<td>sxun</td>
<td>2861</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 spin</td>
<td>M</td>
</tr>
<tr>
<td>kraan</td>
<td>kxeun</td>
<td>590</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 kyn</td>
<td>M</td>
</tr>
<tr>
<td>non</td>
<td>non</td>
<td>590</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 non</td>
<td>M</td>
</tr>
<tr>
<td>graan</td>
<td>xran</td>
<td>514</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 gras</td>
<td>M</td>
</tr>
<tr>
<td>boon</td>
<td>bon</td>
<td>645</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 bon</td>
<td>M</td>
</tr>
<tr>
<td>kern</td>
<td>kœrn</td>
<td>1842</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 knan</td>
<td>M</td>
</tr>
<tr>
<td>kraan</td>
<td>kxean</td>
<td>629</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 knan</td>
<td>M</td>
</tr>
<tr>
<td>baan</td>
<td>ban</td>
<td>3121</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 bran</td>
<td>M</td>
</tr>
<tr>
<td>snor</td>
<td>snœr</td>
<td>784</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 snor</td>
<td>M</td>
</tr>
<tr>
<td>kier</td>
<td>kir</td>
<td>561</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 kner</td>
<td>M</td>
</tr>
<tr>
<td>ster</td>
<td>stir</td>
<td>545</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 ster</td>
<td>M</td>
</tr>
<tr>
<td>kar</td>
<td>kor</td>
<td>683</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 ksr</td>
<td>M</td>
</tr>
<tr>
<td>spier</td>
<td>spœir</td>
<td>1521</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 spier</td>
<td>M</td>
</tr>
<tr>
<td>teer</td>
<td>ter</td>
<td>574</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 tar</td>
<td>M</td>
</tr>
<tr>
<td>zaar</td>
<td>zœar</td>
<td>1345</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 zar</td>
<td>M</td>
</tr>
<tr>
<td>geur</td>
<td>geœr</td>
<td>2947</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 geur</td>
<td>M</td>
</tr>
<tr>
<td>nier</td>
<td>nœr</td>
<td>500</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 nuer</td>
<td>M</td>
</tr>
<tr>
<td>ster</td>
<td>ster</td>
<td>2664</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 sur</td>
<td>M</td>
</tr>
<tr>
<td>boer</td>
<td>bur</td>
<td>4245</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 boer</td>
<td>M</td>
</tr>
<tr>
<td>schuur</td>
<td>sxœur</td>
<td>922</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 sxer</td>
<td>M</td>
</tr>
<tr>
<td>bier</td>
<td>bir</td>
<td>2725</td>
<td>plœyf</td>
<td>fpl tlœyf f1</td>
<td>tlœyf f1 bier</td>
<td>M</td>
</tr>
</tbody>
</table>

Table A.3: Experiment 3: experimental items that were embedded in the Initial position. The two contexts show the phonemes that followed the Word / Non-word. Frequency counts are derived from CELEX. Type: M for items copied from McQueen (1998), MA for items copied with adaptations, ME for items based on the same specifications but newly constructed.
### A.2 Full statistical models

<table>
<thead>
<tr>
<th>Orthography</th>
<th>IPA</th>
<th>Freq</th>
<th>Aligned Context</th>
<th>Misaligned Context</th>
<th>Non-word</th>
</tr>
</thead>
<tbody>
<tr>
<td>rem</td>
<td>rem</td>
<td>307</td>
<td>wot tr</td>
<td>wod dr</td>
<td>rrw</td>
</tr>
<tr>
<td>ren</td>
<td>ren</td>
<td>3707</td>
<td>vit tr</td>
<td>vid dr</td>
<td>rrij</td>
</tr>
<tr>
<td>rib</td>
<td>rep</td>
<td>391</td>
<td>lty tr</td>
<td>lyl dr</td>
<td>rdx</td>
</tr>
<tr>
<td>riem</td>
<td>rim</td>
<td>939</td>
<td>veit tr</td>
<td>veid dr</td>
<td>rix</td>
</tr>
<tr>
<td>rrij</td>
<td>rri</td>
<td>3280</td>
<td>wot tr</td>
<td>wod dr</td>
<td>ro</td>
</tr>
<tr>
<td>roos</td>
<td>ros</td>
<td>1213</td>
<td>veit tr</td>
<td>veid dr</td>
<td>ro</td>
</tr>
<tr>
<td>lap</td>
<td>lap</td>
<td>662</td>
<td>syt tl</td>
<td>syd dl</td>
<td>loj</td>
</tr>
<tr>
<td>lelf</td>
<td>lef</td>
<td>5449</td>
<td>rvt tl</td>
<td>rvd dl</td>
<td>lik</td>
</tr>
<tr>
<td>lies</td>
<td>lis</td>
<td>624</td>
<td>not tl</td>
<td>nod dl</td>
<td>li</td>
</tr>
<tr>
<td>lolf</td>
<td>lof</td>
<td>586</td>
<td>jut tl</td>
<td>jad dl</td>
<td>lax</td>
</tr>
<tr>
<td>lomp</td>
<td>lom</td>
<td>328</td>
<td>syt tl</td>
<td>syd dl</td>
<td>lan</td>
</tr>
<tr>
<td>loon</td>
<td>lon</td>
<td>1127</td>
<td>sut tl</td>
<td>sud dl</td>
<td>loj</td>
</tr>
<tr>
<td>lam</td>
<td>lam</td>
<td>390</td>
<td>vsyp pl</td>
<td>vsyb bl</td>
<td>lar</td>
</tr>
<tr>
<td>leak</td>
<td>lek</td>
<td>4026</td>
<td>top pl</td>
<td>tob bl</td>
<td>lex</td>
</tr>
<tr>
<td>leas</td>
<td>leas</td>
<td>143</td>
<td>fop pl</td>
<td>Fsb bl</td>
<td>lef</td>
</tr>
<tr>
<td>leer</td>
<td>leer</td>
<td>200</td>
<td>dop pl</td>
<td>dab bl</td>
<td>luk</td>
</tr>
<tr>
<td>long</td>
<td>long</td>
<td>867</td>
<td>fyp pl</td>
<td>fyb bl</td>
<td>luj</td>
</tr>
<tr>
<td>loom</td>
<td>loom</td>
<td>350</td>
<td>kep pl</td>
<td>keb bl</td>
<td>lo</td>
</tr>
<tr>
<td>waas</td>
<td>waas</td>
<td>267</td>
<td>hep pw</td>
<td>heb bw</td>
<td>waf</td>
</tr>
<tr>
<td>wal</td>
<td>wal</td>
<td>830</td>
<td>dryp pw</td>
<td>dryb bw</td>
<td>wap</td>
</tr>
<tr>
<td>wiel</td>
<td>wiel</td>
<td>501</td>
<td>tropp pw</td>
<td>trob bw</td>
<td>wuk</td>
</tr>
<tr>
<td>wier</td>
<td>wier</td>
<td>1465</td>
<td>top pw</td>
<td>tob bw</td>
<td>wus</td>
</tr>
<tr>
<td>wol</td>
<td>wol</td>
<td>423</td>
<td>myp pw</td>
<td>myb bw</td>
<td>wuf</td>
</tr>
<tr>
<td>wond</td>
<td>wond</td>
<td>823</td>
<td>doup pw</td>
<td>doub bw</td>
<td>wuj</td>
</tr>
<tr>
<td>wang</td>
<td>wang</td>
<td>1185</td>
<td>forys sw</td>
<td>foryz aw</td>
<td>wum</td>
</tr>
<tr>
<td>wieg</td>
<td>wieg</td>
<td>450</td>
<td>bes sw</td>
<td>bez zw</td>
<td>wum</td>
</tr>
<tr>
<td>wolf</td>
<td>wolf</td>
<td>727</td>
<td>tys sw</td>
<td>tylz aw</td>
<td>war</td>
</tr>
<tr>
<td>walk</td>
<td>walk</td>
<td>2040</td>
<td>juss sw</td>
<td>juz aw</td>
<td>waf</td>
</tr>
<tr>
<td>worm</td>
<td>worm</td>
<td>427</td>
<td>xes sw</td>
<td>xez aw</td>
<td>war</td>
</tr>
<tr>
<td>woud</td>
<td>woud</td>
<td>971</td>
<td>doxs sw</td>
<td>doxyz zw</td>
<td>wouj</td>
</tr>
<tr>
<td>lalf</td>
<td>lalf</td>
<td>430</td>
<td>rouz sl</td>
<td>roues sl</td>
<td>lan</td>
</tr>
<tr>
<td>lauw</td>
<td>lauw</td>
<td>582</td>
<td>zos sl</td>
<td>zos sl</td>
<td>lauw</td>
</tr>
<tr>
<td>lied</td>
<td>lied</td>
<td>1578</td>
<td>klyz sl</td>
<td>klzys sl</td>
<td>li</td>
</tr>
<tr>
<td>lier</td>
<td>lier</td>
<td>137</td>
<td>not sl</td>
<td>nos sl</td>
<td>lin</td>
</tr>
<tr>
<td>lift</td>
<td>lift</td>
<td>1189</td>
<td>pot sl</td>
<td>pos sl</td>
<td>lif</td>
</tr>
<tr>
<td>lucht</td>
<td>lucht</td>
<td>7735</td>
<td>bes al</td>
<td>bes al</td>
<td>lyx</td>
</tr>
<tr>
<td>reep</td>
<td>reep</td>
<td>245</td>
<td>tex zr</td>
<td>tch br</td>
<td>rrw</td>
</tr>
<tr>
<td>rek</td>
<td>rek</td>
<td>367</td>
<td>lvy zr</td>
<td>lvy br</td>
<td>ref</td>
</tr>
<tr>
<td>riet</td>
<td>riet</td>
<td>613</td>
<td>fuy zr</td>
<td>fub br</td>
<td>rin</td>
</tr>
<tr>
<td>rit</td>
<td>rit</td>
<td>540</td>
<td>saaz zr</td>
<td>sah br</td>
<td>rxi</td>
</tr>
<tr>
<td>roet</td>
<td>roet</td>
<td>146</td>
<td>kloaz zr</td>
<td>klob br</td>
<td>row</td>
</tr>
<tr>
<td>rund</td>
<td>rund</td>
<td>191</td>
<td>kouaz zr</td>
<td>kubb br</td>
<td>rvp</td>
</tr>
<tr>
<td>raaf</td>
<td>raaf</td>
<td>251</td>
<td>bes sr</td>
<td>bep pr</td>
<td>hon</td>
</tr>
<tr>
<td>rat</td>
<td>rat</td>
<td>1409</td>
<td>forys sr</td>
<td>foryp pr</td>
<td>run</td>
</tr>
<tr>
<td>ring</td>
<td>ring</td>
<td>1442</td>
<td>zos sr</td>
<td>zoup pr</td>
<td>rj</td>
</tr>
<tr>
<td>rok</td>
<td>rok</td>
<td>1378</td>
<td>fas sr</td>
<td>fop pr</td>
<td>rj</td>
</tr>
<tr>
<td>room</td>
<td>room</td>
<td>232</td>
<td>tos sr</td>
<td>top pr</td>
<td>roj</td>
</tr>
<tr>
<td>rot</td>
<td>rot</td>
<td>593</td>
<td>mos sr</td>
<td>mep pr</td>
<td>roj</td>
</tr>
</tbody>
</table>

Table A.4: Experiment 4a and Experiment 4b: experimental items. The two contexts show the phonemes that preceded the Word or Non-word. The frequency count is derived from CELEX.
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.706 (0.224)</td>
<td>-12.07</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>-0.770 (0.035)</td>
<td>21.80</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>-0.137 (0.036)</td>
<td>-3.79</td>
<td>0.0002***</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>0.357 (0.051)</td>
<td>6.94</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Linear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>60.298 (7.452)</td>
<td>8.09</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>-10.169 (1.663)</td>
<td>-6.34</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>-31.694 (0.321)</td>
<td>-98.86</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>9.174 (1.619)</td>
<td>5.66</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Quadratic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>23.333 (4.703)</td>
<td>4.96</td>
<td>0.0002***</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>1.853 (1.266)</td>
<td>1.45</td>
<td>0.1455</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>-11.180 (1.231)</td>
<td>-9.08</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>7.111 (1.777)</td>
<td>4.00</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Cubic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>8.527 (2.750)</td>
<td>3.10</td>
<td>0.0016 **</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>-6.424 (1.275)</td>
<td>-5.04</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>-1.486 (1.141)</td>
<td>-1.30</td>
<td>0.1930</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>2.245 (1.707)</td>
<td>1.32</td>
<td>0.1884</td>
</tr>
</tbody>
</table>

### Table A.6: Experiment 3 distractor fixations, modelled logit of correspondence to Phonotactic condition. Bonferroni corrected for two comparisons. Log-likelihood: -16044 (54216 observations).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-2.720 (0.388)</td>
<td>-7.0</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>5.765 (0.085)</td>
<td>67.7</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>0.666 (0.072)</td>
<td>9.2</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Misaligned x Word</td>
<td>-0.352 (0.098)</td>
<td>-3.6</td>
<td>0.00032***</td>
</tr>
<tr>
<td><strong>Linear</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-19.654 (10.626)</td>
<td>-1.8</td>
<td>0.06437</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>39.702 (3.968)</td>
<td>10.0</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>-4.731 (3.408)</td>
<td>-1.4</td>
<td>&lt;0.1650</td>
</tr>
<tr>
<td>Misaligned x Word</td>
<td>-12.796 (4.627)</td>
<td>-2.8</td>
<td>0.00568*</td>
</tr>
<tr>
<td><strong>Quadratic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>9.314 (9.668)</td>
<td>1.0</td>
<td>0.33533</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>25.671 (2.863)</td>
<td>9.0</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>11.992 (2.543)</td>
<td>4.7</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Misaligned x Word</td>
<td>7.906 (3.430)</td>
<td>2.3</td>
<td>0.02119*</td>
</tr>
<tr>
<td><strong>Cubic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>6.307 (6.123)</td>
<td>1.0</td>
<td>0.29616</td>
</tr>
<tr>
<td>Phonotactics-Misaligned</td>
<td>7.406 (3.218)</td>
<td>-1.1</td>
<td>0.27667</td>
</tr>
<tr>
<td>Lexicality-Word</td>
<td>6.003 (3.871)</td>
<td>-2.1</td>
<td>&lt;0.0366</td>
</tr>
<tr>
<td>Misaligned x Word</td>
<td>6.430 (3.878)</td>
<td>-1.7</td>
<td>0.0973</td>
</tr>
</tbody>
</table>
Table A.7: Experiment 4A modelled logit of target fixations. Significance markers are Bonferroni corrected for two comparisons and four time levels (α = 0.05/8). Log-likelihood: -72321 (214266 observations). The model also took the different fixation levels for the cluster pairs into account as a fixed effect (as not many more cluster pairs are possible), as well as a Lexicality main effect, and the values are thus corrected for these effects.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 tr-dr:Misaligned</td>
<td>-0.567 (0.061)</td>
<td>-9.216</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>A2 *tl-*dl:Misaligned</td>
<td>-0.860 (0.066)</td>
<td>-12.052</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>B1 pl-bl:Misaligned</td>
<td>0.132 (0.076)</td>
<td>1.728</td>
<td>0.0841</td>
</tr>
<tr>
<td>B2 *pw–*bw:Misaligned</td>
<td>-0.953 (0.087)</td>
<td>-10.775</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>C1 *sw–zw:Misaligned</td>
<td>-0.048 (0.066)</td>
<td>-0.695</td>
<td>0.4873</td>
</tr>
<tr>
<td>C2 *z1–z2:Misaligned</td>
<td>-0.260 (0.086)</td>
<td>-3.037</td>
<td>0.0024 *</td>
</tr>
<tr>
<td>D1 *sr–pr:Misaligned</td>
<td>-0.319 (0.071)</td>
<td>-4.472</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>D2 *sr–pr:Misaligned</td>
<td>-0.747 (0.075)</td>
<td>-10.008</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Linear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 tr-dr:Misaligned</td>
<td>5.198 (2.463)</td>
<td>2.111</td>
<td>0.0348</td>
</tr>
<tr>
<td>A2 *tl-*dl:Misaligned</td>
<td>-19.112 (2.652)</td>
<td>7.208</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>B1 pl-bl:Misaligned</td>
<td>-7.058 (3.405)</td>
<td>2.089</td>
<td>0.0222</td>
</tr>
<tr>
<td>B2 *pw–*bw:Misaligned</td>
<td>38.150 (3.475)</td>
<td>10.980</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>C1 *sw–zw:Misaligned</td>
<td>3.473 (2.805)</td>
<td>1.238</td>
<td>0.2195</td>
</tr>
<tr>
<td>C2 *z1–z2:Misaligned</td>
<td>-18.362 (3.517)</td>
<td>-5.220</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>D1 *sr–pr:Misaligned</td>
<td>-22.680 (2.793)</td>
<td>-8.114</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>D2 *sr–pr:Misaligned</td>
<td>4.104 (3.023)</td>
<td>1.357</td>
<td>0.1746</td>
</tr>
<tr>
<td>Quadratic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 tr-dr:Misaligned</td>
<td>-21.475 (2.280)</td>
<td>-9.420</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>A2 *tl-*dl:Misaligned</td>
<td>25.234 (2.539)</td>
<td>9.940</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>B1 pl-bl:Misaligned</td>
<td>-2.842 (2.745)</td>
<td>-1.035</td>
<td>0.3064</td>
</tr>
<tr>
<td>B2 *pw–*bw:Misaligned</td>
<td>-12.617 (2.662)</td>
<td>-4.659</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>C1 *sw–zw:Misaligned</td>
<td>-12.735 (2.464)</td>
<td>-5.168</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>C2 *z1–z2:Misaligned</td>
<td>11.265 (2.958)</td>
<td>3.809</td>
<td>0.0003 ***</td>
</tr>
<tr>
<td>D1 *sr–pr:Misaligned</td>
<td>1.565 (2.564)</td>
<td>0.610</td>
<td>0.5420</td>
</tr>
<tr>
<td>D2 *sr–pr:Misaligned</td>
<td>-8.365 (2.764)</td>
<td>-3.029</td>
<td>0.0025 *</td>
</tr>
<tr>
<td>Cubic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 tr-dr:Misaligned</td>
<td>14.189 (1.753)</td>
<td>8.097</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>A2 *tl-*dl:Misaligned</td>
<td>-6.336 (1.870)</td>
<td>-3.372</td>
<td>0.0007 **</td>
</tr>
<tr>
<td>B1 pl-bl:Misaligned</td>
<td>-1.336 (1.964)</td>
<td>-0.673</td>
<td>0.5007</td>
</tr>
<tr>
<td>B2 *pw–*bw:Misaligned</td>
<td>1.134 (1.052)</td>
<td>1.081</td>
<td>0.2814</td>
</tr>
<tr>
<td>C1 *sw–zw:Misaligned</td>
<td>4.167 (1.794)</td>
<td>2.362</td>
<td>0.0182</td>
</tr>
<tr>
<td>C2 *z1–z2:Misaligned</td>
<td>-8.178 (1.662)</td>
<td>-4.919</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>D1 *sr–pr:Misaligned</td>
<td>16.339 (1.660)</td>
<td>9.357</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>D2 *sr–pr:Misaligned</td>
<td>3.776 (2.117)</td>
<td>1.784</td>
<td>0.0745</td>
</tr>
</tbody>
</table>
### Table A.8: Experiment 4a distractor fixations, modelled logit of correspondence to Phonotactic condition, i.e. looking at distractor present in auditory stream when looking at any of the two distractors. Significance markers are Bonferroni corrected for two comparisons and four time levels ($\alpha = 0.05/8$). Log-likelihood: -32049 (69312 observations). The model also took the different fixation levels for the cluster pairs into account as a fixed effect (as not many more cluster pairs are possible), as well as a Lexicality main effect, and the values are thus corrected for these effects.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 tr-drMisaligned</td>
<td>2.5816 (0.1366)</td>
<td>17.510</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>A2 *fl-*dlMisaligned</td>
<td>2.1133 (0.0933)</td>
<td>22.660</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>B1 pl-blMisaligned</td>
<td>1.8096 (0.0880)</td>
<td>20.560</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>B2 *pw-*bwMisaligned</td>
<td>4.5502 (0.1303)</td>
<td>34.930</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>C1 *sw-*zwMisaligned</td>
<td>3.8537 (0.1037)</td>
<td>37.170</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>C2 *zl-*slMisaligned</td>
<td>4.7717 (0.1229)</td>
<td>38.830</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>D1 *zr-*brMisaligned</td>
<td>3.2890 (0.1209)</td>
<td>29.690</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>D2 *se-prMisaligned</td>
<td>1.9951 (0.1019)</td>
<td>19.580</td>
<td>&lt; 0.0001 ***</td>
</tr>
</tbody>
</table>

#### Linear

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 tr-drMisaligned</td>
<td>-46.1180 (5.4796)</td>
<td>-8.420</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>A2 *fl-*dlMisaligned</td>
<td>9.8037 (3.7725)</td>
<td>2.600</td>
<td>0.0094</td>
</tr>
<tr>
<td>B1 pl-blMisaligned</td>
<td>14.4666 (3.8179)</td>
<td>4.000</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>B2 *pw-*bwMisaligned</td>
<td>-46.4621 (5.4443)</td>
<td>-8.860</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>C1 *sw-*zwMisaligned</td>
<td>-17.1555 (4.0958)</td>
<td>-4.230</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>C2 *zl-*slMisaligned</td>
<td>-35.8726 (5.0543)</td>
<td>-7.100</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>D1 *zr-*brMisaligned</td>
<td>35.3434 (4.3323)</td>
<td>8.120</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>D2 *se-prMisaligned</td>
<td>-30.3520 (4.1377)</td>
<td>-7.310</td>
<td>&lt; 0.0001 ***</td>
</tr>
</tbody>
</table>

#### Quadratic

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 tr-drMisaligned</td>
<td>39.8640 (4.5946)</td>
<td>8.680</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>A2 *fl-*dlMisaligned</td>
<td>4.9217 (3.4617)</td>
<td>1.420</td>
<td>0.1551</td>
</tr>
<tr>
<td>B1 pl-blMisaligned</td>
<td>12.1421 (3.2872)</td>
<td>3.690</td>
<td>0.0002 **</td>
</tr>
<tr>
<td>B2 *pw-*bwMisaligned</td>
<td>14.3790 (4.7143)</td>
<td>3.040</td>
<td>0.0024 *</td>
</tr>
<tr>
<td>C1 *sw-*zwMisaligned</td>
<td>32.3607 (3.6510)</td>
<td>8.860</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>C2 *zl-*slMisaligned</td>
<td>43.1734 (4.2605)</td>
<td>10.130</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>D1 *zr-*brMisaligned</td>
<td>18.6571 (4.0776)</td>
<td>4.560</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>D2 *se-prMisaligned</td>
<td>3.6175 (3.7266)</td>
<td>0.960</td>
<td>0.3371</td>
</tr>
</tbody>
</table>

#### Cubic

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 tr-drMisaligned</td>
<td>-6.9191 (3.1331)</td>
<td>-2.210</td>
<td>0.0271</td>
</tr>
<tr>
<td>A2 *fl-*dlMisaligned</td>
<td>11.7241 (2.9770)</td>
<td>3.940</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>B1 pl-blMisaligned</td>
<td>0.7366 (2.7550)</td>
<td>0.270</td>
<td>0.7862</td>
</tr>
<tr>
<td>B2 *pw-*bwMisaligned</td>
<td>-9.5319 (3.4211)</td>
<td>-2.900</td>
<td>0.0037 *</td>
</tr>
<tr>
<td>C1 *sw-*zwMisaligned</td>
<td>-26.9033 (3.9063)</td>
<td>-9.080</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>C2 *zl-*slMisaligned</td>
<td>-6.0210 (3.3333)</td>
<td>-1.810</td>
<td>0.0709</td>
</tr>
<tr>
<td>D1 *zr-*brMisaligned</td>
<td>-13.4616 (3.9668)</td>
<td>-3.370</td>
<td>0.0008 **</td>
</tr>
<tr>
<td>D2 *se-prMisaligned</td>
<td>16.0848 (2.7264)</td>
<td>5.950</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>Predictor</td>
<td>Estimate (SE)</td>
<td>z-value</td>
<td>p-value</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.3799 (0.2984)</td>
<td>-4.620</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>0.1096 (0.5688)</td>
<td>0.200</td>
<td>0.8364</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>0.1121 (0.0647)</td>
<td>1.730</td>
<td>0.0834</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>-0.0811 (0.0706)</td>
<td>-1.150</td>
<td>0.2510</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>0.1372 (0.0920)</td>
<td>1.490</td>
<td>0.1360</td>
</tr>
<tr>
<td>Linear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>7.0135 (6.3994)</td>
<td>1.100</td>
<td>0.2708</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>-1.2191 (8.1892)</td>
<td>-1.370</td>
<td>0.1715</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>15.9775 (2.5759)</td>
<td>6.200</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>-4.9270 (2.8149)</td>
<td>-1.720</td>
<td>0.0854</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>-15.1237 (3.6752)</td>
<td>-4.120</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>Quadratic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-6.4637 (5.7317)</td>
<td>-1.110</td>
<td>0.2685</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>3.6901 (8.0848)</td>
<td>0.450</td>
<td>0.6533</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>5.9800 (4.5555)</td>
<td>2.280</td>
<td>0.0229</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>20.3298 (2.6721)</td>
<td>7.620</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>-10.7174 (3.5066)</td>
<td>-3.060</td>
<td>0.0022 *</td>
</tr>
<tr>
<td>Cubic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-1.1181 (3.0177)</td>
<td>-0.370</td>
<td>0.7140</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>4.7385 (4.0144)</td>
<td>1.180</td>
<td>0.2388</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>-10.9891 (1.9724)</td>
<td>-5.570</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>-5.8589 (2.1288)</td>
<td>-2.710</td>
<td>0.0695 *</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>2.7723 (2.8147)</td>
<td>0.980</td>
<td>0.3247</td>
</tr>
</tbody>
</table>

Table A.9: Experiment 4b modelled logit of target fixations. Significance markers are Bonferroni corrected for two comparisons and four time levels ($\alpha = 0.05/8$). Log-likelihood: -92090 (291048 observations).
### A.2 Full Statistical Models

#### Table A.10: Experiment 4b: distractor fixations, modelled logit of looking to the distractor that is present in the auditory stimulus when looking at one of the two distractors (the other distractor is then not present in the auditory material). For Legal clusters, this distractor corresponds to the onset of a phonotactically legal chunk; for Illegal clusters it does not. Log-likelihood: -16044 (54216 observations). Significance markers are Bonferroni corrected for two comparisons and four time levels (\( \alpha = 0.05/8 \)). Log-likelihood: -56473 (103694 observations).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (SE)</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-2.9215 (0.3584)</td>
<td>-8.151</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>-0.8175 (0.5143)</td>
<td>-1.589</td>
<td>0.1120</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>0.4814 (0.0579)</td>
<td>8.615</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>-0.1371 (0.0646)</td>
<td>-2.433</td>
<td>0.0150</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>-0.4749 (0.0842)</td>
<td>-5.641</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td><strong>Linear</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>52.6608 (12.9955)</td>
<td>4.067</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>10.8885 (18.7149)</td>
<td>0.582</td>
<td>0.5607</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>-3.9958 (2.2322)</td>
<td>-1.741</td>
<td>0.0817</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>-7.3219 (2.6185)</td>
<td>-2.796</td>
<td>0.0052 *</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>8.4350 (3.4073)</td>
<td>2.476</td>
<td>0.0133</td>
</tr>
<tr>
<td><strong>Quadratic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-17.2628 (7.2546)</td>
<td>-2.380</td>
<td>0.0173</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>-14.6842 (10.0635)</td>
<td>-1.459</td>
<td>0.1445</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>0.0157 (2.1298)</td>
<td>0.242</td>
<td>0.8087</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>8.3945 (2.4157)</td>
<td>3.475</td>
<td>0.0005 **</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>-14.7093 (3.1407)</td>
<td>-4.683</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td><strong>Cubic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-0.3221 (2.9184)</td>
<td>-0.110</td>
<td>0.9121</td>
</tr>
<tr>
<td>L1-Slavic</td>
<td>1.2821 (3.8793)</td>
<td>0.305</td>
<td>0.7606</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal</td>
<td>8.5073 (1.6472)</td>
<td>5.165</td>
<td>&lt; 0.0001 ***</td>
</tr>
<tr>
<td>L1-Slavic \times Lexicality-Word</td>
<td>5.0821 (1.8296)</td>
<td>2.778</td>
<td>0.0055 *</td>
</tr>
<tr>
<td>L1-Slavic \times Cluster-Legal \times Lexicality-Word</td>
<td>-0.0322 (2.4078)</td>
<td>-0.013</td>
<td>0.9893</td>
</tr>
<tr>
<td>Predictor</td>
<td>Estimate (SE)</td>
<td>z-value</td>
<td>p-value</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.901 (0.240)</td>
<td>3.76</td>
<td>0.00017***</td>
</tr>
<tr>
<td>Dutch-Misaligned</td>
<td>-0.524 (0.058)</td>
<td>-9.03</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Misaligned</td>
<td>-0.165 (0.060)</td>
<td>-2.72</td>
<td>0.00645**</td>
</tr>
<tr>
<td>Word</td>
<td>-0.935 (0.063)</td>
<td>-15.40</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>English</td>
<td>-0.651 (0.039)</td>
<td>-1.64</td>
<td>0.05236</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word</td>
<td>0.029 (0.080)</td>
<td>0.37</td>
<td>0.71409</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>-0.053 (0.094)</td>
<td>-0.62</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Dutch-Misaligned × English</td>
<td>1.344 (0.086)</td>
<td>15.67</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Misaligned × English</td>
<td>-0.218 (0.080)</td>
<td>-2.72</td>
<td>0.00605**</td>
</tr>
<tr>
<td>Word × English</td>
<td>-0.975 (0.081)</td>
<td>-11.75</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word × English</td>
<td>-0.497 (0.117)</td>
<td>-4.33</td>
<td>0.000002***</td>
</tr>
<tr>
<td>Misaligned × Word × English</td>
<td>1.345 (0.126)</td>
<td>11.23</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td><strong>Linear</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>33.485 (7.516)</td>
<td>4.46</td>
<td>0.00001***</td>
</tr>
<tr>
<td>Dutch-Misaligned</td>
<td>-35.654 (5.987)</td>
<td>-5.95</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Misaligned</td>
<td>-28.245 (2.053)</td>
<td>-13.67</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Word</td>
<td>-28.616 (2.066)</td>
<td>-13.65</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>English</td>
<td>33.433 (10.496)</td>
<td>3.18</td>
<td>0.00345**</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word</td>
<td>40.453 (2.722)</td>
<td>14.86</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>70.814 (3.346)</td>
<td>21.17</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Dutch-Misaligned × English</td>
<td>11.475 (3.056)</td>
<td>3.75</td>
<td>0.000017***</td>
</tr>
<tr>
<td>Misaligned × English</td>
<td>13.497 (2.763)</td>
<td>4.88</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Word × English</td>
<td>29.962 (2.865)</td>
<td>10.68</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word × English</td>
<td>-13.707 (4.127)</td>
<td>-3.22</td>
<td>0.00090***</td>
</tr>
<tr>
<td>Misaligned × Word × English</td>
<td>-71.809 (4.214)</td>
<td>-17.06</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td><strong>Quadratic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.978 (6.624)</td>
<td>0.01</td>
<td>0.99979</td>
</tr>
<tr>
<td>Dutch-Misaligned</td>
<td>-1.884 (1.883)</td>
<td>-1.00</td>
<td>0.31729</td>
</tr>
<tr>
<td>Misaligned</td>
<td>-2.239 (1.918)</td>
<td>-1.18</td>
<td>0.23747</td>
</tr>
<tr>
<td>Word</td>
<td>-3.228 (1.665)</td>
<td>-1.90</td>
<td>0.05977</td>
</tr>
<tr>
<td>English</td>
<td>-5.823 (2.251)</td>
<td>-2.63</td>
<td>0.000002***</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word</td>
<td>6.435 (5.952)</td>
<td>2.00</td>
<td>0.01245**</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>-17.649 (3.010)</td>
<td>-5.86</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Dutch-Misaligned × English</td>
<td>20.308 (2.864)</td>
<td>7.09</td>
<td>0.000000***</td>
</tr>
<tr>
<td>Misaligned × English</td>
<td>-0.714 (1.686)</td>
<td>-0.42</td>
<td>0.67356</td>
</tr>
<tr>
<td>Word × English</td>
<td>11.067 (2.631)</td>
<td>4.21</td>
<td>0.000003***</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word × English</td>
<td>-28.842 (3.868)</td>
<td>-7.46</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td>Misaligned × Word × English</td>
<td>22.710 (3.370)</td>
<td>6.70</td>
<td>&lt; 0.00001***</td>
</tr>
<tr>
<td><strong>Cubic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-6.209 (4.221)</td>
<td>-1.47</td>
<td>0.14127</td>
</tr>
<tr>
<td>Dutch-Misaligned</td>
<td>6.339 (1.662)</td>
<td>3.81</td>
<td>0.000014***</td>
</tr>
<tr>
<td>Misaligned</td>
<td>-6.355 (1.700)</td>
<td>-3.73</td>
<td>0.000010***</td>
</tr>
<tr>
<td>Word</td>
<td>7.396 (1.740)</td>
<td>4.21</td>
<td>0.000002***</td>
</tr>
<tr>
<td>English</td>
<td>8.641 (1.884)</td>
<td>4.57</td>
<td>0.14104</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word</td>
<td>-3.645 (2.321)</td>
<td>-1.59</td>
<td>0.11137</td>
</tr>
<tr>
<td>Misaligned × Word</td>
<td>7.105 (2.624)</td>
<td>2.74</td>
<td>0.00811**</td>
</tr>
<tr>
<td>Dutch-Misaligned × English</td>
<td>-22.071 (4.410)</td>
<td>-5.12</td>
<td>&lt; 0.000001***</td>
</tr>
<tr>
<td>Misaligned × English</td>
<td>-3.208 (2.279)</td>
<td>-1.41</td>
<td>0.15917</td>
</tr>
<tr>
<td>Word × English</td>
<td>-15.858 (3.131)</td>
<td>-5.01</td>
<td>&lt; 0.000000***</td>
</tr>
<tr>
<td>Dutch-Misaligned × Word × English</td>
<td>25.909 (3.331)</td>
<td>7.76</td>
<td>0.000000***</td>
</tr>
<tr>
<td>Misaligned × Word × English</td>
<td>-11.252 (3.360)</td>
<td>-3.33</td>
<td>0.00081***</td>
</tr>
</tbody>
</table>

Table B.1: Modelled logit of target fixation. Log-likelihood: -39897 (75078 observations).
REFERENCES


Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes* [Computer software manual]. (R package version 0.999375-39)


AUTHOR INDEX

Adams, T., 37, 205
Adriaans, E., 37, 61, 90, 115–117, 135, 155, 181, 182, 203
Albright, A., 33, 38, 56, 58, 60, 62, 203
Allen, J., 95, 204
Allopenna, P., 101, 102, 203
Anthila, A., 24, 178, 203
Aslin, R. N., 40, 101, 209, 212
Auer, E. T., 31, 36, 39, 153, 203, 213
Baayen, R. H., 44, 69, 72–75, 108, 162, 203, 211
Bailey, T. M., 36, 38, 42, 61, 70, 203
Barr, D. J., 109, 164, 203
Bataille, F., 105, 205
Bates, D., 73, 203
Bates, D. M., 72, 203
Berent, I., 1, 2, 26, 43, 44, 57, 58, 153, 203
Bergh, H. van den, 72, 108, 162, 212
Bever, T., 7, 212
Bever, T. G., 36, 38, 211
Blomert, L., 17, 210
Boardman, I., 11, 207
Boersma, P., 30, 34, 50, 56, 70, 147, 152, 178–181, 184, 187, 203, 204, 206
Bolker, B., 73, 203
Boll-Avetisyan, N., 40, 204
Brennan, S. E., 92, 208
Brent, M. R., 114–117, 204
Broe, M. B., 25, 206
Broersma, M., 64, 145, 204
Brown, R., 36, 41, 56, 204
Butterfield, S., 13, 210
Cairns, P., 114, 116, 117, 204
Cappa, S., 210
Cartwright, T. A., 114, 116, 117, 204
Chambers, C., 101, 212
Chambers, K. E., 90, 210
Charles-Luce, J., 31, 213
Chater, N., 114, 204
Chomsky, N., 22, 23, 27, 204
Christiansen, M. H., 95, 204
Church, K. W., 101, 158, 204
Coetzee, A. W., 30, 58, 147, 178, 181, 204
Cohen, M., 11, 207
Cohen, M. M., 13, 36, 41, 56, 60, 178, 180, 209
Coleman, J., 32, 36–38, 116, 204
Colomé, A., 90, 210
Cooper, F., 12, 208
Cutler, A., 8, 13, 18, 37, 46, 50, 65, 95, 96, 98–100, 114, 134, 138, 141, 144, 145, 151, 154, 155, 173, 186, 187, 204, 205, 210, 213
Dahan, D., 18, 101, 212
Daland, R., 116, 117, 205
Davidson, D. J., 72, 203
Davidson, L., 37, 42, 43, 86, 205
Davis, M. H., 95, 205
Dehaene, S., 37, 208
Dehaene-Lambertz, G., 56, 205
Dixon, J. A., 109, 209
Dijkstra, T., 8, 152, 206, 207
Donati, C., 210
Dupoux, E., 36, 37, 41, 42, 44, 57, 60, 65, 86, 91, 153, 205, 208
Dutoit, T., 105, 161, 205

215
Eberhard, K. M., 101, 213
Eckes, T., 68, 205
Elman, J. L., 5, 7, 13, 209
Ernestus, M., 10, 44, 205, 206, 211
Escudero, P., 50, 56, 152, 187, 206

Fazio, F., 210
Fisher, C., 90, 210
Fleischhacker, H., 62, 85, 87, 206
Fox, P. T., 41, 211
Frankish, C., 40, 206
Frauenfelder, U., 8, 36, 60, 206, 207, 212
Frieda, E. M., 152, 206
Friederici, A. D., 155, 208
Frisch, S. A., 25, 30, 36, 38, 61, 206
Fudge, E., 23, 206

Ganong, W. F., III, 8, 13, 57, 64, 206
Garrod, S., 17, 211
Gaskell, M. G., 95, 205
Gathercole, S. E., 40, 57, 206
Gaygen, D., 37, 45, 206
Goldinger, S. D., 9, 10, 207, 209
Greenberg, J. H., 23, 37, 207
Griffith, B. C., 7, 208
Grosjean, F., 157, 207
Grossberg, S., 10, 11, 14, 15, 17, 18, 39, 207
Grotjahn, R., 68, 205
Gulikers, L., 69, 203

Hagoort, P., 152, 207
Hahn, U., 36, 38, 42, 61, 70, 203
Halle, M., 22, 23, 27, 204, 208
Hallé, P. A., 36, 41, 56, 57, 59, 60, 153, 207, 212
Harris, K. S., 7, 208
Hayes, B., 30, 32, 38, 147, 178, 204, 207
Heuven, W. J. B. van, 152, 207

Hildum, D., 36, 41, 56, 204
Hintze, F., 14, 207
Hirose, Y., 36, 205
Hockema, S. A., 116, 117, 207
Hoffman, H. S., 7, 208
Homa, D., 10, 207
Hout, M. C., 10, 207
Idsardi, W. J., 56, 58, 60, 208
Itô, J., 65, 208
Jacquemot, C., 37, 41, 65, 208
Jaeger, J. J., 7, 56, 208
Jaeger, T. F., 74, 108, 208
Jakobson, R., 22, 27, 208
Jenkins, J. J., 23, 37, 207
Jusczyk, A. M., 155, 208
Jusczyk, P. W., 155, 208
Kabak, B., 56, 58, 60, 208
Kakehi, K., 36, 205
Kemmerer, D., 31, 213
Kessler, B., 25, 208
Kisseberth, C. W., 27, 208
Kolinsky, R., 56, 210
Kraljic, T., 92, 208, 212
Large, N. R., 36, 39, 57, 61, 84, 172, 206, 209
LeBihan, D., 37, 208
Legendre, G., 30, 147, 184, 208
Lenneertz, T., 1, 203
Levy, J., 114, 204
Liberman, A. M., 7, 12, 208
Lloyd, P. M., 65, 209
Lönngren, L., 135, 209
Luce, P. A., 9, 13–16, 19, 20, 31, 36, 37, 39, 45, 49, 57, 61, 62, 84, 85, 87, 99, 116, 153,
AUTHOR INDEX

172, 178, 181, 203, 206, 209, 213
Llugt, A. H. van der, 98, 101, 116, 209

Maechler, M., 73, 203
Magnuson, J. S., 101, 109, 203, 209, 212
Marslen-Wilson, W., 5, 209
Marslen-Wilson, W. D., 95, 205
Masaro, J., 26, 116, 214
Massaro, D. W., 13, 36, 41, 56, 60, 178, 180, 209
Mattys, S. L., 8, 18, 37, 45, 95, 101, 115, 143, 151, 209
McCarthy, J. J., 25, 209
McClelland, J. L., 5, 7, 13, 209
McMurray, B., 101, 209
McQueen, J. M., 5, 8, 13, 14, 17–19, 36, 37, 40, 45, 46, 95–98, 100–103, 107, 111, 117, 118, 143, 145, 146, 151, 153, 154, 158, 176, 191, 192, 209–212
Mehler, J., 36, 205
Melhorn, J. F., 8, 209
Mester, A., 65, 208
Meunier, C., 36, 207
Milliken, A. M., 10, 207
Milliken, L., 10, 207
Mirman, D., 109, 164, 209
Mitterer, H., 17, 177, 209, 210
Miyata, Y., 30, 208
Morais, J., 56, 210
Morciniec, N., 14, 210
Moreno Sandoval, A., 88, 210
Moreton, E., 1, 21, 43, 56, 58, 62, 210
Moro, A., 41, 60, 210
Myers, C. W., 11, 18, 207
Náñez, J. E., 40, 213
Newport, E. L., 40, 212
Norris, D., 5, 8, 13, 14, 17–19, 45, 96, 97, 100, 114, 143, 146, 172, 176, 205, 210
Nozawa, T., 152, 206
Ohala, J. J., 36–38, 210
Ohala, M., 36–38, 210
Onishi, K. H., 90, 210
Oostdijk, N., 66, 67, 69, 210
Pagel, V., 105, 205
Pallier, C., 36, 37, 90, 145, 205, 208, 210
Peaker, S., 40, 206
Perani, D., 210
Perruchet, P., 114, 211
Pertz, D. L., 36, 38, 211
Petersen, S. E., 41, 211
Pickering, M., 17, 211
Pickering, S. J., 40, 206
Piepenbrock, R., 69, 203
Pierrehumbert, J. B., 10, 25, 32, 36–38, 56, 66, 116, 204, 206, 211
Pierret, N., 105, 205
Pisoni, D. B., 36, 153, 206, 213
Pitt, M. A., 36, 40, 117, 153, 209, 211
Pluymaekers, M., 44, 211
Polivanov, E., 40, 56, 59, 65, 211
Polka, L., 56, 206
Prince, A., 26–29, 147, 177, 183, 211
Quené, H., 7, 18, 72, 95, 105, 108, 162, 211, 212
R Development Core Team, 73, 212
Raichle, M. E., 41, 211
Saffran, J. R., 40, 95, 117, 151, 212
Salverda, A. P., 18, 212
Samuel, A. G., 8, 57, 92, 208, 212
Sasaki, Y., 40, 213
Savin, H., 7, 212
AUTHOR INDEX

Schnitzer, R. D., 65, 209
Scholes, R. J., 23, 24, 37, 212
Schriefers, H., 152, 207
Sebastián-Gallés, N., 90, 210
Sedivy, J. C., 101, 213
Segui, J., 8, 36, 60, 206, 212
Seidenberg, M. S., 96, 204
Seitz, A., 40, 212
Selkirk, E. O., 23, 212
Shankweiler, D., 12, 208
Shaw, J., 37, 205
Shillcock, R., 114, 204
Smith, J. D., 10, 212
Smolensky, P., 26–29, 31, 147, 177, 183, 208, 211
Snyder, A. Z., 41, 211
Spivey-Knowlton, M. J., 101, 213
Steriade, D., 1, 203
Stone, G., 10, 14, 17, 207
Studdert-Kennedy, M., 12, 208
Svenkerud, V. Y., 155, 208
Tanenhaus, M. K., 101, 203, 209, 212, 213
Tees, R. C., 56, 214
Tettamanti, D., 210
Thiessen, E. D., 117, 212
Trapman, M., 141, 213
Treiman, R., 25, 208
Trubetzkoj, N. S., 7, 14, 64, 213
Umeda, N., 162, 213
Vaknin, V., 1, 203
Viebahn, M., 102, 209
Vinter, A., 114, 211
Vrecken, O. van der, 105, 205
Waals, J., 105, 162, 213
Warren, R., 8, 13, 213
Watanabe, T., 40, 212, 213
Werker, J. F., 56, 214
Wessels, J. M. L., 155, 208
Wetzels, W. L., 26, 116, 214
White, L., 8, 209
Wilson, C., 32, 38, 178, 207
Zonneveld, W., 26, 116, 214
INLEIDING

Kennis van mogelijke woordvormen is een mal voor spraakwaarneming, die deels de vorm van de waarneming bepaalt. De klankregels van een taal geven ten eerste aan of combinaties van klanken toegestaan zijn en onder welke voorwaarden (bv. ‘rm’ aan het einde van een lettergreep, maar niet aan het begin). Ook zijn er relatieve, probabilistische klankregels die weergeven dat sommige klankcombinaties heel vaak voorkomen (bv. ‘sp’), terwijl andere bijna nooit worden gebruikt (bv. ‘rîst’ bijna alleen in het woord ‘herfst’).

Klankregels kunnen worden bestudeerd als taalkundig fenomeen, maar omdat ze ook een effect hebben op de waarneming van spraak, moeten ze ook als psychologisch verschijnsel worden gezien. Dit proefschrift verkent de eigenschappen van bestaande taalkundige theorieën over klankregels. De vraag is of die theoretische eigenschappen terug te vinden zijn in het psychologische proces van spraakwaarneming.

Spraakwaarneming is door taalpsychologen uitvoerig bestudeerd. Er zijn diverse modellen van spraakwaarneming. Deze modellen gaan uit van een binnenkomend geluid dat gekoppeld moet worden aan representaties van woorden. Dit gebeurt in de meeste modellen door eerst de klanken te herkennen als fonemen, de betekenisvolle klanken van een taal. Aangezien woorden uit fonemen bestaan, kunnen woorden door fonemen geactiveerd worden. Spraak is ook lineair: klanken volgen elkaar op in de tijd. De volgorde van klanken is een belangrijk onderdeel van de representaties van woorden. De temporele structuur van taal is echter om nog een andere reden belangrijk: woorden moeten herkend worden zonder dat hun grenzen bekend zijn.

Als je naar spraak in je eigen taal luistert, meen je losse woorden te horen, maar dit is een illusie; de spraak bevat geen duidelijke markering van de grenzen tussen woorden, zoals de spaties in geschreven taal. De perceptie van spraak als een stroom woorden komt maar ten dele door herkenning van de woorden zelf. De illusie treedt ook op voor nonsens-spraak die geen woorden bevat. Een luisteraar gebruikt dus ook kennis van mogelijke woorden, oftewel de klankregels van de taal, om in de spraakstroom de meest ‘woordvormige’ stukjes te isoleren. Als een klankcombinatie volgens de klankregels niet binnen een woord kan voorkomen, kan de spraakstroom worden opgebroken in het midden van die combinatie.

Klankregels kunnen ook gebruikt worden om onduidelijke spraak waar te nemen. Als een stukje spraak klinkt als twee verschillende klankcombinaties,
waarvan er maar een legaal is volgens de klankregels, dan is die legale combinatie waarschijnlijk bedoeld door de spreker. Uit eerder onderzoek blijkt dat luisteraars inderdaad klankregels toepassen op onduidelijke spraak, maar ook dat ze woorden met veel voorkomende klankcombinaties makkelijker verwerken.

Aangezien spraakwaarneming bijna altijd woordherkenning tot doel heeft, wordt in de inleiding de architectuur van verschillende woordherkenningsmodellen en de mogelijke plaats van klankregels in die modellen besproken. Het uitbreiden van modellen van woordherkennings met klankregels maakt het mogelijk dat een aantal observaties de effecten van klankregels op spraakherkenning in die modellen opgenomen kunnen worden, waarmee ze accurater worden.

De vraag is dan op welke manier klankregels in modellen van woordherkenning moeten worden opgenomen. In de fonologie worden klankregels op verschillende manieren beschreven. In theorie kunnen klankregels absoluut zijn (‘mr’ is nooit toegestaan in een Nederlandse lettergreep, ‘pr’ wel), maar ook relatief (‘sp’ komt vaker voor dan ‘sl’). Ook kan kennis van klankregels bestaan uit verboden of juist kennis van de welgevormdheid van klankcombinaties. Het is theoretisch niet ingewikkeld om één theorie te construeren waarin alle kennis over klankcombinaties is vastgelegd als een waarde op een graduele welgevormdheidsschaal; om het verschil tussen legaal en illegaal uit te drukken kan een bepaalde waarde op de welgevormdheidsschaal als de ondergrens van welgevormdheid worden beschouwd. Alle combinaties die onder die ondergrens vallen, zij dan niet legaal.

Het meest geschikte model van klankregels is een model waarin verschillende beperkingen (constraints) de klankregels, zowel positief als negatief, uitdrukken. De interactie tussen de beperkingen kan gemodelleerd worden met gebruik van Optimaliteitstheorie. Een klankgrammatica in termen van constraints is echter nog steeds een theoretische constructie. Om deze constructie te toetsen aan empirische gegevens over spraakwaarneming zijn experimenten nodig. Dit proefschrift beschrijft een serie experimenten waar mee de theoretische verschillen tussen de waarschijnlijkheid van klankregels, illegaliteit van klankcombinaties, graduele welgevormdheid en gemaakteerdheid getest zijn. De vraag is of al deze theoretische begrippen hun eigen effect op spraakwaarneming hebben.

**WAARSCHIJNLIJKHEID EN ILLEGALITEIT BIJ WOORDHERKENNING**

De eerste vraag die wordt aangepakt is of leerders van een tweede taal die worden geconfronteerd met de frequenties van klankcombinaties in een vreemde taal eerst hun kennis van gemaakteerdheid van die combinaties in de
eerste taal opzij zetten, voordat ze graduele verschillen opmerken en daarmee de welgevormdheid van klankcombinaties leren.

Als alle grammaticaliteit van klankcombinaties uit hun waarschijnlijkheid of frequentie wordt afgeleid, dan zou eerst grammaticale illegaliteit met haar effecten moeten verdwijnen voordat meer graduele verschillen in welgevormdheid hun effect laten zien. Dit bleek echter niet het geval te zijn.

Spaanstalige en Japanstalige leerders van het Nederlands werden in hoofdstuk 2 vergeleken met luisteraars met Nederlands als hun moedertaal. In het Spaans en Japans zijn combinaties van een ‘s’ en een medeklinker niet legaal en de tweedetaalluisteraars horen deze combinaties mogelijk verkeerd, met een extra klinker die er eigenlijk niet is. Dit bleek het geval voor de Spaanstalige luisteraars; zij activeerden een woord als ‘stad’ even makkelijk wanneer ze ‘stad’ hoorden als wanneer ze ‘estad’ hoorden. Dit was een voorspeld effect: omdat ‘st’ niet legaal is volgens de klankregels van het Spaans, voegen Spaanstalige luisteraars in hun waarnemingen een illusionaire klinker toe, een ‘e’, die het cluster splitst over twee lettergrepen (‘es’ en ‘tad’). De afwezigheid of aanwezigheid van die klinker in het geluidsfragment leidt dus tot dezelfde waarneming.

Nederlandse luisteraars lieten zien dat ze profiteerden van de hogere frequenties van sommige s-medeklinker-combinaties; woorden met die combinaties herkenden ze makkelijker, behalve als er een klinker aan was geplakt (zoals bij ‘estad’). Spaanstalige luisteraars die beter waren in het Nederlands waren ook beter in de combinaties met hogere frequenties, maar hoorden desalniettemin het verschil tussen de afwezigheid en aanwezigheid van een klinker voor de s-medeklinker-clusters niet. Dit duidt op de aanwezigheid van twee verschillende mechanismen: één mechanisme is het wegfilteren van illegale waarnemingen uit de spraak, terwijl het andere mechanisme de snellere herkenning van veelvoorkomende combinaties verzorgt. Deze twee mechanismen spreken elkaar bij de Spaanstalige leerders tegen. Dit toont aan dat de kennis van luisteraars over klankregels van twee verschillende types is: categoriale en graduele kennis.

SPRAAKSEGMENTATIE

Woordgedreven of klankregelgedreven segmentatie

Spraaksegmentatie is een gebied waar klankregels invloed kunnen hebben, die precies tegengesteld is aan het effect op losse woorden. Kennis dat een spraakcombinatie illegaal is kan spraaksegmentatie helpen en bepaalde segmentaties sterker activeren dan andere, terwijl dat soort kennis bij losse woorden juist de waarneming van de illegale combinaties onderdrukt en dus
een negatief effect heeft. In hoofdstuk 3 werd vastgesteld dat klankregels zorgen voor de splitsing van de spraakstroom voordat er woorden in worden herkend.

Om dit vast te stellen is gebruik gemaakt van eye-tracking. Gedurende het experiment werden de oogbewegingen van de proefpersonen vastgelegd. Ze waren gevraagd om te luisteren naar korte stukjes spraak en te reageren als ze een woord herkenden. Ze konden reageren door naar de eerste letter van het woord te kijken. Op een computerscherm was die letter te zien, samen met drie anderen. Deelnemers aan het experiment probeerden zo snel mogelijk naar de goede letter te kijken. Hun oogbewegingen onthullen welke letter ze de meest waarschijnlijke woordgrens vinden.

De proefpersonen herkenden woorden zoals ‘rok’ in ‘fiemrok’ en ‘fiedrok’, maar in het laatste geval veel moeilijker. Dit was al bekend, maar er zijn meerdere verklaringen voor. Het experiment liet zien dat ook zonder een woord, bijvoorbeeld in het geluid ‘fiemruupt’ of ‘fiedruupt’, de klankregels invloed uitoefenen op de plek waar de proefpersonen de woordgrens vermoeden. Omdat ‘mr’ illegaal is, zullen ze bij de ‘r’ een woordgrens vermoeden in ‘fiemrok’ en ‘fiemruupt’. De klank ‘d’ is in het Nederlands echter niet toegestaan aan het einde van een woord. In ‘fiedrok’/‘fiedruupt’ is het dus extra moeilijk om te splitsen bij de ‘r’, omdat in de segmentaties ‘fied.rok’ en ‘fied.ruupt’ het illegale residu ‘fied’ is ontstaan (met een ‘d’, niet uitgesproken als een ‘t’, aan het einde). Het is echter de vraag of luisteraars weet hebben van beide klankregels, of slechts van één. Bovendien is het ook mogelijk dat ze, omdat ze veel woorden met ‘dr’ en geen woorden met ‘mr’ kennen, de fragmenten ‘fiedrok’ en ‘fiedruupt’ willen splitsen bij de ‘d’. Als dat zo is, zou het effect helemaal niet veroorzaakt zijn door kennis van klankregels, maar juist door kennis van woorden.

Categoriale en graduele welgevormdheid

Met de eye-tracking-methode die hierboven beschreven is, kon de invloed van klankregels op segmentatie nog duidelijker worden bestudeerd. Experiment 4a in hoofdstuk 3 liet zien dat het effect voor een deel te danken is aan het vermijden van illegale segmentaties, zoals ‘lud.rib’, vergeleken met ‘lut.rib’. In deze vergelijking is het feit dat er veel woorden beginnen met ‘dr’ niet van invloed, omdat er ook veel woorden met ‘tr’ beginnen. Die alternative verklaring gaat dus niet op.

De resultaten van Experiment 4a laten verder zien dat luisteraars altijd eerder proberen een illegaal cluster te splitsen dan een legaal cluster. Echter, alle resultaten zijn gebaseerd op een vergelijking tussen legale en illegale
clusters. Het verschil is daarom zowel te verklaren met illegaliteit als met legaliteit.

Eerder onderzoek heeft uitgewezen dat leerders van een vreemde taal hun kennis van illegaliteit van de eerste en tweede taal combineren en dus clusters splitsen als ze illegaal zijn in één van de twee talen. Indien de kennis die gebruikt werd de welgevormdheid van legale clusters zou zijn, dan zouden clusters die alleen illegaal zijn in één taal niet gesplitst worden, maar juist bij elkaar gehouden worden. Toch is het nog niet uitgesloten dat kennis van graduele welgevormdheid, naast kennis van illegaliteit, ook een invloed heeft. Om daar achter te komen werden ook luisteraars naar een tweede taal getest, in Experiment 4b. Hun eerste taal was een Slavische taal, waarin alle clusters die in Experiment 4a gebruikt werden legaal waren. Experiment 4b was gelijk aan Experiment 4a, waardoor de moedertaalluisteraars en tweedetaalluisteraars vergeleken konden worden. Het bleek dat alleen de eerste groep segmentaties waarin de illegale clusters intact waren helemaal verwierpen. De tweedetaalluisteraars hebben ook de voorkeur voor het splitsen van deze clusters. Echter, als dat geen woord oplevert, dan proberen ze alsnog de segmentatie uit waarin een illegaal cluster niet gesplitst is. Dit toont aan dat ook graduele verschillen in welgevormdheid gebruikt worden in spraaksegmentatie.

TAALMODUSEFFECTEN

De resultaten van de experimenten met tweedetaalleerders in de hoofdstukken 2 en 3 suggereren dat het erg moeilijk is om de klinkregels van twee talen uit elkaar te houden. De vraag is of klinkregels zijn geregistreerd als onderdeel van de taal waar ze betrekking op hebben, of dat er maar één set klinkregels is, waar tweedetaalleerders de klinkregels van twee talen in hebben opgeslagen. In een experiment met Nederlandse moedertaalsprekers die een Engels taal of Nederlandse taal uitvoerden, met dezelfde eye-tracking-methode, blijkt dat deze tweedetaalluisteraars de klinkregels toepassen van de juiste taal. Dit toont aan dat klinkregels, die taalkundig bij één bepaalde taal horen, ook psychologisch gezien aan één taal zijn gekoppeld.

DISCUSSIE EN CONCLUSIE

Met een serie experimenten laat dit proefschrift zien dat zowel strikte kennis van illegale combinaties als graduele kennis worden gebruikt om de spraakstroom in woordvormige stukjes te dwingen. Illegale combinaties worden weggefilterd of aangepast, of gebruikt om woordgrenzen te vinden, zoals in ‘de lamp viel’: omdat ‘pv’ illegaal is in Nederlandse lettergrepen, wordt de
De luisteraar heeft daarnaast een voorkeur voor stukjes spraak met veelvoor-
komende klankcombinaties, om zo snel en makkelijk mogelijk woorden te
horen. Luisteraars kunnen de klankregels van twee talen die ze goed kennen
uit elkaar houden. Meestal wordt een vreemde taal echter wel gefilterd door
de strikte regels van de eerste taal, wat het moeilijk maakt om woorden te
isoleren bij het luisteren naar vreemde talen.

De klankregels van een taal kunnen worden beschouwd als zowel po-
sitief als negatief geformuleerd. Deze kennis kan beschreven worden in
de vorm van constraints, beperkingen. Deze beperkingen geven aan welke
klankcombinaties vermeden moeten worden in perceptie en welke makkelijk
herkend moeten worden. Soms zijn de beperkingen in tegenspraak. Het
raamwerk van Optimaliteitstheorie biedt een manier om de tegenspraak en
de oplossing daarvan te formaliseren. In het laatste hoofdstuk wordt een
dergelijke formalisering in globale termen voorgesteld. Het belangrijkste
kenmerk van de formele grammatica is dat verschillende oplossingen voor
de waarneming van binnenkomende spraak worden geëvalueerd. De meest
welgevormde oplossing is echter niet altijd de enige die in de rest van het
spraakwaarnemingsproces wordt gebruikt. Indien de waarneming die vol-
gens de klankregels het beste is, geen woorden bevat, wordt een volgende
waarneming uitgeprobeerd.

Modellen van woordherkenning kunnen worden uitgebreid met de formele
klankgrammatica die is voorgesteld in hoofdstuk 5. De klankgrammatica moet
een plaats krijgen tussen de herkenning van fonemen en de herkenning van
woorden.
Tomas Ostar (Tom) Lentz was born in Zaanstad in 1981. At the age of four, he moved to Gouda, where he grew up. He attended the Coornhert Gymnasium and graduated in nine courses, including five languages. He pursued his interests in languages during his study of Artificial Intelligence at the University of Amsterdam in 1999; he also moved to Amsterdam. During his study in Amsterdam, Tom added a Minor in Cognitive Science (which would later become the Research Master in Cognitive Science) and most of the Bachelor in Spanish Language and Culture to his curriculum. In 2003, he spent a semester abroad, studying Translation and Interpretation at the Autonomous University in Barcelona.

In 2006, Tom wrote a thesis on the algorithmic formalisation of the semantics and pragmatics of tense and aspect and obtained his MSc in Artificial Intelligence: Language and Speech, with distinction. In 2007, Tom started working on his PhD research at the Utrecht Institute of Linguistics OTS, part of Utrecht University.

Currently, Tom works on statistical methods for data analysis with Paula Fikkert at the Radboud University Nijmegen, as a teacher of Psycholinguistics at the Amsterdam University College and as a research assistant at the University of Amsterdam.